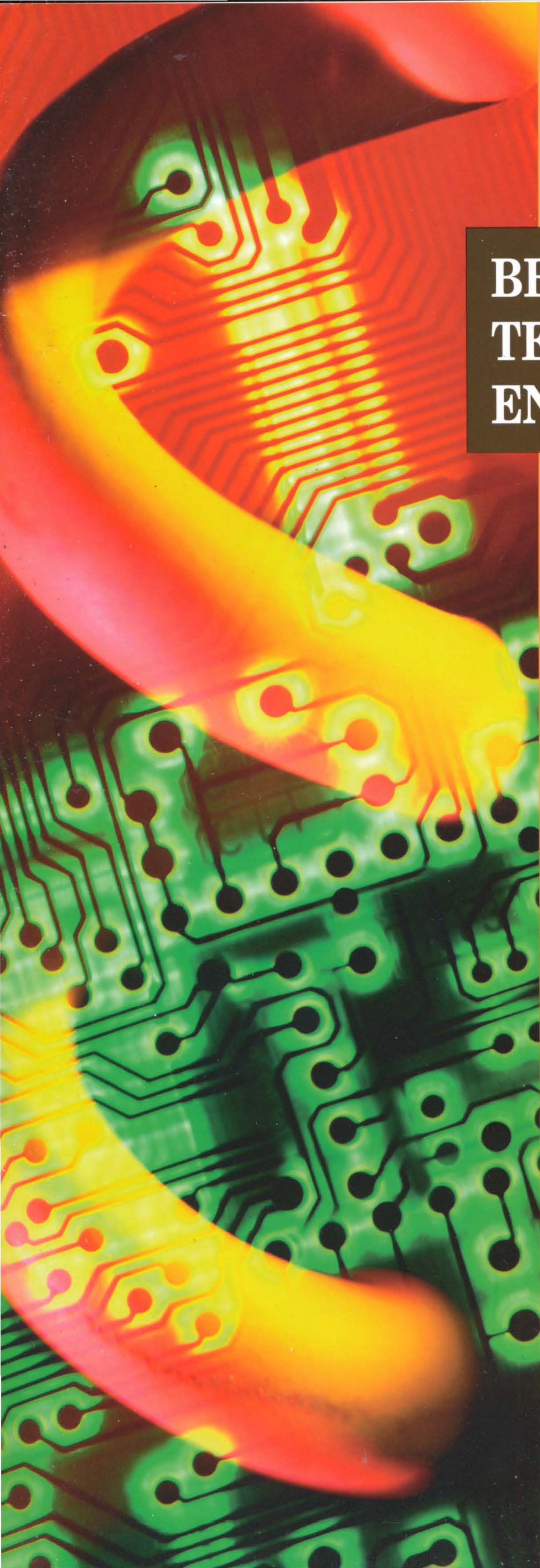


VOLUME 18 PART 2 SPECIAL AUGUST 1999 ISSUE

BRITISH TELECOMMUNICATIONS ENGINEERING

Included in this Issue

Proceedings of FITCE '99— the 38th European Telecommunications Congress 'Networking the Future'

The Journal of The Institution of British Telecommunications Engineers

BRITISH TELECOMMUNICATIONS ENGINEERING

VOL 18 ■ PART 2 ■ SPECIAL AUGUST 1999 ISSUE

Dear Reader

This edition of *British Telecommunications Engineering Journal* has been given over to the Proceedings of this year's Congress of the Federation of Telecommunications Engineers of the European Community (FITCE), held in Utrecht, The Netherlands, from 24–28 August 1999. The congress has the theme 'Networking the Future'. We have been pleased, once again, to be able to work with FITCE in producing the Proceedings for the Congress and to make the papers available to the wider IBTE audience through this special edition. The editors would like to thank the Comité de Direction of FITCE for giving permission to publish the papers.

Since the size of this edition is much larger than normal, it has not been possible to include the *Structured Information Programme*. However, this will continue with the next, October 1999, edition of the *Journal*.

I would like to draw readers' attention to our special Millennium Article Competition (see panel below). This is a wonderful opportunity for readers to be part of a very special edition. So start writing and get your articles to us before 15 October.

Yours sincerely,

Paul Nichols (Managing Editor)

Millennium Article Competition

Write a captivating article on the theme
'Communications—Into the New Millennium'
for the Special Millennium Edition of the *British Telecommunications Engineering* journal and win £400 plus a trophy.

The January 2000 edition of the *British Telecommunications Engineering* journal is to be a Special Millennium Edition containing specially commissioned articles. The winning article from this competition will also be published in this unique edition.

Articles for this competition are welcome from anybody, Members and non-Members of the IBTE alike. Articles can tackle any topic related to communications and discuss how it is responding to the changing environment. They should be well illustrated and contain no more than 5500 words. A 150-word (approx.) biography and colour photograph of each author must also be included. Articles not selected as the winning article may be published in subsequent editions throughout 2000. The closing date is **15 October 1999** and entries (which should be in electronic form) should be sent to: The Managing Editor, IBTE Office, Post Point 2D05, The Angel Centre, 403 St John Street, London EC1V 4PL. E-mail: paul.e.nichols@bt.com.

For further information, call the IBTE HelpLine on 0800 028 0209, or on +44 171 843 7622 outside the UK.

All entries will become the property of the IBTE. The Editor's decision is final.

British Telecommunications Engineering Journal, Post Point 2D05, The Angel Centre, 403 St. John Street, London EC1V 4PL **ISSN 0262-401X**

38th European Telecommunications Congress
Utrecht, The Netherlands
24–28 August 1999

Proceedings

NETWORKING THE FUTURE

FITCE Secretariat
Belgacom Towers, T9
Boulevard Emile Jacqmain, 177
B-1030 Bruxelles
Belgium
Tel: +32 2 202 77 05
www.fitce.org

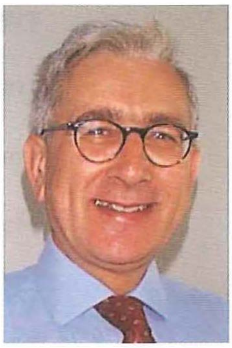
Publication produced by
The Institution of
British Telecommunications
Engineers
Post Point: 2D05
The Angel Centre
403 St. John Street
London EC1V 4PL
UK
Tel: +44 20 7843 7622
www.ibte.org

Contents

Contents continued

Guntram Kraus

Foreword



Ladies and Gentlemen,

The FITCE Congress in London in 1998 showed us that it is possible to organise a congress where the presentations are not too long. Of course, it is possible to say something important within a quarter of an hour. As a result, the paper selection committee has succeeded in preparing an event with over 40 presentations in a compact form. My thanks go to the colleagues responsible for this.

I am delighted by the high number of speakers and, in particular, that we have been able to secure speakers from the USA as it is one of our aims to reflect the globalisation in telecommunications within the FITCE and the FITCE Congresses, too. A first step in this direction is the invitation of speakers from the USA to report on the latest developments in their country.

I think the selection of the presentations and the spread between European countries means we can look forward to a particularly interesting event.

As in previous years, the FITCE will once again honour the best presentations with a prize.

It only remains for me to wish you, the listeners and readers, interesting new findings and many fruitful personal discussions.

Guntram Kraus

Wim van der Bijl

Networking the Future

Welcome to Utrecht. A city where past and present come together in an interesting environment to see and enjoy.

This year's conference, being the last one of this millennium, will bring together present and future of telecommunications. Many authors will give their views on the development of telecommunications in the coming years of the next decade.

Looking further than the first years in the next millennium is very difficult since we are experiencing the same major changes as society did a few hundred years ago. A period we now call the *industrial revolution*. It is difficult to imagine how people saw the future in those days when the steam engine was invented and deployed. Probably no one could predict that we all would leave for work every morning to go to the factories where the machines dictated the speed of working. And still today we leave everyday home to go to the office, first spending a lot of time in traffic jams, and sitting behind a computer screen where e-mails and telephone calls are dictating the speed of work.

Who can predict how this will evolve during the coming *information revolution*? In the current turbulent years of telecommunication we try to cope with deregulation, globalisation, enormous growth of mobile and Internet users and growing capabilities created by breakthroughs in technology. How long will we continue with doubling the processing power every 18 months, doubling the bit rate per fibre every 12 months? Digitalisation and software capabilities in all equipment leads to a convergence between the traditional worlds of telecommunications, IT and broadcasting leading to a plethora of new services based on combinations of the different worlds. Viewing your favourite TV programme via the Internet on your multimedia PC is one example. Buying new books using web TV is another. These are just a few examples of some combinations. Clearly many new combinations will ultimately lead to the situation that all individuals can communicate everywhere with everyone at any time using a personal communication device. This same device will also be used to entertain oneself or together in common interest groups. Soon enough it will be technologically and economically viable. At the same time it is fulfilling a basic requirement of the human being: communicating with other human beings.

The developments in technology, the changes in economic and social environments are all based on the fact that we as human beings are able to build upon the knowledge from other human beings through our sophisticated ability to communicate thoughts, ideas and experiences. This is merely also the basic reason for our conference: coming together, listening to the presentations, exchanging ideas during the breaks and the enjoyable evening sessions. All activities focused on broadening the mind, sharpening the vision and expanding the personal network.

Networking the future is the theme the coming days during the conference. I thank all authors for their effort in sharing their vision on the future or the problems of today with all participants. I thank also you as a participant for joining and interacting in this interesting event focused on our common future in the next decade, centennial and millennium.

Wim van der Bijl

List of Authors

Session Chairmen

Wim van der Bijl	1. Voice/Data Convergence
Bruno Jacobfeuerborn	2. Mobile Communications
Christine Jominet	3. Managing the Network
Prof. Gerry White	4. Futures
Bruno Jacobfeuerborn	5. Legal Framework
Antonio Como	6. IP Technology
Antonio Como	7. Planning the Network
Lise Baudouin	8. Market Dynamics
Prof. Keith Ward	9. Charging Innovations
Egied Dekoster	10. Access Network
Egied Dekoster	11. IP-PSTN
John Lysaght	12. Applications
Wim van der Bijl	13. E-Commerce

Opening Ceremony—Keynote Speakers

Mrs. J. M. de Vries	Vice Minister of Traffic and Watermanagement
Mr. Peter Scott	Department Head of DGXIII of the EU
Mr. Peter Morley	Chief Technical Officer KPN
Mr. Dieter Kaiser	Vice-Chairman of the ETSI Board

Workshop on Internet2

Chairman/moderator	Mr. Frans Heitkamp	Intercai
Keynote speakers	Mr. John Silvester	University of Southern California
	Mr. George Böhm	Deutsche Telekom AG
	Mr. Kees Neggers	SURFnet BV

Francis Pereira and Elizabeth Fife

The Internet: Business Efficiencies of Merged Voice and Data Communications

Companies that have deployed Internet protocol (IP)-based applications in their core business processes have realised business efficiencies emanating from supply chain integration, just-in-time inventory management, real-time price and availability quotes for products, increased productivity and improved, albeit less costly, customer support. Several of these applications, particularly web-based customer support, will be adopted by a majority of companies within the next 3–5 years. In leading-edge companies, however, the re-engineering and integration of internal and external business processes was a prerequisite for the successful adoption of these applications. Although such re-engineering and integration are fraught with technical and organisational complexities, it may well reflect how successful global corporations have responded to the challenges and pressures of the new networked, and competitive business environment.

Francis Pereira and Elizabeth Fife:
Center for Telecommunications Management (CTM),
DCC 217
Marshall School of Business
University of Southern California
University Park,
Los Angeles, California 90089-0871
Tel: + 1 213 740 0980
E-mail: pereira@rcf.usc.edu **or** fife@rcf.usc.edu

Introduction

The phenomenal growth of the Internet, coupled with stellar increases in Internet-based sales by companies such as Amazon.com and Dell Computers, has generated tremendous interest in the consumer benefits of e-commerce, and the Internet consumer market. However, while some retail companies will experience meteoric increases in Internet-based sales in the next 3–5 years, many will not. Indeed, even the most optimistic projections of $200 billion in Internet sales by the year 2000 would represent less than 5% of total retail sales by US consumers, and would be negligible if it accrues to all sectors.

However, a more pervasive, though less trumpeted, use of the Internet has already been occurring in the business-to-business sector. A large number of companies have already deployed IP-based applications, from the now common place web-based customer support functions to more elaborate applications, such as inventory management. By studying Marshall Industries, the Automotive Network Exchange (ANX), Cisco Systems and Wal-Mart, this paper examines cases where the Internet and IP-based applications have been reshaping the business landscape and revolutionising the business model, and identifies business efficiencies that have been attained through the merging of voice and data applications.

In summary, such efficiencies extend beyond cost savings from a single voice and data network, to include more extensive benefits accruing to just-in-time inventory management, real-time price and availability quotes for products and improved, albeit less costly, customer support. For example, this paper estimates that Marshall Industries, through its use of IP-based applications, has achieved cost savings of some $145 million from reduced sales, general and administrative costs between 1994 and 1998. Similarly, Wal-Mart has been able to shave some 0·04% annually off its sales, general and administrative costs, which translates to some $400 million in savings for the retail giant. Based on a survey of US companies by CTM, this paper shows that several of these IP-based applications will be deployed extensively by a majority of businesses within the next 3–5 years. The results of this survey, together with a review of findings from other surveys, support the view that the Internet revolution is rapidly and 'insidiously', albeit unglamorously, revamping business processes.

Marshall Industries

With revenues of $1·46 billion in 1998, Marshall Industries, is the fourth largest electronic components distributor in the United States and operates in an industry which is highly competitive and characterised by rapid technological change and fast product cycle times†. The industry, in general, has been experiencing tremendous growth, and many companies have seen their annual revenues double between 1993 and 1998, due to the increased global demand for electronics and computer equipment*. Marshall has

† The largest two distributors, Arrow Industries and Avnet Incorporated, had net sales of some $8·34 billion and $5.92 billion in fiscal 1998, respectively.

* Part of the growth in net sales for some of the companies were due to acquisitions.

developed a strong intranet/extranet strategy that has made the company more successful than its competitors at delivering some 125 000 different products from more than 150 suppliers to over 100 000 customers, and tracking these deliveries each day.

Marshall's reorganisation

Although many observers have credited Marshall Industries' growth and performance to its implementation of information technology (IT) systems, it is important to note that several significant changes preceded the implementation of IT. First, beginning in 1991, Marshall Industries reorganised its corporate structure and adopted a 'very flat' organisational structure to speed up decision making and direct contact with customers. The IT department at Marshall was eliminated in 1993. Many IT projects were late or had failed to meet user expectations, and Marshall Industries felt that the best way to overcome these problems was to integrate IT into the business units[1]. Secondly, in response to growing competition in the industry, it abolished the Management by Objective System and adopted a customer-oriented global performance measure. It abolished all sales commissions and management incentives based on specific objectives in favor of one criteria: overall corporate profits. This was done to prevent friction between traditional sales processes and the new sales channel of cyberspace. Marshall Industries made the customer the primary focus with the object of delivering the lowest possible price, with the highest quality and service, and in the fastest possible time, as captured by its impossible mantra, 'Free, Perfect, Now'.

Since its web efforts began in 1994, Marshall has gained nearly 200 000 new distinct customers through a network which handles nearly 700 000 transactions per day. In addition, productivity per employee has increased from \$360 000 in 1991, to \$635 375 in 1998[2].However, this represents a decline from \$846 150 in 1997, primarily due to an increase in staff as a result of its acquisition of Sterling Electronics Corporation. As Figure 1 also shows, Marshall's sales, general and administrative expenses, as a percentage of net sales, have decreased with the introduction of its web efforts, from 16·6% in 1993 to 11·2% in 1998[3].

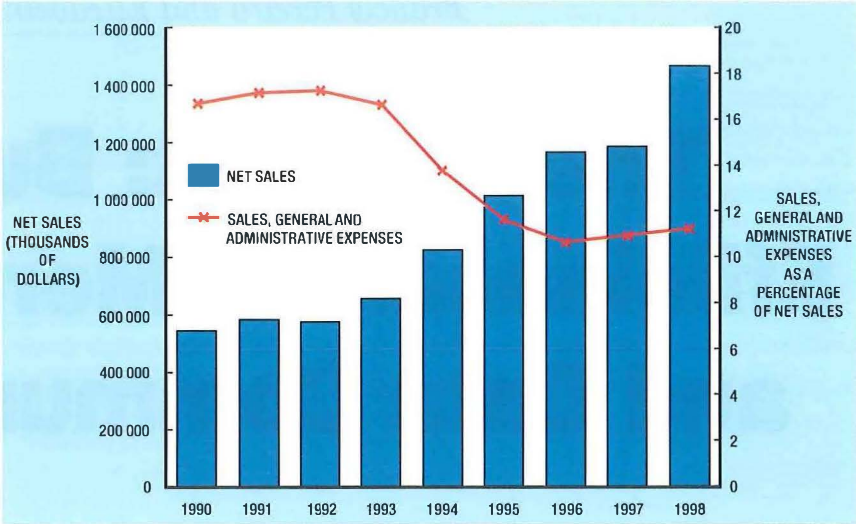


Figure 1 – Marshall Industries

To a large extent, Marshall's success is the result of effective management of technology—applications are tailored to their specific operations, and are thoroughly integrated throughout all units of the company—sales, marketing, and corporate management. Furthermore, greater communication among its customers, suppliers, and partners was achieved through: a restructuring of the company which forced employees to focus on the customer; implementation of a 24 hour help desk for customers to talk to a live person via telephone or the Web; and through Marshall's international partnerships and alliance activities. The corporate redesign was facilitated by the implementation of the IT systems[4].

Marshall's intranet

Marshall's web site was designed to provide customers with information on products, pricing and availability via an object-based relational database. The site currently registers over 1 million hits each week from customers in over 59 different countries. It contains information about 170 000 part numbers, 100 000 pages of data sheets and real-time inventory pricing and availability from over 100 suppliers. The site provides customers with their purchase history and allows them to order parts, request samples and track their orders on-line, in conjunction with United Parcel Services. Marshall's site also provides customers with electronic industry news using RealAudio broadcast. In addition, visitors can chat with Marshall's engineers on-line and in real-time 24 hours of the day to attain product information or help in trouble-shooting problems. Suppliers are also provided with customised pages on the site.

Marshall Industries' intranet, christened GlobalNet, was commissioned on 24 July 1995 at an undisclosed cost, although industry estimates put it at some \$50 million. Marshall's intranet is part of a global wide area network (WAN) upon which 13 client/server platforms sit. It includes an IBM 3090 400J mainframe running a DB2 relational database. The company has been deliberately vague about the specific technical aspects of its intranet for proprietary reasons. Mobile workers use Lotus Notes when they are on the road. A systems integrator DPI Services coordinated the project of automating of the sales force. Key applications include[5] the Distribution Resource Planner (DRP) system, a database management system that supports customers and suppliers' inventory and shipping scheduling[2]; the Quality, Order, Booking, Resell Application (QOBRA), a proprietary system used for day-to-day order management; the Automated Shipping and Receiving System (AS/RS) that interfaces with the automated warehouse for robotic movement of the inventory[2]; and the Education, News and Entertainment Network (ENEN) that allows suppliers to give live seminars about their products, and archive the presentations for potential customers in the future[2]. GlobalNet thus affords Marshall's salespeople real-time or near real time price quotes and inventory status from customer premises and provides virtual interaction between potential customers and suppliers to design new products through use of web-casting technology.

Marshall Industries' performance

The increasingly competitive nature of the electronics distribution industry has generated intense downward pressure on gross margins. Following Marshall's lead many distributors have implemented, to varying degrees, IT solutions in their business processes, such as Arrow Industries' Worldwide Stock Check System and, Pioneer Standard's Systems Integration Value-Added Center (SIVAC). Although use of IP-based solutions in business processes is not unique to Marshall Industries, they have utilised these applications more comprehensively and more effectively.

Marshall Industries' seemingly phenomenal success is testimony to the power of customised IT solutions and the Internet as a means to differentiate companies from competitors. As net sales have increased in the electronics distribution industry, many of the larger firms have responded by increasing the number of employees. However, Marshall has managed to accommodate a near doubling of its revenue since it began its comprehensive shift to e-commerce four years ago, while keeping the number of employees relatively constant. Although accurate quantification of gains in employee productivity and customer service attributable solely to IT-based applications is not easily calculable, the evidence strongly suggests that greater efficiency and service have been achieved. This is reflected aptly in Figure 2.

Two features in Figure 2 are noteworthy. First, Marshall Industries has been able to achieve higher levels of employee productivity compared with some of its competitors. Secondly, it has been able to achieve greater increases in productivity between 1992 and 1998 than its competitors. Increased competition in the industry has pressured distributors to find ways to lower or contain costs, particularly through the adoption of IP-based applications in business processes. Marshall appears to have been successful in this respect, as shown in Figure 3. Juxtaposing Marshall Industries' sales, general and administrative costs against the industry average, and accepting a natural downward trend of these costs for the industry, Marshall's ability to lower its costs below the industry average translates to an estimated total savings of some \$143·7 million between 1994 and 1998.

Wal-Mart

Wal-Mart is considered the leader in quick-response merchandise replenishment, using massive parallel processing, and supply-chain management to ensure the uninterrupted flow of goods to its 3400 stores around the world. Tightly knit relationships have been formed with business partners in order to streamline the supply chain, and to keep inventories low. In 1997, Wal-Mart, embarked on an aggressive programme to increase IT development staff by 40% and quickly deliver applications aimed at reducing store inventories and speeding the supply chain. As a result, profits grew 14% on a 12% sales jump for the first nine months of Wal-Mart's 1997 fiscal year, and third-quarter profits jumped by 16% on an 11% revenue advance. More crucially, third-quarter inventories at Wal-Mart's US stores were lower than year-earlier levels, an achievement directly attributable to better and faster information in the supply chain[6]. Currently, more than 90% of Wal-Mart's software is developed internally by 1000 full-time programmers. Despite the company's expertise in making software, they spend less on technology as a percentage of sales than other retailers, 0·5% in 1998, versus 1·43 % for the industry[7].

Data warehousing

Wal-Mart's extranet, Retail Link, is a private network for sharing inventory and sales data with some 5000 merchandise suppliers and is one of the first large-scale examples of vendor-managed inventory over the Internet. Wal-Mart has built one of the world's largest data warehouses which is used to track everything possible about their merchandise[8] and meant to essentially streamline the information supply chain[9]. Its multimillion dollar data warehouse stores some 43 terabytes of data and enables detailed analyses of data on point of sale, inventory, products in transit, market statistics, customer demographics, finance, product returns, and supplier performance. The data is used for three broad areas of decision support: analysing trends, managing inventory, and understanding customers. What emerges are 'personality traits' for each of Wal-Mart's 3000 or so outlets, which Wal-Mart managers use to determine product mix and presentation for each

Figure 2—Net sales per employee for selected distributors, fiscal 1992–1997 (in thousands of dollars)

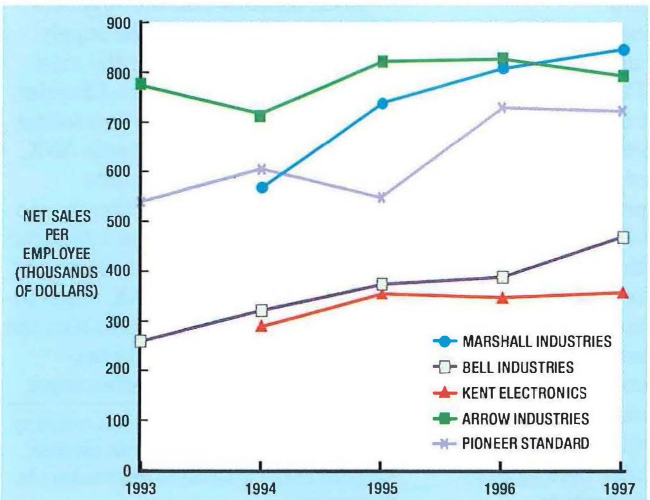
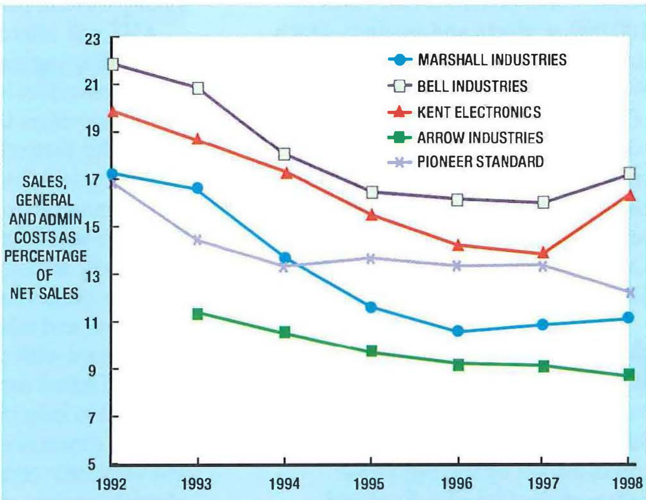


Figure 3—Sales, general and administrative costs as percentage of net sales, fiscal 1992–1998



store. The information helps Wal-Mart determine the right product on the appropriate shelf at the lowest price and the optimum placement of goods in the store so as to maximise sales of related products and cross-marketing. The purchasing information is sent to product manufacturers over the Internet which allows manufacturers to re-supply or cut prices if a product is not selling well. This affords Wal-Mart a major advantage over most of its competitors[10].

Since information from the individual operational databases is copied into the data warehouse at regular intervals, anyone in a retailer's organisation can access information from the data warehouse without disrupting the often critical applications the donor databases were designed to support. Techniques using three-dimensional visualisation tools show data and make predictions about what consumers will buy based on factors like ethnicity, weather patterns, local sports affiliations, and 10 000 other traits[10]. The company has a network called *CarrierLink* for firms that haul goods to its distribution.

Wal-Mart's cost of sales as a percentage of net sales for fiscal year 1998 was 79·2% as compared to 79·6% for its 1997 fiscal year, as shown in Figure 4. The company attributes the 0·4% decrease to improvements in the mix of merchandise sold and to better inventory management[11]. This translates to a savings of some $456·56 million. Because of its success, Wal-Mart is beginning to introduce a demand-forecasting application, based on neural networking software and a 4000 processor parallel computer from NeoVista Solutions Inc. The application will look at individual items for individual stores to decide the seasonal sales profile. The NeoVista system keeps a year's worth of data on the sales of 100 000 products and predicts which items will be needed in each store. Wal-Mart also plans to expand its use of market-basket analysis. Data is collected on items that comprise a shopper's total purchase so that the company can analyse relationships and patterns in customer purchases. Currently some 3500 users make 10 000 database queries a day[12].

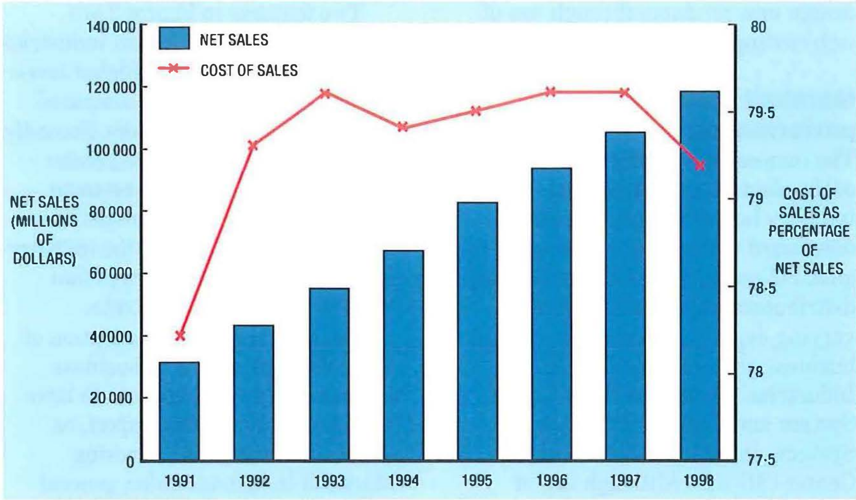


Figure 4 — Wall-Mart Industries

Automotive Network Exchange (ANX)

ANX represents an unprecedented collaboration by the Big Three automakers, Chrysler, Ford and General Motors and their top suppliers. ANX is an extranet that will establish a standard way for parts suppliers to communicate with manufacturers. The system is owned by the Automotive Industry Action Group, (AIAG). It has been under development since 1996, and began full operation at the end of 1998. ANX could ultimately link as many as 40 000 automakers, parts suppliers, dealerships, and financial-service companies to share everything from computer-aided design (CAD) files and groupware applications to e-mail and electronic data interchange (EDI) over a single IP network, replacing a complex and costly network system of multiple connections formerly existing in the automotive supply industry. Although private TCP/IP networks are currently well-established and reliable in the automotive industry, they fall short of industry-wide connectivity, which is the aim of ANX. Because ANX combines industry-wide connectivity with a protocol that is application-independent, it will encourage higher speed connections among trading partners.

ANX will interconnect automakers and trading partners using certified service providers (CSPs)—Internet service providers that meet ANX criteria for network performance, reliability, security, and administration—connected to a private ANX exchange point. An 'overseer' will continually measure the connected CSPs for packet loss, latency, link utilisation and other metrics, and will be equipped with disaster recovery plans[13]. Digital certificates are intended to help industry overcome the security concerns over the Internet. The electronic certificates used by ANX allow its various participants to authenticate one another's identities as they pass supply information back and forth over the extranet. Digital Signature Trust Company (DST), one of the premier providers of digital signature authentication and certification services, is the exclusive provider of digital certificate directory services to ANX.

ANX is using the Internet protocol security (IPSec) standard to secure its virtual private network (VPN). As a result, subscribers can conduct business-critical applications such as EDI via a ubiquitous and cost-effective alternative to private leased lines. Limited entry and exit points will let automakers securely share real-time data, such as changes to a production design, and confidential business information with suppliers. ANX is designed to be more reliable, more secure and faster than the public Internet[14]. By limiting access to the network and by using extra encryption, ANX hopes to move beyond security concerns that have hobbled the industry's move to increased reliance on computer networking in the past.

ANX, believes that the suppliers furthest downstream in the supply chain will stand to benefit the most from ANX[15]. In this respect, Chrysler expects to save millions by consolidating communications links onto ANX, through fewer T1 and satellite connections, and through the use of a standardised IP protocol in infrastructure support[16]. Similarly, LucasVarity, another participant in ANX, hopes to save at least 20% of the $6 million the company spends on private networks†. In addition, UT Automotive,

† LucasVarity is a $7 billion company based in London whose Detroit division supplies brake systems to automakers in 20 countries.

the $3 billion auto parts subsidiary of United Technologies, asserts that ANX relieves it financially from having to build a telecommunications infrastructure[17]. Table 1 compares the cost structure for traditional telecommunication connections with those that would be incurred under the ANX system. As shown, for a large automotive company with annual revenues of $8·3 billion, this translates to an annual cost of $151 532 under the current structure as compared to $54 680 under the ANX system[18]. For a medium company with annual revenues of some $1·7 billion, this amounts to an annual cost of some $78 000 using traditional connections as compared to some $25 850 under the ANX system.

Cost-savings through streamlining the supply-chain

Usually, when a supplier is connected to an original equipment manufacturer in order to access an application, a new link is established. Consequently, there are redundancy and higher costs, and many different skill sets must be maintained in order to handle every communications flavour that customers require[18]. Currently, 50–60% of a car's cost is due to the supply chain. If ANX is successful in cutting the time it takes for information to move through layers of the supply chain to 4 days from the 4–6 weeks that is normally spent, cost savings will be enormous. Since substantial cost is embedded in the supply chain, the challenge of redesign lies outside of each company rather than internally[19]. Overall, estimates of annual cost savings for the auto industry range from $1 billion to $2 billion, with savings per car estimated at $70 to $76[20]. Through ANX, auto suppliers can reduce the time it takes to turn around an order. The faster the parts come in, the faster the cars leave the assembly line. Ford, for example, hopes to compress some work order communication from three weeks to five minutes[21].

ANX has announced that it intends to link up with the European Network Exchange (ENX), although some European industry officials are doubtful if the ANX system can bridge the cultural, political and legal barriers separating the two markets[22]. As of March 1999, ANX had signed up only 34 customers.

Table 1 Telecommunications Connections Cost Comparison for Company (annual revenue of $8·billion)

	Existing Connections	
Purpose/Applications	**Types of Service**	**Annual Costs ($)**
X.400 Mail	19·2 kbit/s X.25 leased line	9 072
Chrysler SMART	9·6 kbit/s SNI leased line	10 272
Ford CAD file transfer	56 kbit/s IP VAN leased line	7 752
Ford DDL	9·6 kbit/s SNI leased line	1 272
Ford supplier CAD file transfer	56 kbit/s X.25 leased line	9 672
Navistar EDI on-line	19·2 kbit/s SNI leased line	5 052
Mack Trucks EDI on-line	56 kbit/s SNI leased line	8 232
Volvo EDI on-line	14·4 kbit/s SNI leased line	11 040
EDI VANs	9·6 kbit/s bisynch dial	53 808
Supplier dial-up access	9·6 kbit/s bisynch dial	35 360
Total		**151 532**
	New Connections	
	One-Time Cost	**Annual Cost**
Subscription assessment	2 400	
IPSec software	11 000	
Training	10 000	
Systems integration		
Total	**23 400**	
Digital certificates		250
Registration		1 600
Subscription		4 800
Connection 1·422 Mbit/s leased line		48 000
Total		**54 650**
Payback period (years)		**0·38**

AT&T hopes to sell leased lines and dial-up access to ANX members, although the cost for customers to buy services from ANX certified providers tends to be 4–5 times higher in price than non-ANX certified providers. It is still unclear if auto-industry customers will switch away from value-added networks, with the myriad of proprietary protocols running disparate financial, procurement and engineering design applications, and private lines to IP-based connections for sending mission-critical traffic.

Cisco Systems

Founded in 1984, Cisco Systems is a leading global provider of Internetworking solutions for corporate intranets and the Internet. It primarily manufactures multiprotocol computer communications equipment but also develops prepackaged network communications software. For the 1998 fiscal year, it reported revenues of some $8·46 billion, an increase of 31·3% over fiscal 1997 and a significant increase from the $700 million it posted in 1993. Sales to international customers accounted for some 40·9% of total sales in 1998[23]. Currently, about 60% of Cisco's business is enterprise, 30% is service providers and 10% is small/medium businesses. It should be noted, however, that much of Cisco's growth has come through a spate of acquisitions, including the $4 billion purchase of StrataCom in April 1997, which has boosted Cisco's share in the ATM switching market[24].

Cisco's web site accounts for 70% of its customer support activity and 15% of product orders. Cisco's web site was designed from the ground up and went on-line in early 1994 as a means of providing technical applications support to their customers, including software downloads[25].

Strategic objectives

When the Cisco Marketplace went online in March 1996, it offered customers the ability to configure their own products, price an order and route it within the company without having to depend on a salesperson. Cisco was reputed to be the first company to reach an annualised run-rate of $1 billion worth of products sold over the Internet by June 1997. According to the company, by October 1998, 69% of Cisco's orders were placed on-line which translate into $21 million per day, or $7·7 billion a year[26]. The overall objective has become thorough streamlining of their business, by taking advantage of the Internet's ubiquity, and the suitability of their product portfolio to on-line support and sales.

Cisco's e-commerce site, Cisco Connection Online (CCO), offers 24-hour global support for customers with technical and service questions. By most accounts, this site has been successful in increasing the company's accessibility to customers and in forging and building relationships[27]. Order status was one of the first processes to go on-line, prompted primarily by the frequent telephone call requests by customers. Some 70–80% of inquiries that used to be handled by Cisco's call centre are now answered electronically. The monthly log of incoming calls is increasing at 6·5% per month, but the CCO log-ins are increasing at a monthly rate of 10–15%[28].

Overall, extranet and Internet applications have allowed Cisco to cut costs and intensify its relationships with all its customers. They built up their on-line business by requiring their suppliers to link into Cisco's procurement system so that there would be a common network[28]. In Cisco's case, its extranet connections to its reseller and distributors are the heart of its e-commerce business, in contrast to other companies that usually begin with intranets to improve internal business efficiencies, and then gravitate towards IT solutions for improving customer relationships[25].

Benefits of e-commerce

Through its on-line operations, Cisco has been able to generate several economic advantages for itself as well as its customers:

- *Lower marketing and support costs*: Cisco's marketing and product-support costs are being held in check as a result of having all technical documents and marketing information sent to resellers and customers over the company's intranet for partners, also accessible through the web site. Cisco expects to save $280 million over the year, which in theory should translate to lower costs for customers†.
- *Increased accuracy of orders*: Cisco, by building into the web-based system an intelligent and sophisticated sales-configuration, has eased the burden on its order-entry staff who must typically correct 10% to 15% of all orders that come in by fax, about 350 000 in the last fiscal year. As a result, orders are correct the first time, reducing the delivery time for Cisco's customers by two to three days.
- *Reduced order processing time*: Cisco has reduced the time required for negotiating contracts, determining pricing, calculating lead times, checking on status and verifying shipment dates. With prices of Cisco products posted, customers can choose their prices and configure their products electronically with the Configuration Agent which walks them through the combination of components, and then gives them the price automatically[26]. Customers can pursue tracking orders, payments, reconciliation of returns against invoices and the status on their line of credit. The system now integrates with the largest customers' accounts-payable systems using traditional EDI networks, allowing all transactions to be reconciled without manual intervention.
- *Increased customer support*: Cisco's sales partners, resellers of the company's networking products, as well as customers who have been granted access to do business with the company on-line, can use the e-store 24 hours a day from anywhere in the world. This allows customers to get their own information at their own pace, rather than having to wait for a salesperson to return their calls or to track one down through the traditional mediums of e-mails or voice mails, which significantly benefits the large number of its international customers. Since customers have instantaneous status checks of their orders, they have been able to save time by not having to wait for a Cisco customer service representative to field their call. For example, in December 1996, the company logged some 50 000 status checks on orders in progress or service orders via the web site. In January 1997, 66 000 product order inquiries were handled similarly.

Table 2 E-Commerce Revenues for Cisco Systems

	1997	1998
First quarter	81	629
Second quarter	126	778
Third quarter	231	960
Fourth quarter	339	na

Cisco's performance

Cost of sales for the industry, generally has been increasing since 1994, reflecting the increasing competitiveness in the industry. Cisco's costs of sales at 33% compares favourable with the industry average. Sales, general and administrative costs for the industry ranged from 5% to 48% in 1995 and averaged 19·3% in 1995 which is an increase from 15·3% in 1994. Cisco's sales, general and administrative costs amounted to some 22% of sales. This has prompted many firms in the industry to search for ways to reduce these costs. Cisco's cost of sales amounted to about 22% of total revenues[29]. Cisco, however, has been able to maintain lower comparatively costs of sales than its competitors, as shown in Figures 5 and 6.

Several qualifications on Cisco's success are necessary. First, the actual extent of sales being completed on-line cannot be independently verified*. Initial research suggests that Cisco may be employing a much broader definition of on-line sales than is used in this study, where such sales would include any part of the process, including attaining pricing and product information on-line. Secondly, many companies' current procurement requirements and procedures prevent sales orders actually being placed through Cisco's on-line commerce agent. The growth of Internet commerce thus will depend on companies and institutions altering their procurement procedures and processes. Thirdly, it is highly unlikely that Cisco's commerce agent will completely replace the traditional sales team approach. The commerce agent is not designed to provide some of the services that the sales team does, such as availability of price discounts and after-sales support. Finally, it should be

† According to government data, cost of sales in the computer communications equipment sector range from 5% to 46% of sales. Cisco reports 22% cost of sales in its 10K filing. (Note: 10K is the name of a form that public corporations are required to complete by the Securities Exchange Commission.)

* For example, the network systems manager at Stanford University, in an interview with CTM, stated that the 27 April 1997 *Industry Week* article that cited Stanford University as having purchased a network systems on-line from Cisco was in error. All of the network systems managers interviewed by CTM reiterated that their company's internal procurement procedures currently prevent such on-line purchases.

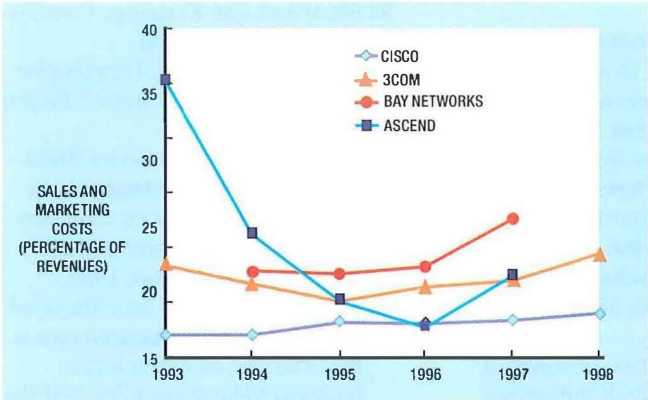


Figure 5—Sales and marketing costs as percentage of total revenues

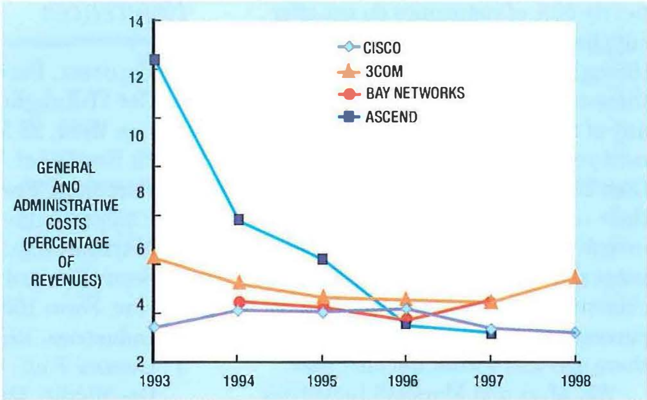


Figure 6—General and administrative costs as percentage of total revenues

noted that Cisco's products and customer bases are unusually well suited to Internet commerce. Cisco's customers are primarily network systems managers who are highly technical, who have the personal ability and organisational resources to conduct business in an on-line environment, and who are generally supportive of such an environment. Clearly, most business and consumer markets lack these particular conditions of success for Internet commerce.

IP Applications and Prospects for Adoption

The above discussion of Marshall Industries and other companies has demonstrated the ability of IP-based applications to lower costs, increase productivity, improve customer support and expand customer base. CTM's survey has found that over 70% rated improvement of the company's image, the use as a promotional tool and the low cost and rapid means of dissemination of information as important or very important reasons for the company to have a web site†. Similarly, some 50% of respondents see the Internet as an important or very important way to increase sales, provide better customer support and expand the geographical scope of the company. Juxtaposed against results from a similar survey in 1997, it is apparent that the main reasons for companies to have web sites have not changed significantly.

Marshall Industries has demonstrated how the use of IP-based applications has enabled it to achieve significant savings in sales, general and administrative costs. However, only 30% of the companies currently see this as an important or very important factor in having a web site. It may well be that the web-based experiences of Marshall Industries and other companies are leavening, albeit slowly, through the economy as there were 10% more companies in 1998 than in 1997 which see cost reduction as an important or very important reason for companies to have web sites.

CTM's survey has also found that some 80% of companies in 1997 and 1998 use their web sites to provide product information. The extent to which companies have adopted other types of IP-based applications is shown in Figure 7. Currently, only 34% of companies allow their customers to order products on-line while only 15% of companies offer their customers the ability to access technical support, make payments or attain real-time pricing and inventory information through their websites. Furthermore, less than 10% of companies currently provide their customers with real-time customer support, order status and checking and purchase history through their web-sites. These are services that Cisco and Marshall Industries currently offer their customers on-line. The growth potentials of these applications, however, are shown by the significant increases in the percentage of companies which intend to offer these applications to their customers by the end of 1999.

The lessons and economic benefits of supply chain integration experienced by Marshall, Wal-Mart and ANX may, however, take a longer time to disseminate through the economy, as shown in Figure 8. Currently,

† CTM's survey of some 70 firms is meant to gauge the extent of penetration of some of these IP-based applications. 51% of these firms are large firms with over $100 million in revenues in 1997 and 61% of the firms employed over 500 people.

Figure 7—Types of on-line services available to customers

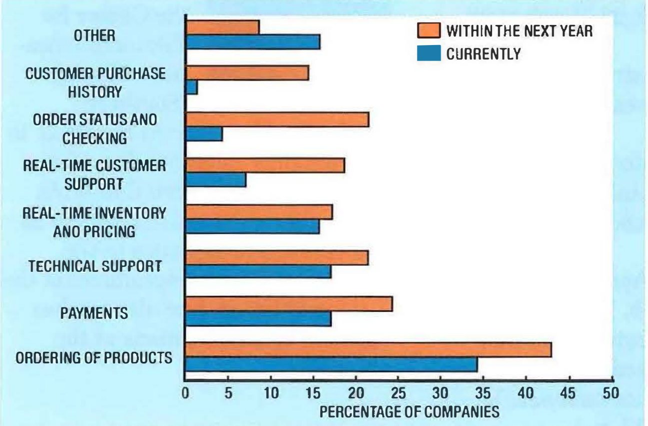
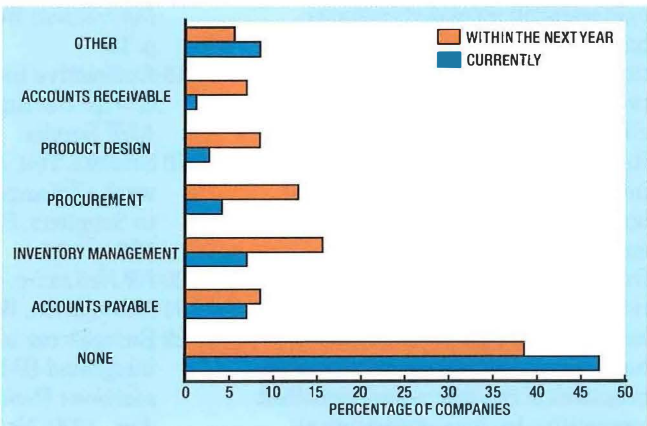


Figure 8—Types of on-line services available to suppliers



nearly 50% of companies do not offer suppliers any business applications through their web sites and most of these companies do not plan to offer any of these applications within the next year. However, although less than 10% of companies currently offer their suppliers account payment, inventory management or procurement services options on-line, Figure 8 shows that a dramatic increase in percentage of firms who will offer these services within the next year.

Wal-Mart and Marshall Industries, particularly, have been able to achieve significant business efficiencies across a wide range of corporate functions through the use of IP-based applications. CTM's survey reiterates two main points. First, a significant percentage of firms have yet to adopt any of these applications in their business processes. Secondly, among those firms that have implemented some of these IP-based applications, a large percentage of them have not been able to fully duplicate the level of success Marshall Industries has achieved. This, in part, may be due to the fact that most firms have not re-engineered their corporate processes first, as in the case of Marshall Industries to fully exploit the advantages of IP-based applications.

Conclusions

Several leading-edge companies have significantly utilised IP-based applications in their business processes and have demonstrated the existence of business efficiencies from the use of these applications. These efficiencies, attained by incorporating IP-based applications and services into the business model include: the lack of need to build proprietary telecommunications and data networks; cost-savings by transferring large data files between clients on the Internet; reduction in customer support personnel; better inventory management; increased customer-base; cost savings in business transactions; and increased employee productivity. However, these efficiencies have been achieved only through the re-engineering and integration of internal and external business processes. Although such re-engineering and integration are fraught with technical and organisational complexities, it may well reflect how successful global corporations have responded to the challenges and pressures of the new networked albeit competitive business environment.

References

1 Caldwell, Bruce. Senior Managers Get IT-Enlightenment. *Information Week,* 23 March 1998.

2 El Sawy, *et.al*, The Relententless Pursuit of 'Free.Perfect.Now': IT-Enabled Value Innovation at Marshall Industries.

3 Securities and Exchange Commission. Form 10K for Marshall Industries, 1998.

4 Gibbons Paul, Laura. Intranet in the Middle; Marshall Industries' GlobalNet Intranet; Company Operations. *PC Week*, **13**(38), p. 54.

5 Wilson, Tim. Wholesale Shifts to the Web. *TechWeb News*, 21 July 1998.

6 Power, Denise. Executives Taking New Look at Role of Technology Retail Systems '98 show Focus on the Shifting Strategies. *Daily News Record*, 4 March 1998.

7 Brown, Eryn; and Costa, Len A. America's Most Admired Companies. *Fortune*, 1 March 1999.

8 Chain Store Age. Data Warehousing: Turning Information into Action. Sept. 1996.

9 Crowley, Aileen. Delivering a Healthy Dose of Sales Data. *PC Week*, 24 Nov. 1997.

10 Wilder, Clinton. Chief of the year: Wal-Mart CIO Randy Mott innovates for his company's and customer's good. *Information Week*, 22 Dec. 1997.

11 Securities and Exchange Commission. Form 10K, Wal-Mart. Jan. 1998.

12 Foley, John. Squeezing more value from data. *Informationweek*, 9 Dec. 1996.

13 Messmer, Ellen. Automakers drive Private IP. *Networld*, 30 Sept. 1996.

14 *Automotive News*, 15 Sept1997.

15 Michel, Roberto. A Common Network. *Manufacturing Systems*, **15**(3), March 1997, p. 178.

16 *Information Week*, 1 Sept. 1997

17 Wilder, Clinton. Global Links. *Information Week*, 23 March 1998, p. 19.

18 Automotive Industry Action Group. The Business Case for ANX Service.

19 Stundza, Tom. Automotive Network eXchange Links Carmakers to Suppliers. *Purchasing*, No.2, **126**, p. 124.

20 *PR Newswire*, 9 April 1998.

21 *Information. Week*, 1 Sept. 1997.

22 Europe's car industry pushes for integrated IP-based net. *Communications Week International*, 18 Jan., 1999, No. 217, p. 1.

23 Securities and Exchange Commission. Cisco 10K, 1996.

24 Bartholomew, Doug. Trawling for $1 Billion. *Industry Week*, 21 April 1997.

25 Jordan, Peter. E-Business Building Blocks—Three steps to help you put your clients on their way to electronic commerce success. *VAR Business*, 23 Nov. 1998.

26 Tully, Shawn. How Cisco Mastered the Net; you think Amazon.com is big? The real money online is business-to-business sales, and the king of routers is writing the manual. *Fortune*, 17 Aug. 1998.

27 Gartner Group. Cisco Reaps Huge Benefits from Superb Web Site; Over $3 Billion in Revenue. http://www.gartner.com/webletter/cisco/article1.html

28 James, Kenneth. Lessons from the leader in Net trade. *Business Times*, p. 18.

29 Market Share Reporter, 1996.

Biographies

Francis Pereira
University of Southern California

Francis Pereira is a research fellow at the Center for Telecommunications Management, University of Southern California. He received his Ph.D. in Political Economy and Public Policy from the University of Southern California and teaches courses in economics and statistics. He has 6 years research experience in the telecommunications field.

Elizabeth Fife
University of Southern California

Elizabeth Fife is a research fellow at the Center for Telecommunications Management, University of Southern California. She received her Ph.D. in International Relations from the University of Southern California and has done extensive research on innovation and IT issues in the RACE and Esprit programmes of the European Union. She also teaches business communications at the university.

Edwin van Tricht and Cor Quist

Voice-Over-IP for Corporate Users

A solution in search of a problem?

What is the real value of voice over IP (VoIP) for corporate users? This is the key question of this paper. Five application areas of VoIP for companies are discussed: VoIP over the local area network (LAN), VoIP over the wide area network (WAN), VoIP for customer contact, VoIP for telecommuters and multimedia collaboration over IP. Competing technologies exist for all five areas. There is, however, a strong tendency towards IP-based solutions and VoIP does offer some unique advantages. The sometimes-poor voice quality of VoIP solutions is a serious barrier for its success. It is also important to note that current VoIP solutions do not yet offer all the well-known features of the classic PBX. It is expected that solutions will become available to solve these limitations.

Edwin van Tricht:
Synergetics IT Consultancy
e.vantricht@synergetics.nl
Cor Quist:
KPN Research
C.P.Quist@research.kpn.com

Introduction

According to some vendors, every company will have implemented voice-over-Internet-protocol (VoIP) solutions in a few years' time, regardless of the application or field of operation.

The following five application areas of VoIP for corporate users are distinguished: VoIP over the local area network (LAN), VoIP over the wide area network (WAN), VoIP for customer contact, VoIP for telecommuters and multimedia collaboration over IP.

For each application, the pros and cons of VoIP are discussed and compared to competing technologies. Based on this comparison, conclusions are drawn regarding the actual value of VoIP for the different application areas. The paper starts with a brief explanation of VoIP and of the quality it can offer.

What is VoIP?

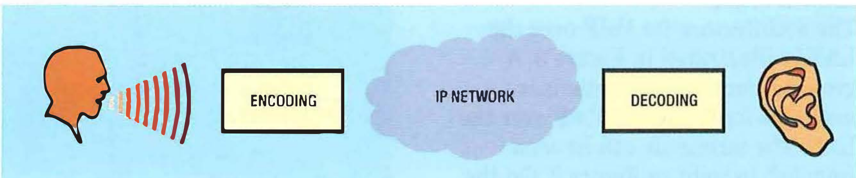
VoIP is a technology that transports voice over an IP network. With the help of special coding methods, voice is converted into IP packages, which are then transported over an IP network. The encoding software and hardware are known as *codecs*. The receiver decodes the packages and converts them back into voice. Figure 1 illustrates this process.

The great advantage of using IP for data or voice transmission is the fact that it can be transported over almost any other transmission protocol. For example, a user who is linked to an Ethernet LAN and someone working at home with a public switched telephone network (PSTN) line can communicate using the same IP application.

Another advantage is the compression that can be applied during the encoding. At the moment, the most popular codecs can compress a 64 kbit/s voice connection by a factor of four, making it a 16 kbit/s connection. Some of the more advanced codecs can even compress to a bit rate of about 5 kbit/s. By way of comparison, global system for mobile communications (GSM) encoding compresses a 64 kbit/s voice signal to 9·6 kbit/s. One should, however, bear in mind that compressed audio data need to be encapsulated in IP packets. This introduces a significant amount of overhead. We found that—depending on the parameters sample frequency, silence suppression and header compression—the overhead can easily result in the required bandwidth being doubled.

There are also disadvantages in using IP networks for real-time applications such as voice or video. IP was originally not developed for real-time applications, which means that the voice quality is lower than the

Figure 1—With the help of codecs, voice is converted into IP packages and transported over an IP network and subsequently decoded



PSTN quality. There is a risk of delays, echo and the loss of various parts of the conversation—if you are having a conversation over the Internet at a bad time, you could end up in a kind of walkie-talkie exchange. On the other hand, laboratory demonstrations have shown that voice quality is satisfactory when using VoIP on a well-dimensioned and well-managed IP network.

KPN Research performs numerous tests on the VoIP quality. Voice quality is not easy to measure. The relevant parameter is the perceived quality by the end-user, expressed as the mean opinion score (MOS). A live conversation has an MOS of 5, PSTN corresponds to MOS 4 and mobile telephony using GSM corresponds to MOS 3. KPN Research has developed a method to predict the MOS factor of voice in an IP network based on the technical parameters of the IP network. During a test carried out in 1998, the highest MOS score obtained was 3·5. So even under perfect conditions, the speech quality is still below PSTN quality. Five parameters determine the perceived audio quality: bandwidth, delay, jitter, echo and packet loss. We found that as the delay increases, the echo problem also increases. The delay of VoIP services will always be higher compared with the delay of the public switched circuits.

VoIP Applications within a Company

This paper discusses five applications of VoIP (see Figure 2):

- VoIP over the LAN,
- VoIP over the WAN,
- VoIP for contacts with customers,
- VoIP for telecommuters, and
- multimedia collaboration

These five application areas are discussed below. First, there is a short description of each application's architecture, then the existing products are described and finally, the competing technologies are highlighted.

VoIP over the LAN

Architecture

The architecture for VoIP over the LAN is illustrated in Figure 3. A growing number of terminals is available for the use of VoIP over the LAN. The terminals can be seen from left to right in Figure 3. On the left is a classic telephone connected directly to a router which has an integrated VoIP-gateway. Secondly, multimedia PCs can be used as telephones. These can include a regular desktop PC or a laptop communicating over a wireless LAN. In the middle you can see a special IP telephone connected directly to the LAN (in combination with an IP PBX). An IP PBX is a telephony server that provides already existing telephony features, such as completion of calls to busy subscriber (CCBS) and call diversion, as well as new computer telephony integration (CTI) features based on the integration between the telephony and the PC/data environment. For example, the number and name of an incoming call can be displayed on the PC screen. At the right side of Figure 3 you can find a solution for mobile telephony: wireless IP telephones that use wave LAN technology to communicate with other terminals.

This IP architecture consists of one single network for both data and voice, which means that only one infrastructure has to be managed. It is also possible to integrate the configuration management. When adding a new user, the new telephone number is created in much the same way as a new e-mail address is created. In this way, user data can be created and modified more efficiently, thereby reducing management costs.

Existing products

Every PBX vendor and IT vendor is currently active in this field. The products they sell can be divided into two categories: IP telephones and (small) gateways for normal telephones. IP telephones can be plugged directly into any Ethernet interface on the LAN. Selsius, meanwhile taken over by Cisco, was the first vendor of IP telephones and IP PBXs. (Small) IP gateways are integrated within the router and standard analogue telephones can be plugged into the router. At the moment, only a small number of services is supported. Frequently-used features such as call completion on busy subscriber are not supported at the moment.

Figure 2—Five applications areas of VoIP within a company

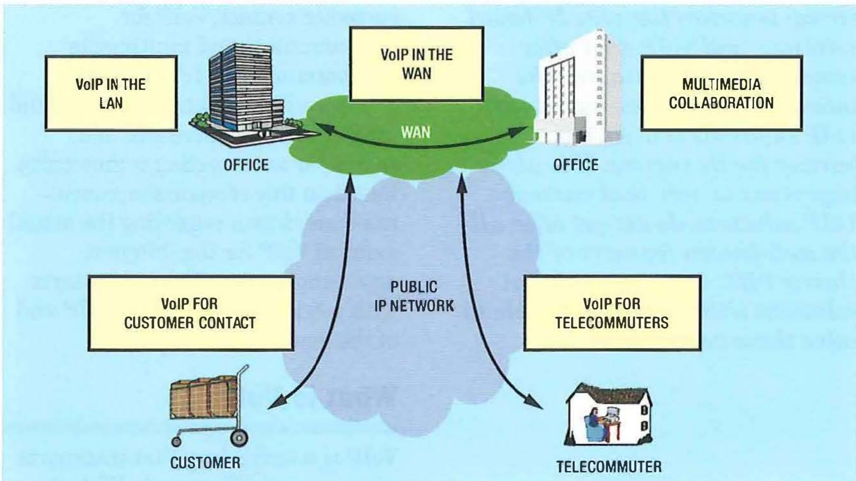
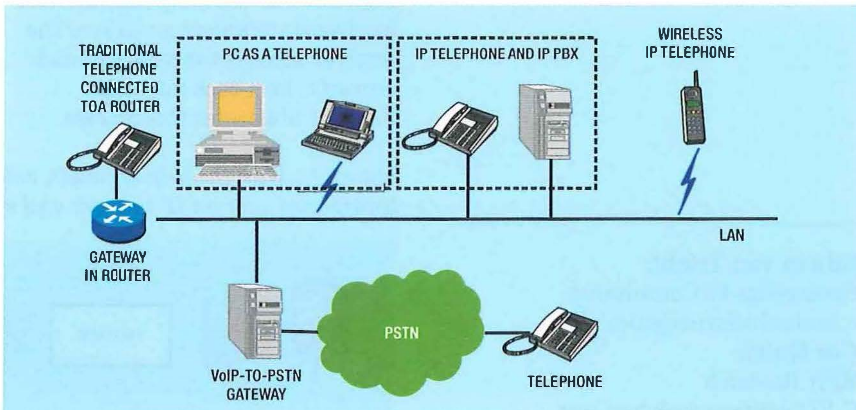


Figure 3—Architecture of a company network using VoIP over the LAN



Voice quality
The present IP installed base within companies is usually not sufficient to guarantee the quality of a telephone call. This means that a conversation might be temporarily interrupted if your neighbour starts up his word processor from the network. Quality guarantees for company networks can only be given if all equipment (for example, routers) supports the same protocol. Preparing your network for high-quality voice therefore often requires upgrading it entirely.

Competing technologies
IP telephony over the LAN will have to compete with existing solutions for business telephony. The PBX is still the preferred solution. The PC-based PBX is a fairly recent development. The various solutions are discussed and compared below.

The PBX is currently the most popular solution for business telephony. In a PBX, the switching matrix for setting up calls between the various extensions and the management software are completely integrated. The system is completely closed—it is not possible, for example, to develop a customer-specific application for a PBX. CTI has been developed to overcome this problem. CTI solutions link a server to a PBX via the hardware.

With a PC-based PBX (PCBX), the link between the telephony environment and the IT environment is not created using the hardware, but using the software. This PCBX is equipped with hardware cards that contain PBX functionality and can be controlled by the software in the PC. The telephone at the desktop is directly connected with the PCBX, while the PC is linked to the PCBX over the LAN. The software connection allows virtually every functionality desired by the user to be built. All of the existing PBX features can be realised, and new services such as video-conferencing and application sharing can be used. The drawback of a PCBX is its limited reliability, which will make it difficult to equal the 99.999% reliability of the traditional PBX.

Use of VoIP over the LAN
The best solution for a company depends greatly on its individual needs. If it wants to make calls using additional services such as voice mail, then a PBX is still the most suitable solution. This will be the case for the vast majority of businesses. Companies that place high demands on functionality will prefer a PCBX solution. By developing a customer-specific PCBX application every desired functionality could be achieved. For the next two or three years, VoIP over the LAN will only be used in niche markets, for example at locations that have a data network but no telephony infrastructure in place.

VoIP over the WAN

Architecture
VoIP also offers new ways of connecting PBXs over an IP network. This option uses VoIP-to-PSTN gateways. The gateway digitises and compresses the voice and splits the audio-bitstream into IP packets. It then sends the data over an IP network to another gateway, where the audio is decoded again. Figure 4 illustrates the architecture.

Integrating voice and data across the same network can offer cost savings—only one network has to be managed by the company, and only one service has to be supplied by the network provider.

Existing products
Current gateways were not specifically developed to link PBXs but to be used in the public network. Transmission of special PBX signals between two gateways is not supported by any supplier at present.

The various suppliers of VoIP-to-PSTN gateways can be divided into three groups: individual gateway suppliers, router suppliers who want to integrate the gateway into the router, and PBX suppliers who try to integrate the gateway into the PBX. The first group is by far the largest, yet this solution is less interesting for businesses, because a stand-alone system cannot be integrated in nor managed with the telephony environment or the data environment. Cisco is a clear example of a company that integrates the gateway into the router, whereas Nortel has chosen to integrate the gateway into the PBX. It is difficult to predict which of the two solutions will prevail in the long term.

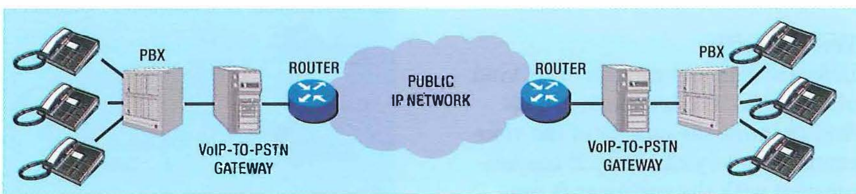
Voice quality
Relatively speaking, it is easier to give quality guarantees for VoIP over the WAN than for VoIP over the LAN because less equipment has to be upgraded or replaced. Only the actual connections with the WAN have to be modified. Depending on the type of network and the available bandwidth, the quality of the voice connection is roughly the same as GSM quality. This quality will increase in the coming years through the use of new protocols, which will make it possible to give quality guarantees.

Competing technologies
The idea behind VoIP over the WAN is that voice and data are integrated over one network. Older technologies that do the same thing include voice over frame relay and voice over ATM, both of which are further developed. For example, the transport of special PBX signalling between sites is already supported for ATM and frame relay, but not for IP.

Neither the IP protocol nor frame relay were originally developed to transport high-quality voice. ATM was. This means that it is easier to obtain a high-voice quality compared to frame relay or IP. Besides, IP always requires a transmission mechanism such as ATM or frame relay. This begs the question why a company would send its voice traffic over IP when it could just as well do so over the underlying transport layers like frame relay or ATM. IP does offer four important advantages:

- IP can be used as an overlay protocol in companies that use different transmission protocols.
- IP is carried to the desktop of the end user. This enables functional integration between the PC/data-environment and telephony applications (CTI).
- The data industry is currently focussed on IP, allocating large research and development budgets to IP-related developments. Furthermore, IP products are improving very rapidly.

Figure 4—Configuration for connecting PBXs over an IP network



- The advent of *all-IP operators* can result in cheaper IP transmission.

Use of VoIP over the WAN
VoIP over the WAN will be used by businesses that simultaneously use several transmission technologies in their network. The main offices can be linked to an ATM network, and the branch offices to a frame relay network. The IP protocol serves as a bridge between the various protocols. Businesses that implement it in the years ahead will have to accept that voice quality will be somewhat poor and that the various PBX features will not be transparent throughout the network. Businesses that use only one transmission technology in the network will probably not use VoIP but direct voice over frame relay or voice over ATM instead.

There will be a steady increase in the number of providers offering IP networks to link sites. The company will be supplied with an IP plug at each site. In this situation, companies may well choose a VoIP solution if they want voice-data integration.

VoIP for customer contact

Architecture
VoIP can be used as an extra way for businesses to come into contact with their customers, for example when a customer is surfing the company's web page and wants to ask a question about a certain product. Using VoIP, the customer can call the company without first having to break the connection. This method uses a single VoIP-to-PSTN gateway that converts the voice into IP packages and vice versa. The customer will need a multimedia PC with a sound card, a microphone and a loudspeaker. The VoIP-to-PSTN gateway can be managed either by the company itself or by a service provider. In the latter case, the gateway may be used by several businesses at the same time. This architecture is illustrated in Figure 5.

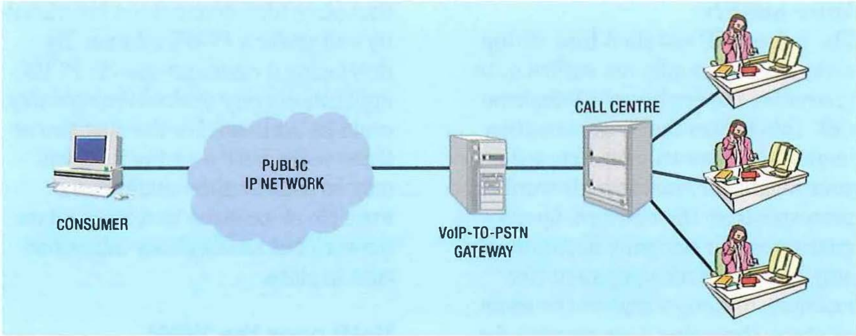


Figure 5 – VoIP architecture for customer contact

Existing products
VoIP-to-PSTN gateways are available as individual gateways. Call centre suppliers are also busy integrating these gateways into their equipment.

Voice quality
Before a company can offer this kind of service to its customers, it must first examine the quality of the voice connection very carefully. If communication with the customer is rendered impossible by a poor connection, this will certainly not be good for the relationship with the customer. The customer will probably blame the company rather than the IP network.

Competing technologies
There are two alternatives for this application:

- call the call centre by a normal telephone once the web surfing session is finished, or
- the use of the so-called *click-to-dial* button whereby a visitor of a web site clicks this button and the call centre call him/her back over the regular PSTN at a specified time.

The advantage of these alternatives is that the telephone call is a regular PSTN telephone call and the quality will be high. The advantages of the VoIP-application are that

- the customer can speak to the call centre agent directly;
- it offers the team-browse feature – the customer and the call centre agent can discuss a specific web page; and
- the customer is still on-line.

Use of VoIP for contact with customers
VoIP for contact with customers is an addition to the existing communication possibilities using web pages. Companies can experiment with this kind of application without big investments if they start with one central gateway managed by a service provider. If the response is satisfactory, the call centre at the customer location can be extended with a gateway.

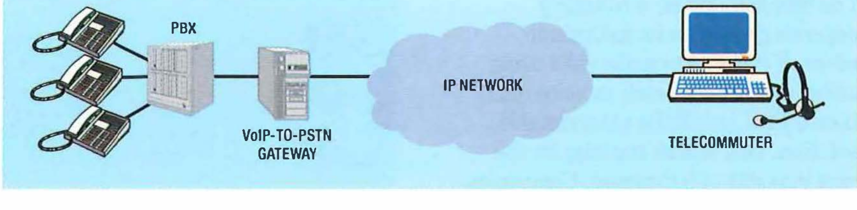
The potential problem of low voice quality might be avoided by giving the customer a choice: using VoIP which might result in a lower quality, or dialling back using the normal high-quality telephone. The latter option has the disadvantage that customers with only one PSTN telephone line have to disconnect from the Internet first.

VoIP for telecommuters

Architecture
Telecommuters often use a dial-in connection to connect to the company intranet. This same IP connection can be used for voice. When the telecommuter logs in to his/her intranet, a special software program on his/her PC logs on the VoIP-to-PSTN gateway. An incoming call, addressed to this telecommuter, is forwarded by the PBX to the VoIP-to-PSTN gateway. The gateway translates the voice into IP packets and sends them to the telecommuter's PC. The telecommuter can also use his/her PC at home to originate telephone calls. Even some PBX features can be supported. This architecture is illustrated in Figure 6.

Existing products
Virtually all gateway suppliers support telecommuter functionality.

Figure 6 – VoIP architecture for telecommuters



The architecture is very similar to the VoIP architecture used for customer contact. The only difference is that telecommuters have to log on to the gateway before they can use their PCs as telephones.

Voice quality

When telecommuters dial directly into the company's intranet, only privately owned network capacity is used and the voice quality approaches that of GSM telephones. However, sometimes public IP networks are used as dial-in infrastructure. In that case, the company does not have to invest in its own modem pool, but uses the modem pool of an Internet service provider. When using the public infrastructure of an Internet service provider, part of the network can no longer be controlled by the company. When using the Internet, voice quality becomes unpredictable.

Competing technologies

Traditional PBX vendors use proprietary telephones to support telecommuter functionality. These telephones are relatively expensive and require a second telephone line. The VoIP solution assumes that the telecommuter uses a multimedia PC which is permanently logged on to the network.

Another competing technology is GSM. Given the fast-growing penetration of mobile telephones, the need for fixed telephones is decreasing. The telecommuter's fixed telephone at the office is forwarded to his/her mobile telephone, without the need for extra gateway investments. When using extra fixed-mobile integration services, it is even possible to dial short numbers on the mobile telephone.

Use of VoIP for telecommuters

VoIP technology is available to automatically transfer calls made to fixed telephones at the office to the PC at home. A voice quality similar to GSM is certainly possible. This can be an interesting solution for telecommuters that have a multimedia PC at home, yet it might come too late, considering the rapid penetration of GSM.

VoIP and multimedia collaboration over IP

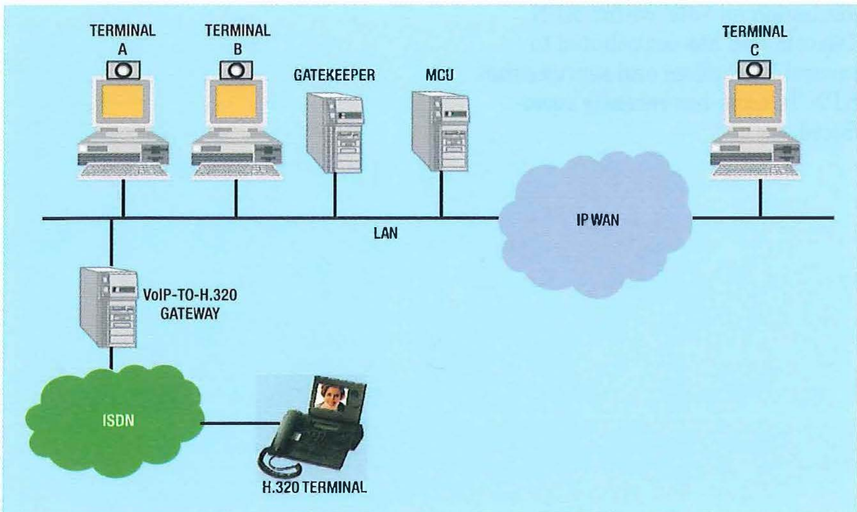
Architecture

It's a done deal: IP is the standard protocol for multimedia communication. Multimedia communication at the desktop will become the standard in the corporate environment. The three traditional key barriers are disappearing: bandwidth becomes available on the LAN and the WAN, desktop PCs are becoming increasingly powerful and the prices of equipment are dropping.

The ITU standard for VoIP, H.323, also supports application sharing and video communication.

A possible architecture is illustrated in Figure 7. We foresee that the desktop multimedia PC will become the de facto multimedia terminal. The gatekeeper is a necessary management tool to control the bandwidth usage and to translate the name of end users into IP addresses on the network. Bandwidth management is a crucial function as real-time traffic (voice, video) is very demanding. The MCU is a multipoint conferencing unit. This server enables multimedia conferences between three or more terminals. And finally, there is the H.323/H.320 gateway, which translates the information of a H.323 terminal to a H.320 (ISDN videoconferencing) terminal and vice versa. This gateway is only useful if you want to communicate with H.320 terminals. It can also be used to make a voice call from any H.323 client to a regular telephone on the PSTN.

Figure 7 – Architecture for multimedia collaboration over IP



Existing products

Multimedia clients come in several forms. An increasing number of H.323 multimedia terminals is available. If you have a multimedia PC, you can use Microsoft Netmeeting, a public domain H.323 client that supports multimedia communication. However, if you prefer to have a better video quality you need to buy special H.323 clients; for example, PictureTel's client. This client uses about 700 kbit/s for the video images. Multipoint conferencing units (for example, PictureTel, VideoServer) and H.323/H.320 gateways have been available since 1998.

Voice quality

The voice quality strongly depends on the characteristics of the network and on the multimedia client. The voice quality can be greatly improved by using a headset instead of a microphone and speakers. Multimedia communication, including video communication, is possible over a well-designed and well-managed LAN/WAN. Special attention needs to be paid to the audio and video synchronisation if video is used. Each client uses a different approach to synchronise the media streams.

Competing technologies

Multimedia collaboration can also be used over the integrated services digital network (ISDN) or asynchronous transfer mode (ATM). Videoconferencing systems for ISDN (H.320 terminals) and ATM are available. Multimedia over IP has one great advantage: IP can be used from the desktop, whereas ISDN and ATM are mainly used to connect videoconferencing studios. An alternative for multimedia collaboration is to use standard telephones for audio conferencing while using the desktop PC for a synchronised video communication and data-/application-sharing session. Although this is feasible from a technical point of view, it is not very popular in the marketplace.

Use of Multimedia collaboration over IP

Multimedia collaboration is particularly useful for multi-site companies with an intranet. It allows employees to participate in multimedia meetings from their desktop PC. Data/application sharing can enrich audio conferences. Video is optional and can be useful for certain types of meetings. Managed bandwidth in the LAN and WAN is required, especially when video is added to the conference. Employees need to have access to multimedia PCs.

Conclusion

This paper argues that using VoIP on the LAN will be a niche market for the forthcoming years. If companies adopt VoIP on the LAN, a gradual migration from their traditional PBX to VoIP on the LAN is expected.

Voice-data integration in the WAN will allow cost savings. This can be achieved using VoIP but also through voice over frame relay and voice over ATM technologies, which are slightly older and have made greater progress than VoIP. VoIP can, however, be a good solution if different tranportation technologies are used within the company network.

VoIP is the only solution for companies dealing with on-line customers who must be able to *push-to-talk* and have access to the Internet simultaneously. Companies should seriously consider the speech quality they can provide to their customers using this technology. Low quality will result in a bad image. To prevent this, companies can give customers two options: dial back using normal telephone lines or VoIP.

VoIP for telecommuters can be a cheap solution compared to proprietary PBX solutions. It enables telecommuters to automatically transfer calls made to fixed company telephones to their PCs at home. The PC can be used as a telephone that even supports some PBX features. However, the need for this solution is decreasing with the fast-growing penetration of mobile telephones.

IP is expected to become the platform for multimedia collaboration applications. Voice, video and data will become completely integrated. The available bandwidth on the public Internet will not be sufficient for these applications to grow substantially. Managed LANs and WANs certainly provide a solution here.

Biographies

Edwin van Tricht
Synergetics IT Consultancy

Edwin van Tricht has worked as a consultant within KPN Research for three years in the field of business communications. Within KPN Research, new upcoming technologies are evaluated, which enables KPN Telecom to offer new services. Edwin worked on several projects to evaluate new technologies, such as VoIP, PC-based switching and multimedia call centres. In the summer of 1999, he left KPN Research and is now working for Synergetics IT Consultancy.

Cor Quist
KPN Research

Cor Quist is a consultant within KPN Research in the field of multimedia and systems integration. He has been working on multimedia communication services, interactive multimedia applications, tele-education and voice-over-IP services. He has been following the developments in voice over IP ever since this evolving technology first presented itself. Over the past few years, he has been closely involved in the strategic discussion on VoIP within KPN Telecom and has contributed to several VoIP pilots and services that KPN Telecom has recently introduced.

Alwyn Lewis and Simon Broom

Feasibility of Non-Intrusive Monitoring of Voice Call Quality in Internet Telephony

In recent years, increasingly sophisticated ways have become available for accurately estimating customer opinions of PSTN call quality using non-intrusive objective measurements. This paper discusses the application of non-intrusive monitoring of end-to-end call quality to voice-over-IP (VoIP) networks. The problems and opportunities of a number of network scenarios are studied, including the impact of interconnections with the public switched telephone network. In particular, interactions between VoIP network elements, call signalling information, the voice signals and the choice of monitoring location are examined.

Introduction

The commercial success of voice-over-Internet-protocol (VoIP) telephony is likely to be strongly affected by customer opinions of connectivity, reliability and call quality that it provides, compared with the performance of the public switched telephone network (PSTN). In recent years, sophisticated ways have become available for accurately estimating customer opinions of PSTN call quality using in-service non-intrusive objective measurements[1]. In particular, such techniques are becoming popular for assessing the quality of service provided by interconnections with other network operators.

Simon Broom:
British Telecommunications plc
E-mail: simon.broom@bt.com
Alwyn Lewis:
British Telecommunications plc
E-mail: alwyn.lewis@bt.com

Internet transmission is a rapidly evolving subject and Internet telephony is an especially fluid part of this growing complexity. Thomas[2] provides a brief history of the Internet and of the semi-autonomous bodies that govern its technical standards and architectures. O'Neill, Sim and Rudkin[3–5] debate real-time Internet protocols, while Catchpole[6] explains current VoIP technology and describes a recent VoIP trial undertaken by British Telecommunications plc.

Despite recent progress, standards for VoIP transmission are not complete or widely agreed, because near-optimum mechanisms for handling real-time Internet traffic are still the subject of widespread research[3–5]. The authors believe that satisfactory methods for the automatic assessment of customer opinions of voice call quality will play a pivotal role in the testing and development of improved real-time mechanisms, and lead to more successful VoIP products and services.

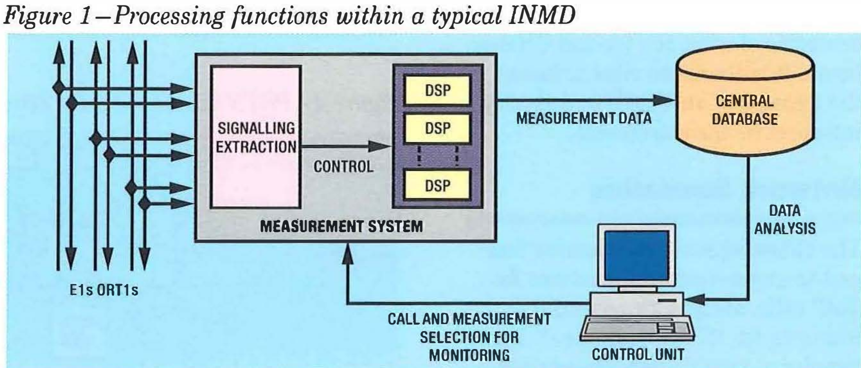
This paper discusses the fundamental issues involved in applying the established methods of in-service non-intrusive monitoring to assess the call quality of person-to-person VoIP connections, using IPv4 protocols. It does not extend to the assessment of conference calls nor are the findings based on the capabilities of any commercially-available non-intrusive monitoring equipment. Four potential network scenarios are studied and the impact of the choice of monitoring location is examined. Finally, questions are presented for further practical development of non-intrusive VoIP monitoring.

Non-Intrusive Monitoring

An in-service non-intrusive monitoring device (INMD) gathers information on network performance by making objective measurements on live traffic without disturbing that traffic or taking circuits out of service. The transmission impairments of a statistically-significant fraction of live traffic can then be assessed in ways that correlate well with customer opinions of voice call quality. If sufficient circuits are monitored for long enough, an accurate indication of average call quality across an entire network can be obtained. INMDs are currently under study by the ITU-T in Study Group 12[7–9] and a related ANSI standard was published in 1997[10].

An INMD is typically composed of three parts, as shown in Figure 1. The

Figure 1 – Processing functions within a typical INMD



measurement system uses digital signal processing (DSP) techniques to make objective measurements on the speech-band signals and extracts signalling information. A control system identifies which calls to monitor and a database system stores the measurement results for later analysis. Basic measurements include active speech level[11], psophometrically-weighted background-noise level and the echo path loss and echo path delay in both transmission directions. More advanced measurements could include estimates of temporal clipping and impulsive noise as well as the use of customer opinion models to predict the voice quality experienced by users.

Interpreting the Measurements

A common method of interpreting INMD data is to set acceptable threshold levels for the individual parameters measured on each network route. However, this can give misleading results because of subjective interactions between the parameters. For example, routes with high delay can provide acceptable voice quality to customers as long as there is also high echo path loss. Yet routes with levels of delay and echo that are each just within the appropriate threshold can give unacceptable voice call quality.

More appropriate interpretation can be achieved with a multi-parameter model of human auditory perception, producing a synthetic measure of customer opinion. One such example is the call clarity index model, developed at BT Laboratories[1].

The location of an INMD within a network significantly affects its measurement accuracy. For example, echo cancellers make it difficult to measure transmission delay and echo path loss through the cancelled path. If an INMD is located before the echo has been cancelled, then delay and echo path loss can be measured but these measurements are not representative of what the far-end listener hears. It is therefore vital to know the location of an INMD to correctly interpret its measurements.

Network Scenarios

The following sections examine four end-to-end network connections for VoIP calls, along with potential locations for INMD equipment. For simplicity, specific implementation details, such as the type of gateway, firewall, protocol or voice terminal in use, are ignored.

Pure IP connections

Figure 2 shows a network with purely Internet transmission of voice signals. At present, a lack of cohesive standards in terminal software limits this scenario to connections between cooperative users. A private intranet, with its potential for router management and control of software standards, is therefore a more plausible example of this scenario than a public Internet link, assuming a normal standard of connectivity and call quality. This scenario might use the public Internet for commercial applications as a 'fall-back' measure or to access mobile workers. As a result, this scenario spans a wide range of voice call quality.

IP-PSTN connections

The network scenario shown in Figure 3 involves a connection to a PSTN telephone, via a PSTN-to-IP gateway. The scenario is asymmetrical and either party could, in principle, originate or answer calls. However, the 'IP-originate' scenario is currently of greater interest owing to the large number of PSTN recipients. A commonly quoted application would be to bypass national or international tariff agreements, when the PSTN section would probably be a distant local network. The 'PSTN-originate' form might be used to access remote call-centres, when the PSTN section could include an international link.

Because one end of the IP connection occurs in the gateway, speech coding software and IP protocols are likely to conform to established standards. Currently, this scenario is unlikely to allow the transmission of PSTN signalling information over the IP section. However, an INMD located at the gateway could use signalling information to estimate the plausible range of some measured parameters, such as delay, for one end of the connection.

PSTN-IP-PSTN connections

The scenario shown in Figure 4 is essentially two examples of the previous scenario in tandem, sharing the same speech coding and IP protocol restrictions. It is commonly promoted for national and international 'tariff bypass' applications with two local PSTN sections. Some PSTN signalling information must travel across the IP section between the gateways, though this need not travel in the voice packets.

IP-PSTN-IP connections

Figure 5 shows an IP-to-IP voice call, connected through an intermediate PSTN network. This scenario is plausible as a VoIP 'overflow' provision in private intranet applications. The PSTN section might include an international link and the speech coding and IP protocols are likely to conform to established standards. In this scenario, the signalling informa-

Figure 2—Pure IP network scenario for VoIP calls

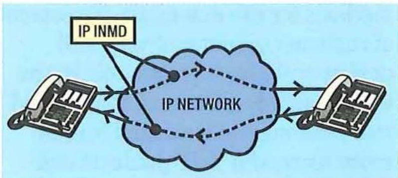


Figure 3—IP-PSTN network scenario

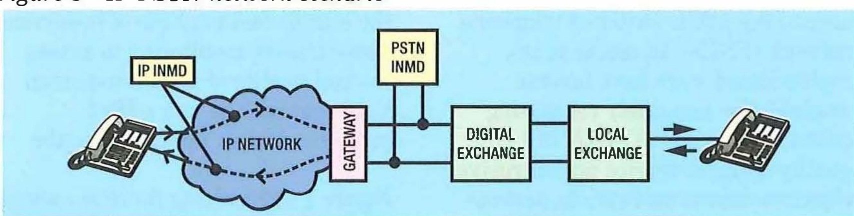
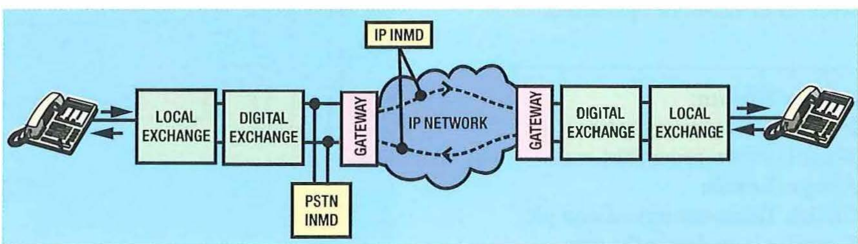


Figure 4—PSTN via IP to PSTN network scenario



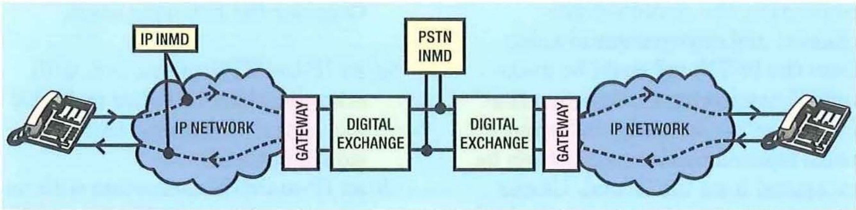


Figure 5—IP via PSTN to IP network scenario

tion that travels over either the IP or the PSTN section is unlikely to give much useful information about the expected parameters of either end of the connection.

Parameters of VoIP Telephony

Table 1 shows the relative perceptual impact of physical factors in telephony. Important characteristics of telephones, in terms of voice quality, are the send and receive loudness ratings[12]. These determine the magnitude of the speech signal reaching the listener's ear. Large deviations from a preferred listening level reduce subjective voice quality and can degrade the objective performance of speech coders. Audible echo signals can seriously interfere with fluent and natural conversation[13]. Noise audible between speech utterances, owing to bit errors or quantisation effects, is generally negligible. In particular, environmental (ambient) noise can degrade the quality and intelligibility of received speech. Such noise problems can be complex, depending on the characteristics of the noise and of the microphones, loudspeakers or handsets in use.

VoIP connections share many of the features of mobile, long-distance and international calls. Delay, echo control, speech coding and environmental noise are important design issues. Interconnection of VoIP telephones with PSTN telephones, in fixed or mobile form, creates the possibility of degradations due to cascade connections of multiple echo cancellers and speech codecs.

Voice Quality Issues

VoIP telephony brings potential impairments and uncertainties that are not present in the PSTN, because IP networks create circuits with mutually interdependent performance. The fundamental processes of VoIP transmission—coding, network transport, buffering and decoding—increase delay and introduce the possibility of packet loss, in ways that depend on network configuration and traffic loading. The variable nature of this extra delay and the unpredictability of packet loss mechanisms can create discontinuous speech signals, with gaps or deletions. Such mutilations not only degrade the call quality experienced by the users, but also upset conventional non-intrusive measurement algorithms, potentially leading to inaccurate or misleading indications of voice quality.

Delay and echo

In the absence of echo, long delay creates difficulty in speech communication. Increased delay also intensifies the subjective impact of echo signals[13]. This creates a need for echo control, especially on PSTN interconnections where 2-to-4-wire converters can be a major source of echo. Acoustic echo can also be troublesome, depending on the terminal and handset design. In the PSTN, handset design is well controlled but in some VoIP applications this aspect of terminal design is, at present, poorly specified. For example, acoustic echo can easily be a problem when the VoIP terminal is a personal computer (PC) using a microphone and loudspeakers of variable sensitivity.

Conventional echo cancellers are unable to cope with the effects of packet loss and variable delay in the echo path. Echo cancellation is most successfully achieved in the VoIP terminal or close to the echo source. Echo audibility is greater when listening with a handset than through loudspeakers, so that user opinions may differ at either end of a connection.

Coding

On heavily-loaded IP networks, speech-coding techniques can improve some aspects of VoIP call quality by reducing the bit-rate capacity needed for transmission. However, speech coding introduces its own degradations which are highly dependent on the amount of compression and the complexity of the coding algorithm used[14]. Additionally, if packets are lost when high compression coding is used then a larger duration of speech is disturbed. The impact on voice quality depends on the coding technique and its ability to reconstruct an estimate of the missing speech.

Some VoIP speech codecs use voice activity detection to avoid transmitting packets when the talker is silent. In this case, so-called *comfort noise* is sometimes synthesised at the decoder, to approximately match the characteristics of the ambient noise from the talker's environment.

Non-Intrusive Monitoring of VoIP Telephony

Several factors in VoIP connections, therefore, increase the difficulty of making accurate non-intrusive measurements, and INMD equipment designed for PSTN use may not give reliable assessments of call quality in VoIP connections. Good predictions of user satisfaction depend on using such measurements to accurately estimate the acoustic experience of the listener.

Table 1 Relative Effect of Physical Factors on Aspects of the Perception of Quality (based on ETSI Technical Report ETR-250, July 1996)

Physical Factors	Relative Effect on Perception			
	Overall Quality	Loudness	Intelligibility	Naturalness
Loss	High	High	High	Low
Bandwidth	High	Medium	Medium	High
Noise	Medium		Medium	Low
Distortion	Medium		Medium	Medium
Sidetone	Low	Low		Medium
Echo (talker)	High		Low	Medium
Delay	Medium			Low

The potential measurement problems, in pure IP or hybrid IP-PSTN networks, are examined in the following sub-sections for a mid-network monitoring location. It should be noted that IP networks also offer the potential for improved kinds of non-intrusive measurement, by exchanging additional data between the terminals along with the speech signals. However, this would require changes in both router and terminal software.

Speech level

Loudness is a key factor affecting ease of communication. Active speech level[11] is the most useful measure of speech signal magnitude in telecommunications, because it essentially removes the impact of the silent periods between utterances. Knowledge of speech levels at a mid-network point, in either IP or PSTN form, does not mean the acoustic levels at the listeners' ears can necessarily be predicted. The gain or loss in gateways, local PSTN links and that due to the electroacoustic sensitivities of the terminals must be known. In VoIP applications that use PC terminals, the electroacoustic sensitivities are usually unspecified and highly variable.

Speech coding

Measuring speech levels means decoding the speech, which demands knowledge about the coding technique for that call and access to object code for all possible decoders. Coding information may or may not be available through inspection of the packets mid-network. In typical VoIP applications, the type of coder used can be determined from the packet header. The subjective performance of certain speech coders has been assessed and can, in principle, be represented by an equipment impairment factor[15].

In some scenarios, speech may be subjected to coding more than once. For instance, in the IP-PSTN-IP scenario speech passes through two gateways and so will be encoded and decoded twice. Such cases are difficult to either detect or measure non-intrusively but are likely to degrade voice quality.

Background noise

Measurement of noise on a VoIP connection is not straightforward when voice activity detection is used to suppress packet transmission during silent periods. In a PSTN–IP connection, the combined idle channel and environmental noise from the PSTN end could be measured if monitoring is located in the PSTN section, but only the comfort noise injected by the gateway can be measured from the IP end. Unless the environmental noise level is high and misclassified as speech, the only noise present in the IP mid-network signals will be that embedded in the speech. An alternative approach would be to use signal-processing methods to estimate the level of noise within active speech[16].

INMDs located in an IP network require knowledge not only of the speech coding algorithms but also of any silence-suppression and comfort-noise algorithms. To provide accurate predictions of voice quality, INMDs for VoIP must be able to duplicate the terminal decoding process.

Packet loss

An accurate way of detecting packet loss is essential for good non-intrusive measurements of VoIP call quality. In typical VoIP applications, packets will contain sequence numbers or timestamps.

The subjective effect of packet loss depends on several factors, including the distribution of packet loss—whether random or bursty—and the speech codec in use. Burst errors are more common and also have a more severe impact on voice quality. Different speech coding algorithms exhibit markedly different robustness to lost packets, depending on the source model employed. If the type of codec is known, it may be possible to predict the impact of lost packets on the voice quality of the codec.

A further important factor is the length of the terminal buffer, since this length limits the acceptable variation in packet transmission delay before degradation occurs. Overflow or underflow of this buffer results in loss or insertion of frames, potentially mutilating the speech signal.

Echo and delay

Non-intrusive measurements of echo path loss and delay are interdependent, because some echo signal must be detected to measure the associated delay. Simultaneous access to the speech signals in both transmission directions is needed. The ease and accuracy of the measurements depend on the monitoring location and, critically, on the design of the VoIP speech codec.

Consider the following cases:

(*a*) an IP-to-PSTN connection, with echo cancellers at either end, that has a large end-to-end transmission delay; and

(*b*) an IP-to-PSTN connection with no echo cancellers, audible echo and medium end-to-end transmission delay.

In case (*a*), little or no echo signal is detectable in either direction with a mid-network monitoring position, making both echo path delays hard to measure. The sum of the two path delays—the end-to-end delay—cannot be reliably determined even though this might be large enough to make conversation difficult.

In case (*b*), echo signals are notionally present in both directions and no measurement problems should occur. However, if the VoIP codec uses silence suppression, it may still be difficult to make useful measurements of echo path loss and delay. Voice activity detection algorithms are designed to classify echo as 'not speech' and block its transmission.

These difficulties may be less significant than they at first seem. If echo signals are suppressed they will probably not be audible, except during changes of talker. The silence-suppression feature has the same action as an echo suppressor. If echo is large enough to be transmitted, it is likely to affect the voice quality but it can readily be measured mid-network. In this way, echo is reliably measured only when it can seriously degrade voice quality.

Opinion models

Individual measurements of speech and noise level, echo path loss and delay are useful indications of voice quality, but they do not account for the psycho-acoustic interactions between parameters that occur in the human perception of voice call quality. A multi-parameter model can calculate these effects, by combining measurements made on a call with information about human perception, producing a single figure estimate of the voice call quality perceived by users at each end of the connection. Such an index can be meaningfully averaged over many calls, unlike an average of any of the individual measured parameters which will not give useful information about voice quality.

Examples of customer opinion models include the call clarity index[1] for non-intrusive measurements and the E-Model for network planning[17].

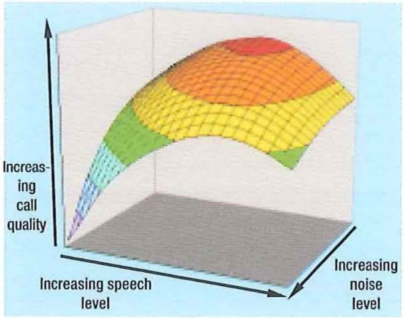


Figure 6—An example of an opinion model prediction of call quality

Figure 6 shows the mean call quality predicted by a model using only two parameters. In practice, many parameters are required and the surface shown in Figure 6 becomes multi-dimensional.

Future Directions

Standards and software for IP transmission of real-time signals are rapidly evolving and some of the technical claims for new regimes may be influenced by commercial factors. The most radical new arrangements require major changes to the software, and potentially hardware, in IP routers, servers, clients and terminals[18]. The technical options for real-time IP services include new protocols, such as Int-Serv and Diff-Serv[19]. New means of transferring PSTN signalling information through IP connections have also been proposed.

VoIP technology creates both problems and opportunities for in-service, non-intrusive measurements of voice call quality. Current customer opinion models may give inaccurate results, because IP transmission features network topology variations not found in the PSTN. Additionally, IP transmission creates impairments whose impact on call quality is not yet fully understood. The questions that arise include:

- How serious is the impact of signal gaps and deletions on speech signal measurements, and can the measurement algorithms be modified to reduce or eliminate this impact?
- How serious is the impact of location on INMD operation for different VoIP network scenarios?
- How will PSTN signalling information be carried through IP gateways?
- Can VoIP terminal configuration (such as buffer size and speech coding scheme) be reliably detected mid-network?
- Can new measurements, such as an estimate of speech coding distortion, be usefully made?
- Can the perceptual models be modified to incorporate the subjective effects of delay jitter and speech coding?

Conclusions

This paper has examined the fundamental architectural and signal processing issues in four network scenarios, spanning pure IP and hybrid IP-PSTN connections with a range of monitoring locations. Direct application of existing in-service, non-intrusive monitoring methods, as used in the PSTN, is not a reliable way of assessing voice-over-IP call quality if delay jitter and packet loss are significant. These degradations are not incorporated in the call quality measurements or models of current INMD equipment.

With appropriate development, it seems likely that useful measurements and models could be devised to deal with these degradations, which mainly affect echo and delay measurements. It is also possible that the degradations of alternative speech coding methods could be classified. Furthermore, IP networks offer potential advantages for improved non-intrusive monitoring techniques, requiring changes in router and terminal software.

The voice quality provided by evolving VoIP technology has been widely recognised as a critical factor in its success. The authors believe that making statistically-reliable non-intrusive call quality assessments, on a large number of VoIP calls under real-life conditions, will play an important role in improving technical standards and in building customer confidence in VoIP services.

Acknowledgments

Thanks go to Richard Reynolds, Phil March, Tim Hunt, Phil Sheppard, James Walker and Adrian Slim of BT Laboratories, for their expertise and assistance.

References

1 Broom, S.; Coackley, P.; and Sheppard, P. J. Getting the Message Loud and Clear—Quantifying Call Clarity. *Br. Telecommun. Eng.*, April 1998, **17**, pp. 66–72.

2 Thomas, J. Governing the Net. *PC-Plus* magazine, May 1999, pp. 35–38, Future Publishing Ltd. (see also http://www.pcplus.co.uk/).

3 O'Neill, A. *et al.* An Overview of Internet Protocols. *BT Technol. J.*, Jan. 1998, **16**(1), pp. 126–139.

4 Sim, S. P.; and Rudkin, S. The Internet—Past, Present and Future. *ibid.*, April 1997, **15**(2), pp. 11–23.

5 Rudkin, S; Grace, A; and Whybray, M. W. Real-Time Applications on the Internet. *ibid.*, April 1997, **15**(2), pp. 209–225.

6 Catchpole, A: Voice-Data Convergence and the Corporate Voice-Over-IP Trial. *Br. Telecomm. Eng.*, Jan. 1999, **17**, pp. 218–224.

7 ITU-T Study Group 12, Q15. In-Service Non-Intrusive Assessment of Voiceband Channel Transmission Performance.

8 ITU-T Recommendation P.561. In-Service Non-Intrusive Measurement Devices—Voice Service Measurements.

9 ITU Study Group 12, draft Recommendation P.5ai. Analysis and Interpretation of INMD Voice Service Measurements.

10 ANSI standard T1.255-1997. Telecommunications— In-Service, Non-Intrusive Measurement Device (INMD)—Methodology for Applying INMD Measurements to Customer Opinion Models.

11 ITU-T Recommendation P.56. Objective Measurement of Active Speech Level. March 1993.

12 ITU-T Recommendation P.79. Calculation of Loudness Ratings for Telephone Sets. March 1993.

13 Richards, D. L. Telecommunication by Speech. Ch. 4.4.2. Butterworths, 1973.

14 Wong, W. T. K., *et al.* Low Rate Speech Coding for Telecommunications. In Westall, Johnston, and Lewis (eds.): Speech Technology for Telecommunications. Chapman & Hall, London, 1997.

15 ITU-T Recommendation G.113. Transmission Impairments. Feb. 1996.

16 Dhanjal, S. Noise Cancellation Techniques in Speech Signals. Proc. Canadian Conf. on Elec. and Computer Eng. (IEEE), Sept. 1993, **1**, pp. 281–4.

17 ITU-T Recommendation G.107. The E-Model , A Computational Model for Use in Transmission Planning. Dec. 1998.

18 Prevost, S.; and Dann, T. IPv6—The Next Generation Internet. *Br. Telecomm. Eng.*, Jan 1999, **17**, pp. 225–231.

19 Ovenden, F. Taming the Internet. *Telecommunications*, Jan. 1999, **33**(1), pp. 65–71.

Biographies

Alwyn Lewis
British Telecommunications plc

Alwyn Lewis is an engineering advisor in Advanced Communications Research at BT Laboratories, with a B.A. in Electrical Sciences from Cambridge University and an M.Sc. in Opto-electronics from Essex University. From 1978 to 1989, he helped develop digital signal processing hardware and software for end-to-end telephony measurements and for an experimental duplex hands-free telephone, using a custom VLSI echo-cancelling chip-set. In 1992, he lead the Speech Coding team in the final stages of the CallMinder™ speech coder development and in the development and testing of DSP software for the Skyphone™ aeronautical facsimile and data service. His interests include speech coding, echo cancellation, beam-steering microphones, voice-over-IP technology, video telepresence systems and the social impact, both historical and current, of electro-technology. He is a chartered engineer and a Member of the IEE and IEEE.

Simon Broom
British Telecommunications plc

Simon Broom graduated from the University of York in 1995 where he gained an M.Eng. in Electronic Systems Engineering. He works for BT within the Advanced Communications Engineering Department where his work centres on voice transmission performance. This includes research into new algorithms for the quantification of transmission impairments and the prediction of call clarity using subjective models. His expertise lies in the application of these tools to non-intrusive network voice-quality assessment and recent work has focused on the application of such tools to the next generation of voice networks. He is an associate member of the IEE, a member of the IBTE and is currently studying for an M.Sc. in Telecommunications.

Wolfgang Groenen

Towards the Needs of Third-Generation Systems: Planning the Evolution of the GSM Platform

This paper gives an overview on the present situation and the future development of cellular mobile services based on the global system for mobile communications (GSM) in Europe and worldwide. It describes market trends and expected changes.

The future evolution as described in Phase 2 and 2+ of the standard will open new market segments and new customer groups for the GSM system.

As GSM continues its rapid growth, work on next-system generations, like the universal mobile telecommunications system (UMTS), point the way into the next millennium. The convergence of networks, applications and services will offer unprecedented mobile and personal telecommunication services to customers worldwide.

Dipl. Ing. Wolfgang Groenen:
Lucent Technologies Network Systems GmbH
P.O. Box 3538
D-90327 Nuremberg, Germany
Tel: +49 911 526 2474
Fax: +49 911 526 3500
E-mail: wgroenen@lucent.com

Cellular Markets

The success of the global system for mobile communications (GSM) can best be illustrated by the numbers of operators, subscribers, countries and networks on air (Figure 1 and Table 1).

Originally intended to be a European system, GSM is on the way to a world standard. In more than 120 countries GSM is in operation, is being introduced or has been chosen as the standard.

Based on that success, the market forecasts show a rapid growth up to 230 million GSM subscribers until the end of the century. Worldwide subscriber figures are forecast to reach more than 500 million cellular subscribers at that time.

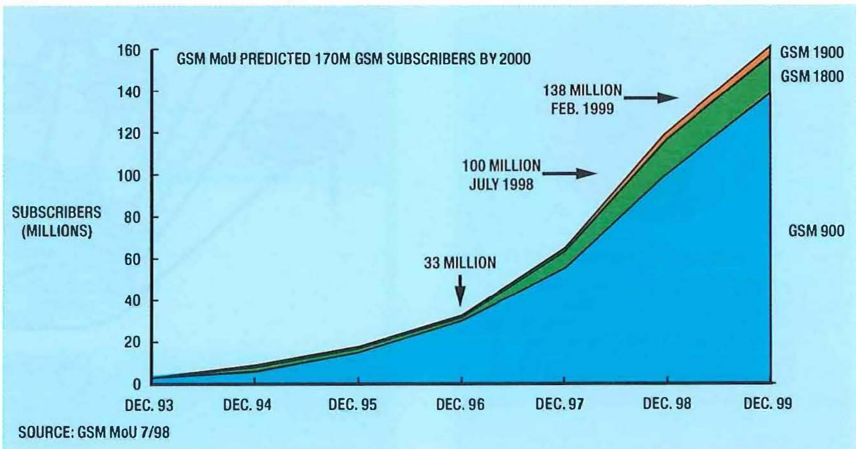
Considering the distribution of cellular subscribers by geographic region, a shift in importance can be

Table 1 GSM Platform—Statistics of Success

	Users	GSM MoU Members	Countries	Networks on Air
1991	Trials only	43	23	0
1995	12·5 million	156	86	117
1997	70 million	256	110	239
1999	150 million	323	129	320
2000 forecast	>230 million	>350	140–160	>350

Source: GSM MoU Association

Figure 1 – GSM subscriber forecast



expected as the European and Asia/Pacific markets will gain compared to North America (Figures 2 and 3).

Future System Evolution

The evolution of the GSM standard is targeted on enhancements of

- service capability,
- system performance, and
- application area.

A phased approach (known as *Phases 1, 2, 2+* of the European Telecommunications Standards Institute (ETSI) standard) retains the overall compatibility of GSM.

Figure 2 – Distribution of world cellular subscribers by regions

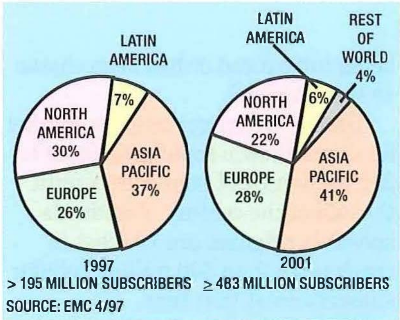
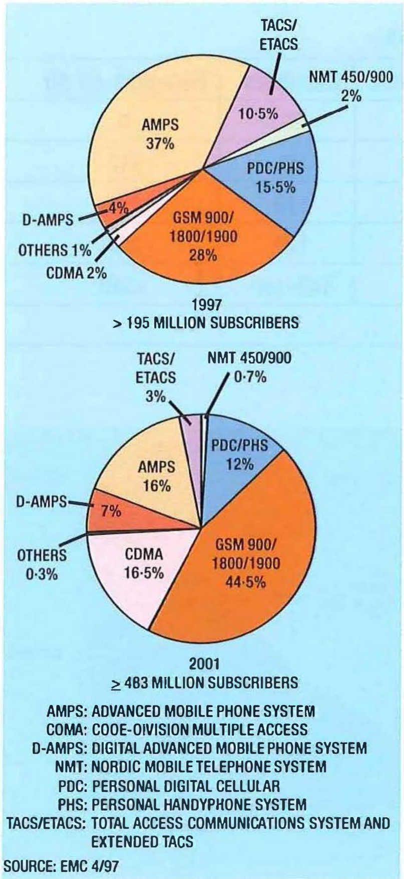


Figure 3 – Distribution of world cellular subscribers by technology



The enhancements in Phase 2+ (see Figure 4) include features which will open new market segments and new customer groups, for example,

- interworking between GSM and the digital cellular system (DCS),
- interworking between GSM and the digital enhanced cordless telecommunications (DECT) system,
- wireless local loop,
- railway applications, and
- trunking, paging, and messaging services.

While personal communications network (PCN) services in countries using the GSM standard will be based on GSM 1800 as a derivative of GSM 900, the next generation of personal communication service (PCS) networks in USA will be based on both time-division multiple access (TDMA) and code-division multiple access (CDMA) standards. GSM 1900, a derivative of the GSM 900/GSM 1800 standards has a strong position in that market.

An outlook on the evolution of GSM with

- new market segments,
- new applications,
- converging of services, and
- converging of networks

GSM Evolution—Phase 2+

- Intelligent network (IN) features (CAMEL—customised applications for mobile network enhanced logic)
- General packet radio system (GPRS) (packet switched data)
- 14·4 kbit/s (single time-slot)
- High-speed circuit switched data (HSCSD)
- Enhanced full-rate speech codec, multi- or variable-rate speech codec
- New supplementary services
- Dual 900/1800 operation
- DECT/GSM interworking; wireless local loop (WLL)
- Optimal routing
- UIC features (railway networks)

Figure 4

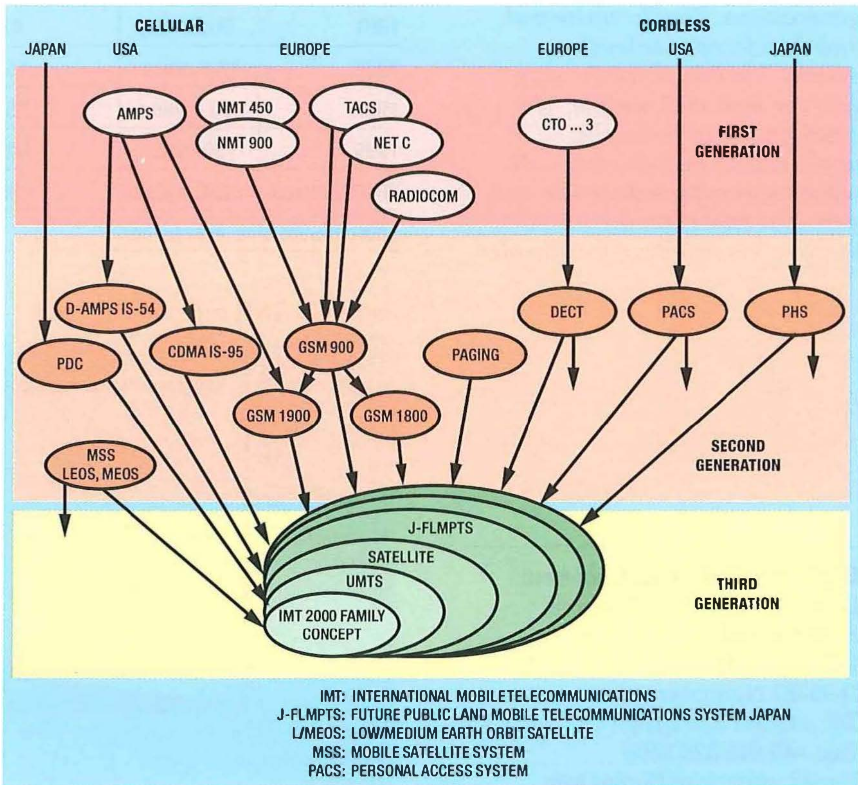
shows that GSM will retain its position as the world's leading digital cellular technology.

Next Generation Systems

Possible migration paths by which different mobile standards and systems could advance from the second to a third generation are shown in Figure 5.

The work towards third-generation systems in Europe, that is, the universal mobile telecommunications system (UMTS) is driven by research

Figure 5 – Evolution to third-generation mobile systems



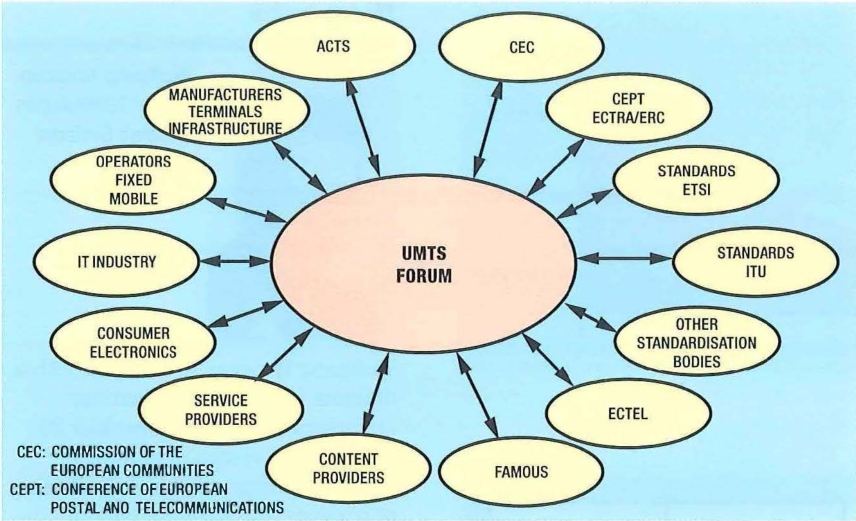


Figure 6—UMTS Forum—spheres of influence

activities in the framework of the Advaced Communications Technologies and Services (ACTS) programme and by standardisation activities in the ETSI/Special Mobile Group.

In parallel, the International Telecommunications Union (ITU) is working on a global base with its IMT 2000 concept. With the Third Generation Partnership Project (3GPP) process in place now a global standardisation and harmonisation procedure pushes the work forward.

The European activities on UMTS have recently been refocused by the report of the "UMTS Task Force" which worked under a mandate of the European Parliament and Council[1].

One of the recommendations was the establishment of the "UMTS Forum" as a central body charged with the elaboration of European policy towards the implementation of UMTS.

The Forum will contribute to better cooperation between all parties involved (Figure 6), taking care of regulatory aspects, frequency needs, market aspects and technology impacts. The tasks include steering the necessary evolutionary steps as well as the necessary common view on fixed and mobile networks and services (Figure 7 and Figure 8). Time schedules up to now show first installations of basic UMTS for broadband needs in 2002 and reach beyond the year 2005/2008 for full UMTS operation[2].

True personal mobility in future networks can only be reached by obeying some key trends; for example,

- an evolutionary approach from second to third generation allowing a smooth introduction,
- the path from standalone to integrated solutions,
- a migration of fixed and wireless networks in a harmonised way leading to common platforms (Figure 9)[3].

Figure 7—Positioning of UMTS

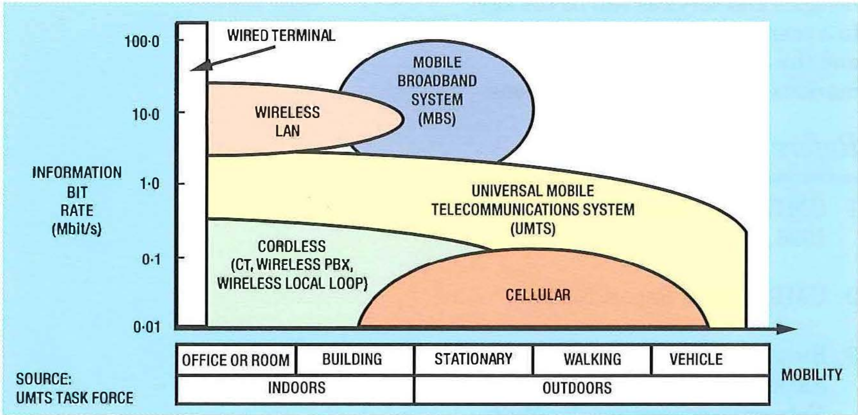


Figure 9—Multi-standard architecture

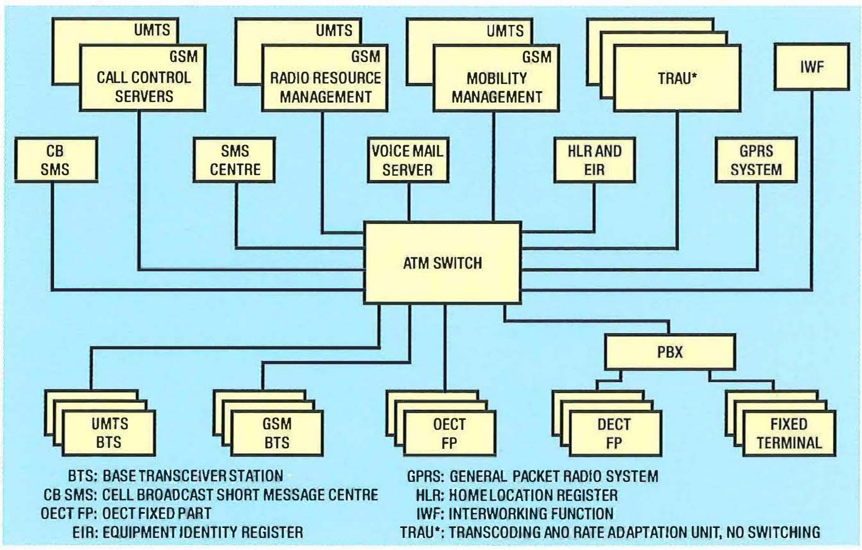


Figure 8

Content of UMTS Phase 1: Operation Possible in 2002

Services
- Multimedia:
  — 144–512 kbit/s for wide area mobility
  — 2 Mbit/s for restricted mobility
- High-quality speech using low bit rates
- Advanced addressing mechanisms
- Virtual home environment for service creation and service portability
- Seamless indoor, outdoor and far outdoor
- Dual mode/band of operation of GSM/UMTS in one network
- Roaming between GSM and UMTS networks

Terminals
- Dual mode/band GSM/UMTS

Access network
- New base station system (BSS) in the UMTS/FPLMTS spectrum
- Spectrum efficient

Core transport
- Evolution of GSM and ISDN
- Mobile/fixed convergence elements

Evolution in annual phases after Phase 1

Source: Chairman ETSI SMG

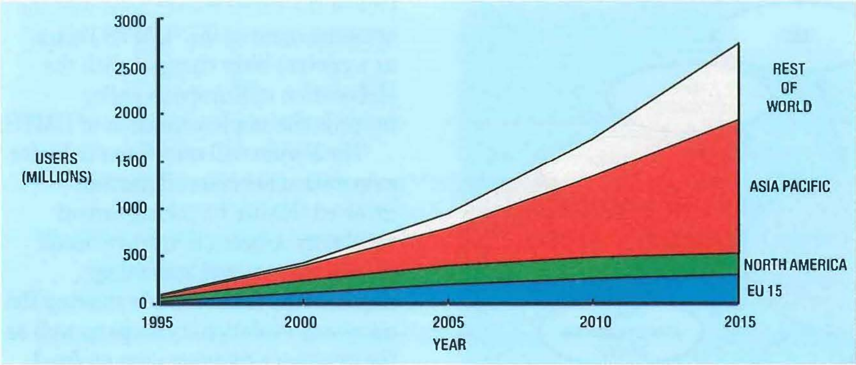


Figure 10 – Mobile users worldwide

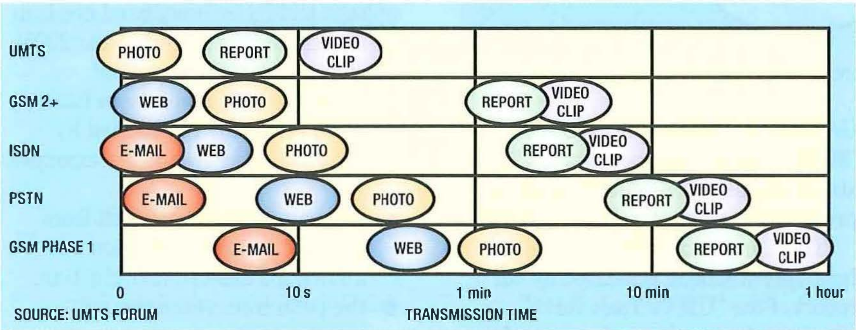


Figure 11 – UMTS speed

- universal personal telecommunications (UPT) functionality, and
- convergence of networks, applications and services.

A keyword for personal seamless mobility is *convergence*; for example, convergence of:

- fixed and wireless networks and applications,
- wireline and wireless services,
- fixed and mobile numbering,
- regional and global network access,
- IT and telecommunications, and
- telecommunications and consumer electronics (mobile multimedia).

Conclusions

Mobile communication services are among the fastest growing market segments in telecommunications. Technological evolution and international standards contribute to that success.

European mobile standards like GSM, are on the way to leading positions in additional regions of the world. Enhancements in the GSM standard enable the penetration of new market segments.

An evolutionary process towards future system generations (UMTS) will be the key for an undisturbed market growth (Figures 10 and 11).

Convergence of networks, technologies and services will be the key to a seamless mobility for the users and the key for opening mass markets in mobile communications.

References

1 UMTS Task Force Report, March 1996.

2 UMTS Forum Report No. 8.

3 Richards, Derek; and Jarvis, Andre. Lucent Technologies. Network Architectures for Multi-Standard Multi-Band Service Provision, June 1996.

Biography

Wolfgang Groenen
Lucent Technologies
Network Systems
GmbH

Wolfgang Groenen graduated with a diploma in telecommunications engineering. He has more than 20 years of experience in planning, product management and marketing of wireline and wireless telecommunications systems. He is currently the Director Industry Relations GSM and UMTS in Lucent Technologies' Wireless Network Group. For many years he has been Vice President of a section of the German Electronic Manufacturers' Association. He is the Chairman of the worldwide UMTS Forum Manufacturers' Group. He is the author of numerous publications and frequently invited as a speaker and/or chairman to international conferences.

Martin Lossie

Towards Mobile Telephony as the Default Solution for Personal Communications

This paper analyses the explosive growth of GSM-based mobile telephony in various countries. Substitution of fixed telephony by mobile telephony has been taking place in some countries and will continue to do so in most countries, albeit at different speeds. Strategies of mobile telephony operators have changed from offering and promoting mobile telephony as a premium service into offering mobile telephony as everyone's default solution for voice and even data communications, at tariff rates that are competitive with fixed network rates.

The strategic, commercial and economical viability of the mobile telephone as the default communications vehicle both at home and on the move is analysed. This includes a review of possible strategies for different types of operators such as incumbents that own both fixed and mobile business, existing second GSM900 operators without fixed network ties and start-up DCS1800 operators. The value proposition for the end-user is reviewed, as well as the business case for this new role of mobile telephony networks.

Enabling technological aspects are discussed such as spectrum requirements, advances in radio engineering, network capacity, homezone solutions, availability of high-speed data and advances in handset technology. Migration to third-generation mobile communications systems is discussed. The paper concludes with an international review of successful end-user applications that have transformed the mobile telephone from a 'talking tool' into a powerful personal assistant that also offers e-commerce transaction support.

Impressive Growth

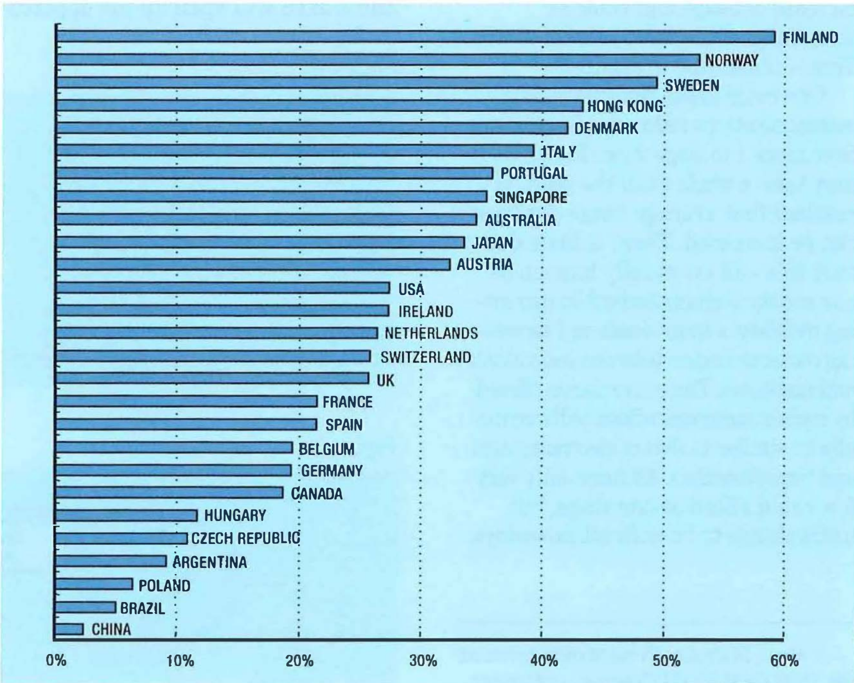
Figure 1 illustrates the impressive growth of mobile telephony users, where 16 countries have achieved penetration levels of more than 25%. These are penetration reaches levels that no one could dream of during the early 1990s, when the global system for mobile communications (GSM) was commercially introduced. In fact, with penetration of fixed lines ranging between 50% and 60% in most countries, mobile penetrations have reached a level that ranges from half or equal that of fixed lines in a number of countries. Growth is expected to continue as illustrated in Figure 2

Figure 2 clearly illustrates that penetration is expected to reach very high numbers, while the growth rate itself is expected to vary per country. It may be surprising that early adopters of mobile telephony such as the UK and the USA are not expected to end up as the countries with highest penetration, but even lag behind the Czech Republic for example.

Despite the growth of the subscriber base in mobile communications, a word of caution is appropriate as financial backers are increasingly worried about the sector's decimated profitability, with decreasing usage, falling tariffs and increasing churn,

Martin Lossie:
PricewaterhouseCoopers Management Consultants
Newtonlaan 115
PO BOX 8283
3503 RG Utrecht
The Netherlands
Tel.: +31 30 219 1140
E-mail:
martin.lossie@nl.pwcglobal.com

Figure 1 – Mobile penetration levels at March 1999 (source: EMC Publications)



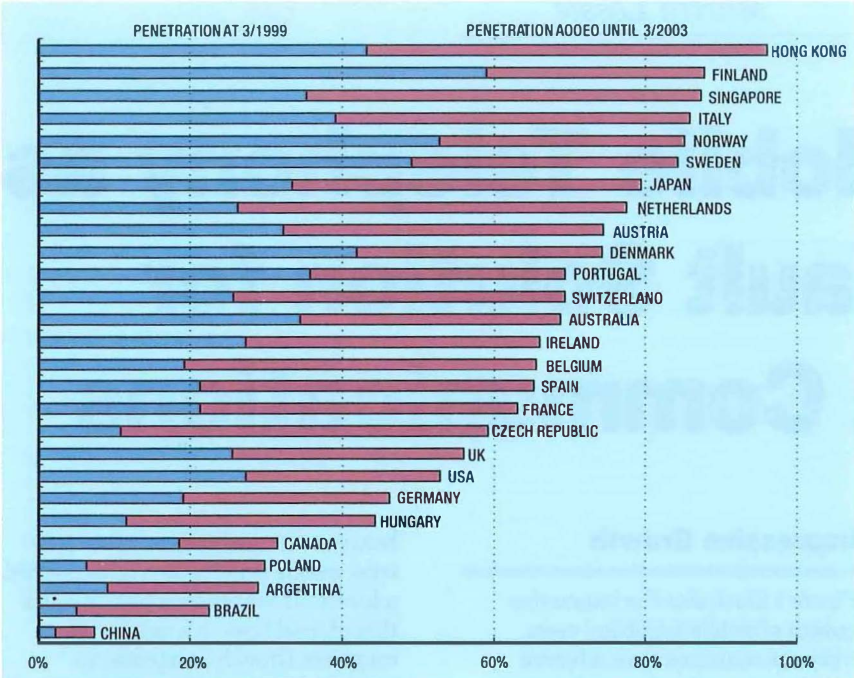


Figure 2—Forecast mobile subscribers by March 2003 (source: EMC Publications)

fraud and bad-debt†. On the other hand, it is perhaps too early to have worries over the long-term success of the mobile communications industry. In other industries, it has been demonstrated that customers are willing to spend a significant share of their wallet on the service as long as convenience is offered. Examples are the electricity industry with very high penetration levels and a multitude of applications, and the automotive industry. Both industries started as a premium service for the rich and happy few, followed by a growth of the customer base to near mass market levels and subsequently followed by increase of usage and value at completion of the mass market status. This is illustrated in Figure 3.

One could argue that the mobile communications industry is on its way from stage 1 to stage 2 (see Figure 3). It may take a while until the point is reached that average usage per user can be increased. There is little doubt that this will eventually happen as our society is characterised by increasing mobility of individuals and increasing communication between individuals and institutes. The convenience offered by mobile communications will eventually be similar to that of electricity, cars and (wrist)watches. All items only very few could afford at one stage, but unthinkable to be without nowadays.

† Ross, Malcolm. Presentation given at the 1999 GSM World Congress in Cannes.

Increasing Convenience

The increase of convenience associated with mobile communications is illustrated in Figure 4. Datacommunications used to be fixed location activity, where local area network (LAN) and wide area network (WAN) technology supported the exchange of high data volumes between locations. With GSM, the ability was introduced to do short messaging (SMS) and datacommunications at 9600 bit/s. That was sufficient for small amounts of data. Next generations of technology such as GPRS, HSCSD and UMTS will open up the opportunity for mobile datacommunications where the amount of data exchanged per event can be substantial. That opens up a range of applications that can drive up the convenience of using a mobile telephone, such as:

- large file exchange from a notebook computer when away from the office,
- browsing knowledge databases when away from the office,
- WWW access,
- e-commerce,
- image messaging (a picture says more than 1000 words doesn't it?),
- multi-party gaming (probably a good service to tap revenue from the scooter generation), and
- gambling (to tap revenue from the rest of the population)

The term *mobile telephone* may not be so correct any longer, since this relates perhaps too much to voice communications. As a matter of fact, users may not want any longer a single portable device that does it all. Perhaps users want a set of end user equipment that fits the particular application. A technology like Bluetooth may play a crucial role in such developments where the user has a network access device (NAD) that provides the prime link to the mobile network while end user devices such as laptop computers, organisers and gaming devices may be connected to the NAD and the other end user devices via Bluetooth technology (Figure 5).

The NAD can potentially be very small, and for example fit into the pocket of a suit without having the

Figure 3

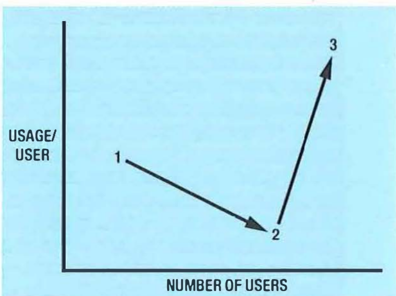


Figure 4

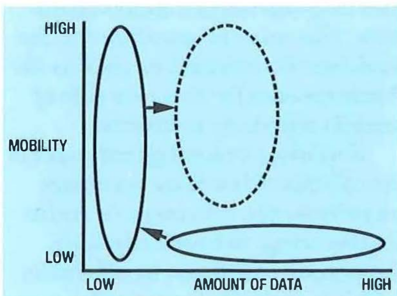
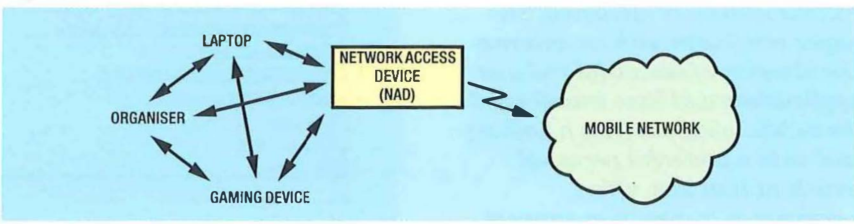


Figure 5



need to take it out other that for recharging batteries, or changing a SIM card. Making a telephone call should be done straight from the organiser: plug in the earpiece into the organiser, look up the number to dial, and have the notepad ready to scribble whatever is needed. As said before, the organiser connects to the NAD and through the NAD to the network.

Commercial, Technical and Economic Viability

A scenario where mobile communications is the default for most of our communications is commercially and technically viable:

Commercial viability

- The use of the mobile network based applications can be made attractive even when being at locations where a fixed network is available by offering differentiated tariff options for being on the move or stationary at home or the office
- The customer will have a transparent customer experience through familiar user interfaces at all times.

Technical viability

- Offering differentiated tariffs for being on the move or stationary at home or the office is already applied in Denmark through so-called *homezoning* technology
- Regulators are likely to free up large amounts of spectrum for value-creating mobile applications
- Technology for handportable end-user devices has been developing quickly:

—increasing battery life;
—large and cost efficient memory;
—full colour liquid crystal display (LCD) technology;
—decreasing power consumption;
—decreasing size and weight;
—Internet user interface under development (WAP);
—WAP and SIM toolkit offers over-the-air programmable menus to support context driven applications (including JAVA); and
—compression technologies keep bandwidth requirements to a minimum.

Economic viability

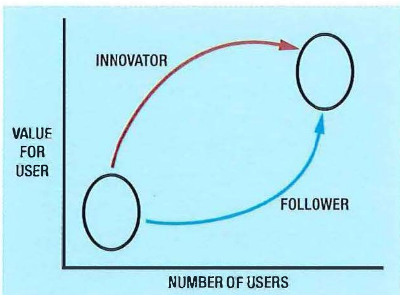
The economic viability is more difficult to assess. New technology developments do not come without risk for those who invest in the infrastructure and end-user devices. The key question is: where is the value going to be created? It is useful to look at the another industry: water. The water network typically is everywhere, and everyone uses it as needed. It goes even as far as people using vast amounts of drinking quality water to maintain their garden or wash the car. The analogy with (mobile) communications is that bandwidth and usage may become virtually unrestricted by the cost of the basic service. In the water industry arena, content is the real creator of value. Think about Coca Cola. Add a very specific taste to a litre water and the value per litre is up by a factor 1000 easily. Content is not even needed. Some mineral waters are said to be in fact not much more than carbonated tap water under a well-developed brand name. Another good example for adding value in the water industry is the health and spa resort. Have people take a bath there and see how much they are willing to pay compared to taking a bath at home.

The real challenge for the mobile communications industry sector is to have applications developed that offer true value to the end-user such that willingness to pay for the application is sufficiently high to make the business case. This may have to include strong branding like 'application connected by XYZ'; compare with 'Intel inside'.

Two core routes towards increased value can be followed to get there: innovation and following, as illustrated in Figure 6.

The innovator will develop true value-added services at an early stage while growing the subscriber base. The follower will focus at growing the subscriber base in the first place, and focus at increased value for the user at a later stage. In particular, the innovator path may be a tricky one to follow. Look at the Apple case: high-value highly-appreciated graphical applications that made the firm a standard in for example market communications and desk-top publishing environments, while DOS was the best Microsoft could offer to its customers at that time. Now Microsoft is said to have about 90% of the PC operating system market with its Windows products with increasing value for the user, leaving Apple far behind.

Figure 6



Innovator strategies may work however. Intel is probably a good example of an innovative firm that shared some of the PC experience with Microsoft in the days that the 8086 and 8087 processors were the flagship products. Innovation at Intel seems to be a never-ending story with generations such as 286, 386, Pentium, Pentium II following each other at a steady pace.

Strategies for the Mobile Operators

Where does all this leave mobile operators?

The answer is: probably confused and surrounded with uncertainty. Key questions for a mobile operator are:

- Should we take the lead in developing innovative value-added applications, or should we assume a role as follower?
- Should we do all the value creation ourselves, or should we involve third parties?

The answer to these questions can be pragmatic: develop innovative value-added applications for market segments where you have a well established position, and take a more passive role for other segments. Such a passive role may be one of being a follower, leaving initiative to competition in these non-priority segments. It may however also be a role of changing the position in the value chain, involving third parties for being the innovator (at their own risk) and reducing the own role to a plain old mobile service (POMS) operator.

It is hard to tell what strategy would work best for different types of mobile operators (first-operator GSM tied to a fixed network, second-operator GSM or DCS1800). Everyone is moving from position 1 to 2 as in Figure 3. A long-term winning position has to be achieved on the way from 2 to 3. Key considerations to take into account when choosing a strategy are:

- nature and needs of the market segments where an established position has been achieved, and
- internal capabilities to develop innovative applications

Operators who choose the evolution towards a highly value-added mobile services and applications environment may need the spectrum for fast-data applications. A UMTS licence is essential in that case. The migration from current GSM services to UMTS-based services will certainly require all kinds of network and radio engineering issues to be addressed. But the real crux will be to offer true benefits to the end-user by means of the value-added applications as discussed before. For migrating towards the wonderful world of unlimited UMTS-based services, it is probably never too early to start working with the priority market segments to develop value-added applications with currently available technology. That can help operators go through some of the learning curve now and create a certain level of intimacy with those priority segments ('working together on increasing the utility of mobile communications'). At the same time, network operators can learn about complex radio network engineering, with services that offer two-wall indoors penetration where needed, supporting all kinds of mobility speeds as well, while providing a capacity that is sufficient to serve the mass market.

Already it is possible to encounter useful applications that make the mobile telephone much more than a talking device:

- automatic teller machine type functionality, to do banking transactions from the telephone, using the identification and authentication capabilities of GSM;
- dial-a-coke—calling a number on the vending machine and getting a coke (charged to the telephone bill);
- dial-a-wash—calling a number to pay for the car wash (charged to the telephone bill); and
- on-demand information services such as traffic, weather, stock exchange

These are services that have one thing in common: they serve the user on particular moments in life and increase convenience levels. As everybody's lifestyle is different, operators should not be hunting for the one big killer application. Personalisation is the killer application and in that sense telecommunications is not different from a pizza delivery service: who wants salami gets salami, who wants mushrooms gets mushrooms. Do your own pizza—do your own telecommunications service. Being able to adapt quickly to changing customer needs, is what will make the real winner in this business. Understanding the customer is therefore key, as well as the ability to use the knowledge on the customer base in the right way.

Conclusion

Mobile telephony as the default vehicle for our personal communications is a scenario for the future that is not at all unrealistic. A large number of countries are expected to have mobile penetration levels exceeding 60% by 2003. It is up to the industry to drive up the value of the service and associated applications such that sustainable revenue streams can be generated at levels that are high enough to support the investments required. The commercial success of GSM is less than 10 years old. It took other industries much longer to arrive at high-volume and high-value positions. Continued focus on the customer and personalised services through exploiting the technology capabilities are likely to be key to future success.

Acknowledgements

All trademarks are acknowledged.

Biography

Martin Lossie
PricewaterhouseCoopers Management Consultants

Martin Lossie graduated as an electrical engineer from the Delft University of Technology in The Netherlands. He has over 11 years experience in the telecommunications industry and is currently with PricewaterhouseCoopers Management Consultants. His current specialism is technology and market oriented strategy for mobile telephony operators. He has been working with leading mobile telephony clients in setting up new mobile businesses as well defending positions of existing mobile businesses in a number of countries.

Delphine Plasse

Contacting a UMTS User: Naming/Numbering and Routing Issues

This paper presents a mobility issue related to a long-term vision of the universal telecommunications system (UMTS). Future UMTS call scenarios introduce the concept of globally unique, portable, non-service provider related user identifiers. This paper analyses the user identifier requirements and measures them against those of existing naming, numbering schemes such as E.164 numbers and domain names. This paper also comments on some of the high-level issues around futuristic UMTS service provision scenarios and call scenarios involving locating users, media negotiation and routing strategies.

Introduction

In the long term, the universal mobile telecommunications system (UMTS) is viewed as the evolution of telecommunications into the 21st century, supporting multimedia services on a converged infrastructure with interworking with current fixed and mobile networks. There will be no distinction between fixed and mobile users. As a UMTS service subscriber, you will be able to contact users on their UMTS identifier, regardless of prior knowledge of their location, the type of terminal they are using at the time, the type of networks they are currently on (such as the Internet, global system for mobile communications (GSM), public switch telephone network (PSTN), etc.), or their current network addresses.

This paper presents a mobility issue related to a long-term vision of UMTS. Future UMTS service provision and call scenarios introduce the concept of globally unique, portable, non-service provider related users' identifiers. This paper comments on some of the high-level issues around futuristic UMTS call scenarios involving locating users, media negotiation and routing strategies. It analyses the user's identifier requirements and measures them against those of existing naming, numbering schemes such as E.164 numbers and domain names.

UMTS

Today, mobile networks mainly provide person-to-person voice communications and are designed and operated as separate entities from the fixed networks, requiring users to accept limited interworking, multiple numbers, multiple subscriptions, multiple service environments, and multiple service providers[1].

Therefore a new mobile communication system, the universal mobile telecommunication system (UMTS) is required to offer significant user benefits including high-quality wireless multimedia services to a convergent network of fixed, cellular and satellite components. It aims to offer mobile personalised communications to the mass market regardless of location, network or terminal used[2]. The European Telecommunications Standards Institute (ETSI) specifications for UMTS phase 1 are a new air interface (wideband code-division multiple access (W-CDMA)), a new access network (UMTS terrestrial radio access), and a core network based on the global system for mobile communications (GSM) and general packet radio service (GPRS). Nevertheless, as it stands, UMTS (phase 1) will not support mobile multimedia nor service/terminal/personal mobility. A GSM-based core network limits the fixed-mobile integration. BT has been supporting a new approach of a core network based on the Internet protocol (IP) to deliver voice and data, and interworking with heterogeneous access technologies (see Figure 1). This vision offers new life scenarios where heterogeneous access networks (GSM, UMTS, and fixed technologies) can internetwork over a flexible platform supporting high data rates, voice, data and multimedia. This fixed-mobile integration vision also affects traditional names and addresses roles.

The Problem

Need for new scenarios

With an IP-based core network, UMTS will support voice, data and multimedia over fixed and mobile networks. Terminal and user mobility in multi-environments will be expected. The core network (GPRS nodes and gateways) will internetwork with fixed existing networks (asynchronous transfer mode (ATM), PSTN, integrated services digital network (ISDN) and mobile legacy networks (GSM), wireless local area networks (LANs)).

Delphine Plasse:
British Telecommunications plc
Post Point: 131C
Adastral Park
B55 Materials Building
Martlesham Heath, Ipswich
Suffolk, IP5 7RE, UK
Tel: +44 1473 642127
Fax: +44 1473 646885
E-mail: delphine.plasse@bt.com

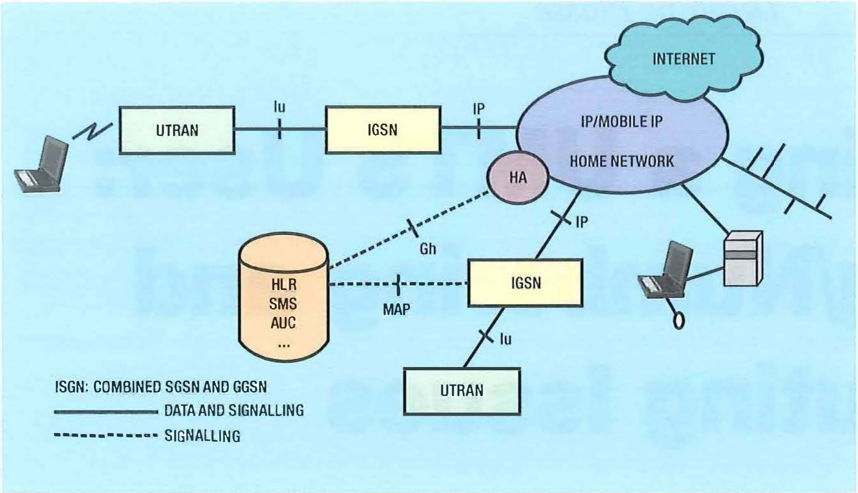


Figure 1—A possible UMTS core network architecture, evolving from the current GPRS backbone elements (ETSI, Draft Technical Report v.0.3, Jan. 1999)

This vision introduces new scenarios. Ideally, users will access their services from any network (Internet, PSTN, GSM, etc.), at any time, with access and delivery via the most appropriate network and terminal for the user at a given time.

New service provision scenarios are introduced since the type of communication is vastly more complicated than in, say, GSM. The network intelligence in conjunction with the terminal intelligence would decide which means of delivery to use based on the caller and called party preferences, charging, urgency and on the network capabilities. Several choices may be available if the user is registered on multiple networks and terminals, and the information may have to be translated from one format to another. For example, you may want to call 'Paul Smith' from your web page; the 'network' finds out 'Paul Smith' subscriber is within a particular area on a GSM network and that he carries a mobile telephone. A ring and a message is displayed on Paul's terminal showing the caller's identity, type of communication (voice, file transfer, e-mail, etc). The same is applied on the caller side together with some charging information. Then the network will route the end-to-end call on the appropriate terminal of 'Paul'.

Similarly, new call scenarios will affect the traditional concept of numbers and addresses. Users could be roaming on different types of networks, using the terminal convenient at the time and, furthermore, may have multiple service providers. If someone wants to contact one of these users then he/she does not know where they are, which network they are on at the time of the call, whether they are fixed or mobile, nor what terminals they are using at the time.

Phases of a multi-media call

Resolution

In order to call these users, a user will first need to dial a unique user identifier (a name or number) and then the network will resolve it into the destination network address and user terminal for the current called party's service provider, in order to route the call (see Figure 2).

The terminal capabilities can be enhanced by some service interfaces such as web-based front-ends and interactive voice systems, which allow users to spell an alphanumeric user's name over an ordinary telephone or keyboard-less mobile terminal. After querying a name directory, the database returns the address of the targeted user's service provider. The called party's current location is stored in a location database either at the service provider or at the network he/she is on. The service provider addresses can be cached on the calling client (phase1). Because the user has neither tariff indications nor network capabilities indications in the number dialled, the network has to inform the user about some characteristics of the call: quality of service, called party's current number, call cost, user availability, terminal capabilities, diversion options, etc.

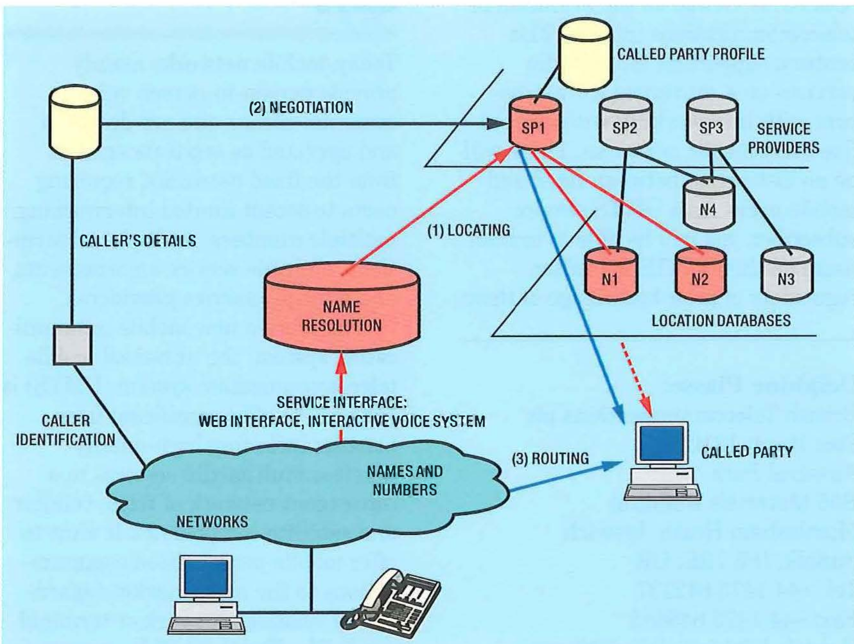
Negotiation

The intelligence in a UMTS negotiation server/database can also compare the call characteristics and charging against the caller's communication subscription (phase 2). It takes the decision to proceed to call (for example, divert it from the PSTN to the called party's PC, or just bar the call) without getting back to the caller. The negotiation server/database is an enhancement of the virtual home environment concept as described in the UMTS concept[3].

Routing

During the negotiation phase the network intelligence will have decided which means of delivery to use (phase 3). But several choices may be available. Two of them are

Figure 2—Locating (1), negotiation (2) and routing (3) phases of a call



that the call can be proceeded from the caller's terminal or from the called party's service provider. If the call is routed from the service provider's network, it can also provide interworking interfaces without affecting the clients but may affect the call tariff. If the call is routed from the caller's terminal, the service provider will need to send back the called party's address plus other call characteristics. Some specific interfaces and intelligence on the client is expected to support the routing.

In a call scenario where the called party's location, service provider, network and terminal used, and cost of calls are not known at the time of the call, none of this information can be embedded within the dialled number. The user's identifier becomes an entity only referring to a person/ persona.

Names, Numbers and Addresses Roles

Traditionally, telephone numbers act as users' identifiers, network locators, routing addresses, and as an indication of the cost of the call. E.164 provides the numbering structure and functionality for telephone numbers used for international public telecommunication: geographic areas, global services, and networks. In E.164 a number indicates the public network termination point and it contains the information necessary to route the call to this termination point either by some of its parts or externally by some mapping functions. E.164 fields for geographic areas are the country code (CC), national destination code (NDC) that is optional and the subscriber number (SN). E.164 fields for networks are composed of a country code, an identification code (IC) for the required network and a subscriber number (SN).

In order to support call scenarios where the network termination point is not known at the time of the call, it is crucial to separate the addressing and routing roles from the user's identifier. BT put forward at the ITU-T and ATM Forum standard bodies the following definition:

The role of a name is to identify end-users and it may be portable. It can be an alpha label, a numeric label (a number obtained from a recognised numbering plan) or an alphanumeric label (for example, 07074 Hertz). The role of an address is to locate a physical end-point within the network. It is used for routing information; it is technology-related and is not portable. It can be a numeric string of characters, alphanumeric characters, related to the network topology. The key difference between the name and address concept is that you can migrate from address1 to address2 and still keep the same name/ number. A name/number can be associated with many addresses, or multiple names may be associated with a single address and the mapping is stored in a database.

Need for a User's Identifier

At the time of the call, the called party network is not known. The user's identifier acts as a name and it needs to support basic requirements: to uniquely identify the user across fixed and mobile networks, not to identify a terminal (terminal SimCard), to be memorable, to be non service provider related (unlike myself@bt.com), to be non topologic, to be portable, to be free from tariff assumptions.

Personal numbers as users' identifiers

Personal numbers offer a limited approach to the problem. These logical numbers identify a person, they are shared between all GSM/ PSTN network operators, and allow you to have your calls routed to any single telephone you might be using at the time. You can control your routing service to any telephone, fixed and mobile or to a voice mailbox by calling a re-routing number and then by dialing your personal number and PIN number. Then you enter the new destination for your calls to be routed to.

UMTS requirements

But to support the UMTS call scenarios, the user's identifier needs to offer some other features. It needs to break away from traditional tariff assumptions and charging mechanisms. Indeed, a UMTS call will offer in-call variable bandwidth according to the networks it will be delivered on and it will require another charging mechanism (other than time and distance). Furthermore, the called parties should not have to forward their calls themselves and could expect the network to do it automatically on their behalf.

Legacy networks

Finally, one needs to consider having one's calls rerouted to an Internet PC. The GSM/PSTN network intelligence will need to interoperate with the Internet and vice versa. Currently, the Internet domain name service does not resolve E.164 names into network resource IP addresses, which affects the Internet world integration to legacy networks[4].

Possible Solutions

In the UK, one of the regulator's requirements is for new ranges of national E.164 numbers to be portable between network operators; that is, to behave as names[5]. One option is to allocate E.164 names to users. However, uniqueness and portability of the E.164 identifiers could only be guaranteed between UK service operators. Nevertheless, it is possible to extend a service to a global one. An operator can obtain from the ITU 'global numbers' which are E.164 numbers with a special country code and are allocated to a specific global type of service. But these numbers are only allocated to special value-add global services that do not compete with services offered by the same operator using national numbers.

Currently, the most advanced naming scheme comes from the computer world. Internet names offer globally unique identifiers, alphanumeric, widely used and resolvable. In the UK, BT 's Millennium Project will offer to everyone in the country an e-mail address[6]. This will provide a globally unique identifier to every potential user. Nevertheless, since Internet names are designed to identify some components of the network, they are only portable within their domains (Internet service provider domain).

Developments

Friendly names

The portability and friendliness of names is raising a lot of interest. One IETF Working Group is currently looking at naming schemes for Internet resources: the *human friendly names* (HFNs)[7]. Although unified resource locators (URLs) (such as http://guest.btinternet.com/html/about.html to visit BT Internet products) provide a powerful mechanism to resolve the location of resources on the Internet, for many applications URLs are complex, totally unpredictable, and too lengthy to memorise. HFNs aim at simplifying the Internet navigation via a simpler globally-available human-friendly naming system. That scheme is being implemented to replace the use of URL-based navigation. Users can navigate the Web by typing unique and meaningful words like company names (SUN, Opel, etc), product titles (Mondeo, *OneNumber*), trademarks (Nike), or advertising slogans without the need to remember the unfriendly URL syntax (*'just do it'* or *'it's good to talk'*). Although several organisations have deployed services that resolve the friendly names (such as the RealName system developed by Centraal), there have been no efforts to standardise the information going to and coming from the service. The HFN group intends to address these issues and release standards to allow the services to become widespread and interoperable.

Another working group is to be established for the development of the common name resolution protocol (CNRP), which aims to process HFNs and return a URL to the user interface[8]. Browsers would have a CNRP client embedded in them. Some of the anticipated applications of HFNs and of the CNRP are an increased visibility of Internet sites and products, search engines improvements, and the resolution of E.164 numbers (acting as addresses but also as names) into an address (postal address or other address spaces). BT may want to run a 'British' CNRP registry.

Interworking

Interworking between heterogeneous networks is another major area of investigation. With the convergence of the Internet and PSTN, there is an increasing desire for events occurring in the PSTN domain to be propagated to the Internet domain. For example, the IPTEL WG is looking at PIN services for advanced caller ID delivery. Scenarios such as the one following are being studied: a display is to appear on your PC that your telephone is ringing at home, and you can then pick up the call from your PC. A service subscriber has one or more telephone numbers (residential, business, mobile). The service subscriber also has many Internet accounts. The PSTN tries to contact the subscriber and if this fails then it handovers to an advanced PINT gateway, which inspects each of the accounts. If the subscriber is on-line then a popup window appears on the screen with a clickable caller number. If the subscriber is off-line then he/she is be sent an e-mail with the caller ID included[9].

For applications such as Internet telephony, users of the PSTN may want to address an Internet entity on a telephony number. Therefore, using E.164 telephony addresses as IPv6 addresses has been proposed: a unique addressing scheme for future telephony equipment that has simultaneous access to PSTN services and to the Internet via the IPv6 protocol. It can also be used for IP telephony where both IP and PSTN routing information are needed simultaneously[10].

Mobility

Regarding mobility, roaming and user naming schemes on the Internet domain, the ROAMOPS Working Group is to develop procedures, mechanisms and protocols to support user roaming among groups of Internet service providers (ISPs). The ROAMOPS group is specifically concerned with the movements of users. So far, the group has been describing the basic mechanisms required to support user roaming and a description of several existing roaming implementations. The group is also defining a standard username syntax to support roaming[11].

Conclusion

In a world where roaming on heterogeneous networks, terminal mobility, personal mobility and multiple service providers are supported, the user's identifier will break away from traditional numbers, and will act as a name. Existing user naming schemes fail to support uniqueness, portability and globality at the same time. As the user's identifier becomes a name, added complexity to the user interface, service interface and to name resolution mechanisms are expected. Many activities at the IETF have started investigation of naming and addressing issues regarding interworking of heterogeneous networks, roaming between different ISPs, and portability of Internet resources identities. Today, Internet resource naming schemes are becoming necessary, but, in the author's view, new user's naming schemes are soon to require the same attention.

References

1 MACKENZIE, FIONA. The Challenge of Universal Personal Communications. FITCE 97.
2 UMTS Forum, 1997.
3 ETSI TR 22.70 on the VHE concept.
4 E.164 to IP mapping BOF. 42nd IETF meeting. http://www.ietf.org/
5 OFTEL (UK Office of Telecommunications). http://www.oftel.gov.uk/
6 *BT Today*, April 1998.
7 Requirements for Human Friendly Identifiers. http://search.ietf.org/internet-drafts/draft-mealling-human-friendly-identifier-req-00.txt
8 A Resolution Protocol for Common Name Namespaces. http://search.ietf.org/internet-drafts/draft-popp-cnrp-00.txt
9 Proposal by AFIFI, HOSSAM and TOUTAIN, LAURENT (ENST Bretagne); and BOUND, JIM (Compaq Computer Corporation). IP Telephony area. 44th IETF meeting Minneapolis, USA. http://www.ietf.org/
10 PSTN Internet Notification BOF. 44th IETF meeting. http://www.ietf.org/
11 ROAMOPS WG. 44th IETF meeting. http://www.ietf.org/

Biography

Delphine Plasse
British Telecommunications plc

Delphine Plasse graduated from the Institut National des Sciences Appliquees de Toulouse, genie Physique in 1996. She also gained an M.Sc. in Optical Communications from the University of Essex. In 1996, she joined BT in 1996 to work on mobility-related issues. Currently she is developing a virtual home environment architecture for the future UMTS architectures.

Stefan Pütz

Security for the Third-Generation Mobile Radio System UMTS

This paper describes the third-generation (3G) security principles specified by the Third-Generation Partnership Project (3GPP) for the third-generation mobile system UMTS as well as the 3G security architecture itself. This is based on five sets of 3G security features and mechanisms. In particular, the differences between second-generation (2G) and 3G security are discussed. The 3G security architecture addresses and corrects all the weaknesses in the 2G systems.

Introduction

Within the last six months a harmonisation and globalisation process for the third-generation mobile radio systems has taken place. Therefore the standardisation process for the universal mobile telecommunications system (UMTS) has moved from the European Telecommunications Standards Institute (ETSI) to the Third-Generation Partnership Project (3GPP). 3GPP is supported by various standards organisations and other related bodies from different continents and countries, for example, Europe (ETSI), Japan (ARIB, TTC), Korea (TTA) and North America (T1), that have agreed to cooperate for the production of a complete set of globally applicable technical specifications for a 3G mobile system based on the evolved global system for mobile communications (GSM) core networks. Third-generation mobile radio systems will involve more players, for example, content and service providers, and more operators, which will result in more roaming. Therefore 3G systems will exist alongside and interact with a lot of different network types. Also 3G systems will promote wireless as the preferred means of communication.

To ensure that UMTS will work securely and reliably under these conditions, a new 3G security architecture has been defined. The scope of 3G security is described by principles. These principles state what is to be provided by 3G security as compared to the security of 2G systems. Security elements within GSM and other 2G systems that have proved to be needed and robust will be adopted. Also 3G security will improve on the security of 2G systems. Therefore 3G security will address and correct real and perceived weaknesses in 2G systems. Last but not least, 3G security will offer new security features and will secure new services offered by 3G mobile radio systems.

3G Security Principles

The following three key principles form the basis of 3G security[1]:

- 3G security will build on the security of 2G systems. Security elements within GSM and other 2G systems that have proved to be needed and robust will be adopted for 3G security[4, 5]. Therefore 3G security will retain (and in some cases develop) the following security elements of 2G systems:

  —authentication of subscribers for service access;
  —radio interface encryption†;
  —subscriber identity confidentiality on the radio interface;
  —the SIM as a removable hardware security module that is manageable by network operators and independent of the terminal as regards its security functionality;
  —subscriber identity module (SIM) application toolkit security features providing a secure application layer channel at least between the SIM and a home network server; and
  —home environment (HE) trust in the serving network (SN) for security functionality is minimised.

- 3G security will improve on the security of 2G systems by the way that it will address and correct real and perceived weaknesses in 2G systems as listed below:

  —active attacks based on masquerading a base transceiver station (BTS) are possible;
  —cipher keys and authentication data are transmitted in clear between and within networks;
  —encryption does not extend far enough towards the core network resulting in the cleartext transmission of user and signalling data across microwave links;
  —the provision of protection against channel hijack relies on the use of encryption;

† The strength of the encryption will be greater than that used in 2G systems; the strength is a combination of key length and algorithm design. This is to meet the threat posed by the increased computing power available to those attempting cryptanalysis of the radio interface encryption.

Dr. Stefan Pütz:
T-Mobil (DeTeMobil, Deutsche Telekom MobilNet GmbH)
POB 300463, 53184 Bonn, Germany
E-mail: stefan.puetz@t-mobil.de

—data integrity is not provided†;
—there is no HE knowledge or control of how an SN uses authentication parameters for HE subscribers roaming in that SN; and
—2G systems do not have the flexibility to upgrade and improve security functionality over time.

- 3G security will offer new security features and will secure new services offered by 3G. More on this issue is discussed in the 'Outlook' section.

As a priority, 3G security will provide the proven 2G security features and correct the weaknesses in 2G systems. Security for new services and service environments will then be developed as required.

3G Security Architecture

The security architecture for the 3G mobile system, specified in Reference 3, consists of five sets of security features and security mechanisms. A security feature is a service capability that meets one or several security requirements. The complete set of security features addresses the security requirements as they are defined in Reference 2. A security mechanism is an element that is used to realise a security feature. All security features and security requirements taken together form the security architecture.

Figure 1 gives an overview of the complete 3G security architecture.

Five security feature groups are defined. Each of these feature groups meets certain threats, and accomplishes certain security objectives. The security feature groups are as follows:

- *network access security (I)*: the set of security features that provide users with secure access to 3G services, and which in particular protect against attacks on the (radio) access link;
- *network domain security (II)*: the set of security features that enable nodes in the provider domain to securely exchange signalling data, and protect against attacks on the wireline network;
- *user domain security (III)*: the set of security features that secure access to mobile stations;
- *application domain security (IV)*: the set of security features that enable applications in the user and in the provider domain to securely exchange messages; and
- *visibility and configurability of security (V)*: the set of features that enables the user to inform himself whether a security features is in operation or not and whether the use and provision of services should depend on the security feature.

† Data integrity defeats certain false BTS attacks and, in the absence of encryption, provides protection against channel hijack.

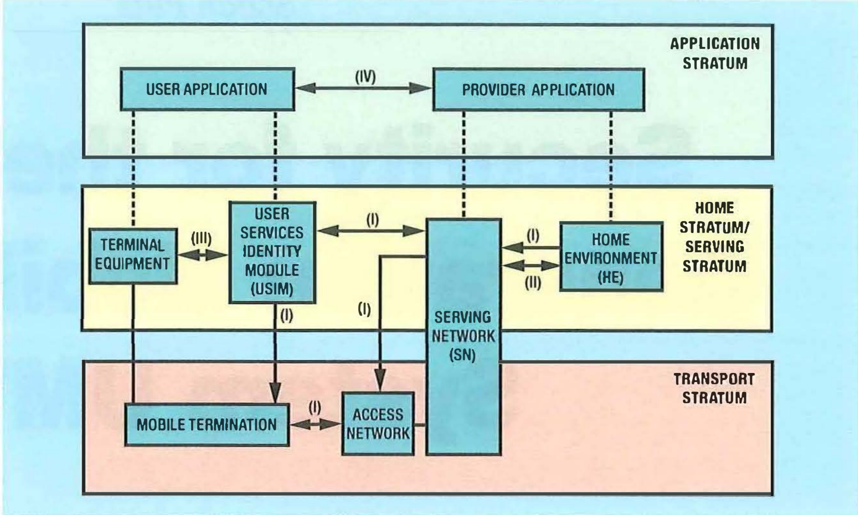


Figure 1—Overview of the security architecture[3]

The following subsections describe the security features belonging to these groups provided by the 3G security architecture.

Network access security

Enhanced user identity confidentiality is the property that the permanent user identity (international mobile user identity (IMUI)) of a user to whom a service is delivered cannot be eavesdropped on the radio access link. Also *user location confidentiality* and *user untraceability* will be provided. Therefore the presence or the arrival of a user in a certain area cannot be determined and an intruder cannot deduce whether different services are delivered to the same user by eavesdropping on the radio access link. To achieve these objectives, the user is normally identified by a temporary identity by which he/she is known by the visited serving network, or by an encrypted permanent identity. To avoid user traceability, which may lead to the compromise of user identity confidentiality, the user should not be identified for a long period by means of the same temporary or encrypted identity. To achieve these security features, in addition it is required that any signalling or user data that might reveal the user's identity is ciphered on the radio access link.

Mutual authentication between user and network will be achieved by the combination of user and network authentication. User authentication is the property that the serving network corroborates the user identity of the user, and network authentication is the property that the user corroborates that he/she is connected to a serving network that is authorised by the user's HE to provide services; this includes the guarantee that this authorisation is recent. To achieve these objectives, it is assumed that authentication should occur at each connection set-up between the user and the network. Two mechanisms have been included: an authentication mechanism using an authentication vector delivered by the user's HE to the serving network, and a local authentication mechanism using the integrity key established between the user and serving network during the previous execution of the authentication and key establishment procedure (see also data integrity).

Confidentiality of user and signalling data is the property that this data cannot be overheard on the radio access interface. This will be achieved by ciphering on the radio interface as well as further back into the network. Cipher key agreement is realised in the course of the execution of the mechanism for authentication and key agreement. Cipher algorithm agreement is realised by means of a mechanism for security mode negotiation between the user and the network. This mechanism also enables the selected ciphering algorithm and the agreed cipher key.

Data integrity and origin authentication of signalling data is the property that the receiving entity is able to verify that signalling data has not been modified in an unauthorised way since it was sent by the sending entity and that the data origin of the signalling data received is indeed the one claimed. Integrity key agreement is realised in the course of the execution of the mechanism for authentication and key agreement. Integrity algorithm agreement is realised by means of a mechanism for security mode negotiation between the user and the network. This mechanism also enables the selected integrity algorithm and the agreed integrity key.

Network domain security

Network element authentication is the property that a network element corroborates the identity of another network element it wants to communicate with. This feature ensures that no malicious operational or maintenance commands can be injected into a network domain by an intruder. It provides network elements, in particular network elements belonging to different network operators, with the possibility to corroborate each other's identities before exchanging data. This goal may be achieved either by an explicit or implicit entity authentication mechanism, to be performed each time data is exchanged between two network entities.

Confidentiality of exchanged data is the property that data exchanged between two network elements cannot be eavesdropped. If authentication data can be eavesdropped in the network domain, serious fraud problems can arise. Therefore, these features are needed to ensure the confidentiality of sensitive data, for example, authentication or other subscriber data inside the network domain. The features may be realised in the course of an authentication and key agreement mechanism performed by the network elements; the agreed cipher key is then used for securing signalling and user data by means of the cipher algorithm.

Data integrity and data origin authentication of signalling data is the property that the receiving network element is able to verify that signalling data has not been modified in an unauthorised way since it was sent by the sending element and that the data origin of the signalling data received is indeed the one claimed. The feature data integrity of signalling data ensures that operation and maintenance commands or user data exchanged between two network elements cannot be modified by an intruder without being detected, and it ensures that no malicious operational or maintenance commands can be injected into a network domain by an intruder. The features may be realised in the course of an authentication and key agreement mechanism performed by the network entities involved; the agreed integrity key is then used for securing integrity of the exchanged data by means of the integrity algorithm.

User domain security

User-to-USIM authentication provides the property that access to the USIM (UMTS SIM) is restricted until the USIM has authenticated the user. Thereby, it is ensured that access to the USIM can be restricted to an authorised user or to a number of authorised users. To accomplish this feature, user and USIM must share a secret (for example, a PIN) that is stored securely in the USIM. The user gets access to the USIM only if he/she proves knowledge of the secret.

USIM-terminal link ensures that access to a terminal or other user equipment can be restricted to an authorised USIM. To this end, the USIM and the terminal must share a secret that is stored securely in the USIM and the terminal. If a USIM fails to prove its knowledge of the secret, it will be denied access to the terminal.

Application security

UMTS is expected to enable operators or third-party providers to create applications which are resident on the USIM or the terminal. Therefore there exists a need to secure messages which are transferred over the 3G network to applications on the USIM or terminal, with the level of security chosen by the network operator or the application provider. *Secure messaging* between the USIM and the network provides a secure application layer channel at least between the USIM and a home network server (similar to the SIM application toolkit in GSM[6, 7]).

Network-wide user traffic confidentiality provides users with the assurance that their traffic is protected against eavesdropping across the entire network, not just on the radio links in the access network, which are particularly vulnerable, but also on the fixed links within the core network.

Security visibility and configurability

Visibility of the operation of security features will be provided in a way that it is transparent to the user if a security feature is enabled or not. This yields to a number of features that inform the user of security-related events, such as:

- indication of access network encryption: the property that the user is informed whether the confidentiality of user data is protected on the radio access link, in particular when non-ciphered calls are set-up;
- indication of network-wide encryption: the property that the user is informed whether the confidentiality of user data is protected along the entire communication path; and
- indication of the level of security: the property that the user is informed on the level of security that is provided by the visited network, in particular when a user is handed over or roams into a network with lower security level (for example 3G to 2G).

Configurability is the property that the user and the user's HE can configure whether the use or the provision of a service should depend on whether a security feature is in operation. A service can only be used if all security features, which are relevant to that service and which are required by the configurations of the user or of the user's HE, are in operation. The following configurability features are suggested:

- enabling/disabling user-USIM authentication: the user and/or user's HE should be able to control the operation of user-USIM authentication; for example, for some events, services or use;
- accepting/rejecting incoming non-ciphered calls: the user and/or user's HE should be able to control whether the user accepts or rejects incoming non-ciphered calls;

Table 1 2G Security Elements Covered by 3G Security Architecture

	Authentication of subscribers	Radio interface encryption	Subscriber identity confidentiality	SIM as a removable hardware security module	SIM application toolkit security features	Minimisation of HE trust in the SN for security functionality
Network access security						
Enhanced user identity confidentiality			x			
Mutual authentication between user and network	x					
Confidentiality of user and signalling data		x				
Data integrity and data origin authentication of signalling data						
User domain security						
Network element authentication						x
Confidentiality of exchanged data						x
Data integrity and data origin authentication of signalling data						x
User domain security						
User-to-USIM authentication				x		
USIM-terminal link				x		
Application security						
Secure messaging					x	
Network-wide user traffic confidentiality						
Security visibility and configurability						
Visibility						
Configurability						
Basic assumption for UMTS				x		

Table 2 2G Security Weaknesses Covered by 3G Security Architecture

	Active attacks based on masquerading BTS	Cipher keys and authentication data are transmitted in clear	Encryption does not extend far enough towards the core network	Protection against channel hijack relies on encryption	Data integrity is not provided	No HE knowledge or control	2G systems do not have enough flexibility
Network access security							
Enhanced user identity confidentiality							
Mutual authentication between user and network	x					x	
Confidentiality of user and signalling data			x				
Data integrity and data origin authentication of signalling data	x			x	x		
User domain security							
Network element authentication							
Confidentiality of exchanged data		x	x				
Data integrity and data origin authentication of signalling data			x			x	
User domain security							
User-to-USIM authentication				x			
USIM-terminal link				x			
Application security							
Secure messaging					x		
Network-wide user traffic confidentiality			x				
Security visibility and configurability							
Visibility							
Configurability							
Basic assumption for UMTS							x

Note for Tables 1 and 2: The first row contains the complete set of security features defined by the 3G security architecture[3]. Columns that are not linked to rows by crosses indicate 3G security features that either do not exist in 2G systems or that are not needed to correct 2G security weaknesses.

- setting up or not setting-up non-ciphered calls: the user and/or user's HE should be able to control whether the user sets up connections when ciphering is not enabled by the network; and
- accepting/rejecting the use of certain ciphering algorithms: the user and/or user's HE should be able to control which ciphering algorithms are acceptable for use.

Mapping Between 2G Security Weaknesses and 3G Security Features

Tables 1 and 2 show that the new 3G security architecture:

- includes all 2G security elements that have been proved to be needed and robust, and
- addresses all real and perceived 2G security weaknesses.

Conclusion

This paper describes in detail the improvement of the new 3G security related to the well known 2G security. It has been shown that all real and perceived 2G security weaknesses are addressed and corrected by the new 3G security features and mechanisms building together the 3G security architecture.

Outlook

Third-generation mobile radio systems will also offer new service capabilities and features that have probably to be secured by additional security features and mechanisms. At the time of writing these cannot be listed here. However, the environment in which these features are likely to be developed can be described. 3G security will secure this environment[1].

The environment in which new services will be developed can be characterised by (but is not limited to) the following aspects. Third-generation mobile radio systems will involve new and different players, for example, content providers or data service providers. These systems will be positioned as the preferred means of communications for users and be preferable to fixed line systems. Probably there will be a variety of prepaid and pay-as-you-go services which may be the rule rather than the exception. A long-term subscription between the user and a network operator may not be the paradigm. Users will have increased control over their service profiles, which they might manage over the Internet and over the capabilities of their terminal. The terminal may support personal authentication of the user, for example, using biometric methods. Also 3G systems will enable new services and functions to be downloaded using extended 2G functions such as MExE (mobile execution environment) and the SIM application toolkit. It is expected that non-voice services will be as important as, or more important than, voice services. Opening the system to the Internet will enable even more services to be used. The terminals will be used as a platform for e-commerce and other applications. Multi-application smartcards, where the USIM is one application among many, can be used with the terminal. The smartcard and terminal will support environments such as Java to allow this. On the other hand there is a higher risk of active attacks† on users that will have to be addressed in the future.

*References**

1 3G TS 33.120 (v3.0.0). Third Generation Partnership Project (3GPP); Technical Specification Group (TSG) SA; 3G Security; Security Principles and Objectives.

2 3G TS 21.133 (v3.0.0). Third Generation Partnership Project (3GPP); Technical Specification Group (TSG) SA; 3G Security; Security Threats and Requirements.

3 3G TS 33.102 (v3.0.0). Third Generation Partnership Project (3GPP); Technical Specification Group (TSG) SA; 3G Security; Security Architecture.

4 GSM 02.09 (ETS 300 920). Digital Cellular Telecommunications System (Phase 2+); Security Aspects.

5 GSM 03.20 (ETS 300 929). Digital Cellular Telecommunications System (Phase 2+); Security Related Network Functions.

6 GSM 02.48 (TS 101 180). Digital Cellular Telecommunications System (Phase 2+); Security Mechanisms for the SIM Application Toolkit; Stage 1.

7 GSM 03.48 (TS 101 181). Digital Cellular Telecommunications System (Phase 2+); Security Mechanisms for the SIM application toolkit; Stage 2.

† In active attacks, equipment is used to impersonate parts of the network to actively cause lapses in security. In passive attacks, the attacker is outside the system and listens in, hoping security lapses will occur.

* See http://www.etsi.org/ to receive ETSI specifications and http://www.3gpp.org/ to receive 3GPP specifications.

Biography

Stefan Pütz
T-Mobil

Stefan Pütz studied electrical engineering at the University of Siegen, Germany. In 1994, he obtained his diploma. From 1994, he worked on a thesis on 'Non-Deniable Authentication for Future Mobile Radio Systems', carried out at the University of Siegen, Germany. In 1998, he received his Ph.D. in Electrical Engineering. In 1997, he joined Deutsche Telekom MobilNet GmbH, the public network operator in Germany, and now works on the area of systems security in mobile radio networks. Since 1998, he has been a delegate on ETSI SMG10, the group which elaborated the first version of the ETSI description of UMTS security, and in 1999 he was elected as a vice chairman of the 3GPP security group.

Baudouin Boreux

The Need for Inter-Domain Management to Control the Network of the Future

The current (r)evolution in the network utilisation from voice to data, from circuit switching to packet switching affects the design of the whole network and creates a number of challenges for established operators as well as a number of opportunities for new entrants. This revolution in the network directly impacts the domain of network management. The new data technologies need to be integrated in the current management infrastructure and in the telecommunications management network (TMN) architecture. Technical issues are not the whole story as process considerations are at least as important. Many different network technologies need to converge at some higher level to allow coordinated and integrated service management, customer care and billing.

This paper proposes a vision for addressing the challenges the operators face in integrating these multiple technologies: how key drivers in the operators' environment force them move the implementation of an inter-domain management (IDM) function above the specific technological domains like synchronous digital hierarchy (SDH), voice switching, asynchronous transfer mode (ATM), Internet protcol (IP)). The paper explains how the inter-domain management function interfaces with other key domains like accounting management or operaions domains.

Ultimately, the purpose is and remains to be able to offer to customers a wide range of network services with a minimum of complexity and a maximum of comfort while enabling the operator to use its infrastructure to the highest efficiency possible.

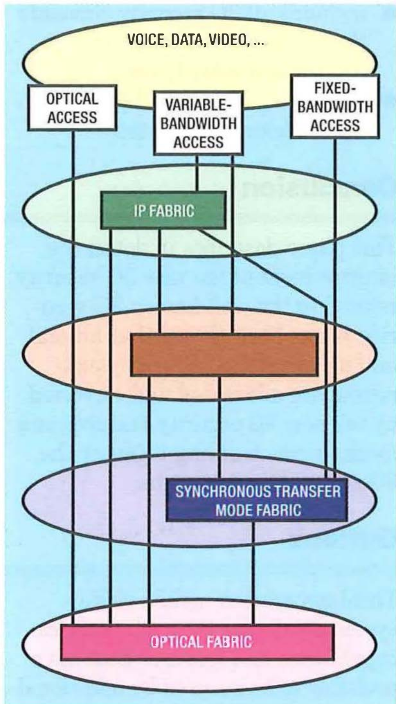


Figure 1 — Multi-technology 'network of networks'

Environment

In this current period of tremendous growth and changes in the telecommunications domain, the evolution and strategies of telecommunications operators are influenced by the main following factors.

- The *regulatory environment* at worldwide, European or country based levels that forces service providers to conform to rules and standards. Examples are the obligation for all operators to offer local number portability capabilities or the complex information regulating the unbundling problem.
- The *evolution of the operators themselves* in their current financial or partnership structures. Mergers, acquisitions, joint ventures heavily impact the way operators evolve and very often create new opportunities in unexplored domains.
- The *competitive pressure* which forces the operator to work on three main dimensions critical to its survival: (1) the control of the cost, (2) the quality of the service provided to the customers, (3) the generation of revenues in current services but more and more in new innovative services.
- The *growing complexity of the technologies* used in the network and the combinations that exist among the new technologies or between the new technologies and the older ones that will remain used for economical reasons. Plesiochronous digital hierarchy (PDH), synchronous digital hierarchy (SDH), asynchronous transfer mode (ATM), Internet protocol (IP), optical, and frame relay technologies will still co-exist for many years. Figure 1 provides examples of the combination of some types of possible technologies.
- *Change in relationship between operator and end customer.* The relationship between service providers and their customers is also evolving. Using software technologies (web technologies, intelligent networks capabilities), customers receive more direct access to the network resources

Baudouin Boreux:
Lucent Technologies Network Systems Belgium
Rue des Deux Gares, 84
B-1070 Brussels, Belgium
Tel: +32 2 556 7217
Fax: +32 2 556 7512
E-mail: bboreux@lucent.com

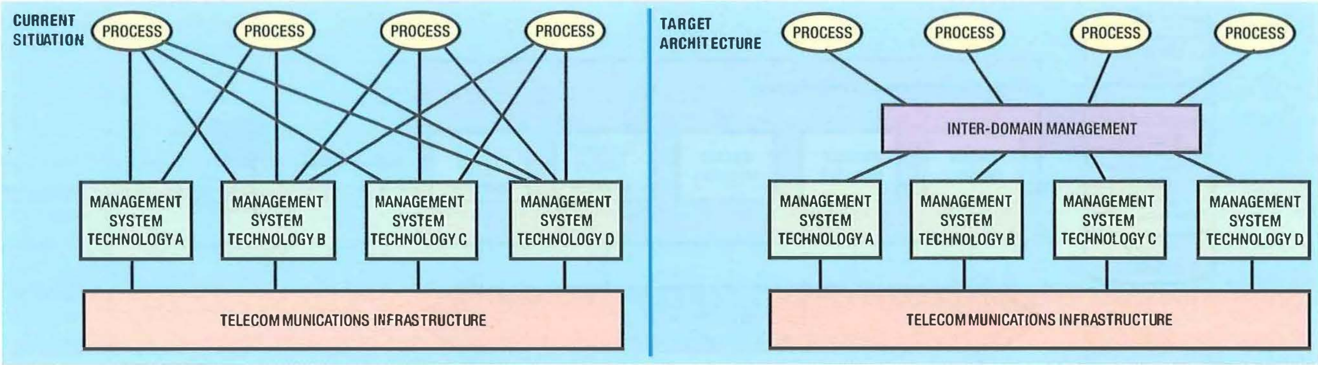


Figure 2—Present versus future mode of operations

they buy from service providers both from a service assurance and resource provisioning standpoint.

In order to face these forces, operators have traditionally organised around three main domains:

- The *accounting domain* involves all the activities relating to the collection of billing information, the generation of invoices, the account receivable domain and all other activities relating to revenue collection and recognition.
- The *operations management domain* provides the skills and tools to design and implement all the processes that manage the services inherent to the life of an operator: customer contact platform, service provisioning, service assurance, planning and capacity management, etc. All these processes are progressively automated in order to provide efficiency gains and reduce redundancy. This domain contains the functions generally performed by enterprise resource planning (ERP) platforms.
- The *network management domain* is directly linked to the infrastructure itself, which has traditionally been handled on a technology-by-technology basis mainly for historical and organisational reasons.

Need for a New Concept

The technology evolution (Figure 1), as well as the need for operators to optimise their operations through completely automated service assurance and flowthrough provisioning processes, make it absolutely necessary to have the relationship between the operations management (ERP) and network management domains as technology independent as possible. Both from an organisational, efficiency and flexibility standpoint, service providers need the network management domain to offer a standard interface to the two other domains whatever technologies these services are implemented on (Figure 2). These considerations force the creation of an additional layer on top of the network management domain. This layer is called the *inter-domain management* layer.

Inter-Domain Management Architecture

The inter-domain management is a management system by which the physical and logical elements of a circuit can be linked as follows:

- to allow the service provider to create a single database which houses current, logical and physical, service/circuit/network design data such that they can make this data available on a near-real-time basis to those fault, performance and traffic management systems they choose to deploy;
- to allow the service provider to design and implement services in a completely automated manner using flowthrough technology;
- to permit the correlation of logical and physical network data, thus yielding a complete end-to-end service/circuit design for those services/circuits which traverse the service providers' hybrid network;
- to allow the service provider to obtain current network data for a given circuit's composition such that maintenance activities can be effected in an expeditious and cost-effective manner;
- allow the service provider physically to display the end-to-end service subscriber's service/network on a near-real-time basis for the purpose of effecting customer service management; and
- to allow the service provider proactively to perform logical capacity management through the establishment of thresholds that when exceeded, result in the automatic creation of an exception report/notice. This capability can be provided as part of the solution itself or it can be provided through the use of an off-line capability which makes use of the data resident within the database proposed.

Figure 3 positions the inter-domain management layer in the global network management structure, in the middle between the domain specific technologies and the service management layer which contains the functions indicated before as the accounting management and operations management functions.

The inter-domain management (IDM) system consists of the following three applications (Figure 3):

- *inter-domain configuration manager*,
- *inter-domain fault manager*, and
- *inter-domain capacity manager*.

Inter-domain configuration manager

The configuration management function is split into the design and the implementation phases which are the key steps in the service activation process. In order for the configuration manager to know the logical structure it is dealing with (Figure 1), it needs a reliable repository of information documenting the logical structure currently in use. This database is called the *logical tree database*. The 'tree' maintains pointers to the domain management systems (this includes the inventory management system) that contain the details of the relevant section capturing the hierarchical relationship.

An example of the logical tree documented in the logical tree

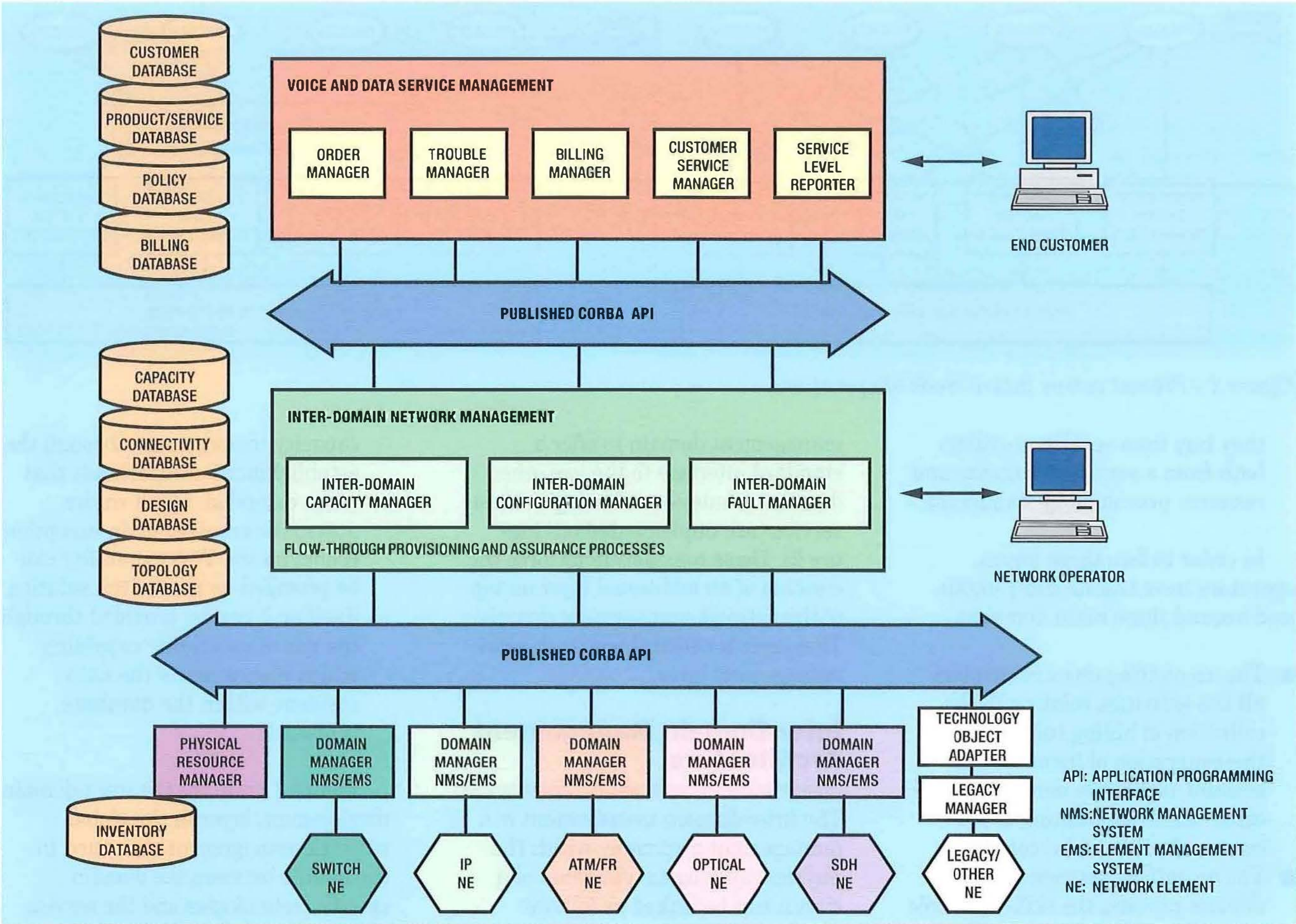


Figure 3—Global network management structure

database is provided in Figure 4. This figure presents an example of a network implementing a combination of SDH and ATM technologies between two locations. For purposes of simplification, only a portion of the network has been shown. The top part of the figure presents the physical implementation and the bottom part indicates the corresponding logical implementation in the tree database with the pointer between the various logical or physical elements.

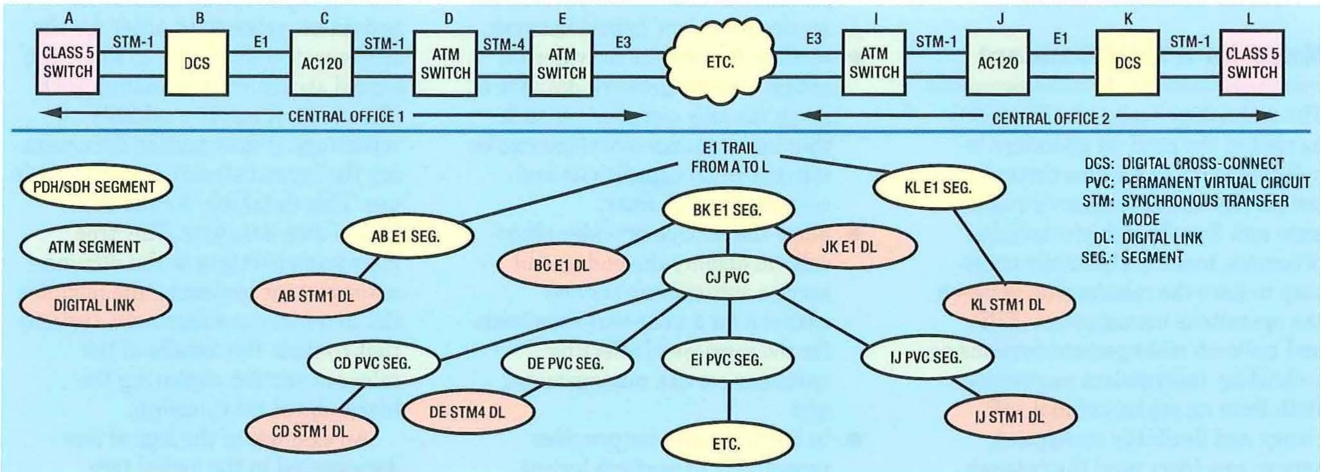
The configuration manager is based on two modules:

- the logical tree manager module, and
- the provisioning manager module.

Inter-domain logical tree manager
The inter-domain logical tree manager provides and maintains a detailed end-to-end view of planned and provisioned transport services and facilities. Its main functions are:

- create and manage an end-to-end service design view and logical hierarchy of facilities supporting the service;
- store the information regarding leased facilities;
- maintain the status of the circuit (pending, implemented, in-effect);
- receive data from the end-to-end design manager and implementation manager;
- service requests for viewing the trail hierarchy information; and

Figure 4 : Sample trail hierarchy



- notify configuration manager the status of changes to the trail's status so that the configuration manager can forward the changes to the 'subscribers' (for example, fault manager).

Inter-domain provisioning manager
This module has the task of supporting seamless provisioning of services/facilities across different technologies as well as multiple domains. It consists of two modules: the design management and implementation management modules.

The design management module:

- performs end-to-end service design across one or more domains of the same or different technology;
- manages inter-domain resources;
- services all requests to create, modify and disconnect intra-domain and inter-domain designs;
- upon successful completion, returns a unique circuit identifier to the order manager, if one has not been provided;
- submits 'pending' designs to the logical tree manager to store in the connectivity database;
- requests port assignments and cable link designs from the physical inventory manager;
- requests and coordinates detailed intra-domain service design from domain management systems; and
- provides sub-activities for high-level design, physical layout, and low-level design.

The implementation management function:

- services all implementation and 'in-effect' requests—intra-domain and inter-domain;
- accesses pending designs from the configuration manager;
- correlates and coordinates implementation and 'in-effect' activities among domain management systems;
- requests trail implementation and 'in-effect' information from domain management systems;
- notifies the physical inventory manager to update cable link and other equipment to 'in-effect' status;
- requests the logical tree manager to update the trail identification status to 'in-effect'; and
- provides sub-activities for implement and in-effect.

Inter-domain fault manager
The inter-domain fault manager collects faults across multiple technologies and domains and determines the root cause domain responsible for the fault. In particular the following functions are supported:

- Creation and management of a network topology database suitable for inter-domain fault correlation.
- Interface with the inter-domain configuration manager to request network information, subscribe to receive updates to the network information, receive network information and create/update the network topology database and finally process network topology information requests from the correlation manager. If the information is not available from the topology database, it requests it from the inter-domain configuration manager.
- Use a correlation management function to:

—interface with domain fault management systems and receive alarms depicting the domain fault manager's view of the service affecting root cause,
—apply the user-defined correlation rules and network topology information to domain root-cause alarms to determine the actual root-cause problem domain,
—correlate the root-cause fault with the effected circuit identification,
—interface with the trouble ticket manager to create a trouble ticket for the root cause fault,
—notify all domains about the root-cause domain,
—optionally receive requests to create trouble tickets on behalf of domain managers for their non-service affecting faults,
—notify the domain manager about the status of their trouble ticket requests along with the trouble ticket identification, and
—service request from the customer service manager for an out-of-service circuit list.

Inter-domain capacity manager
The inter-domain capacity manager enables service providers to perform proactive management of transport capacity across their networks consisting of several technologies. The capacity manager enables providers:

- to interface to the configuration manager to get periodic updates of the new pending and 'in-effect' digital links, facilities, and services;
- to interface to the inventory manager to get spare equipment/ports on network elements for a given service type;
- to provide user-settable threshold crossing alerts on the available service capacity as a percentage of the total between any two service locations over all possible facilities as well as per facility bases;
- to set user-settable threshold crossing alerts on the equipment capacity (that is, number of available slots of a particular kind, etc.);
- to provide pending versus implemented views of the capacity (that is, available bandwidth) on any given date between two service locations or specified facilities;
- to obtain 'what if' view of the capacity between two service locations (that is, if additional facilities are added between two nodes that provide transport between A and Z); and
- to provide notification to the users of capacity threshold crossings; and
- to provide periodic and on-demand reports of the monitored capacity.

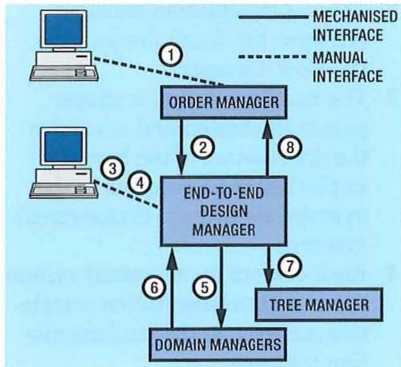
Operational Examples

This scenario illustrates a generic procedure required to fulfil the provisioning request issued by a customer for one of the transport services. It is assumed that the order manager controls the overall execution of the request. The request from the order management perspective is broken down into two steps:

- service design, and
- service implementation.

Service design flow steps (Figure 5)
1 The customer specifies the transport service required. The

Figure 5—Provisioning flow: design



order manager captures the request in the form of a service order and initiates the appropriate provisioning task model.
2 The order manager sends a design request to the end-to-end design manager. The request includes the service identifier (that is, circuit identification), service end points (A, Z), type of service, route requirements (for example, avoidance criteria, protection, etc.), required bandwidth, etc.
3 The end-to-end design manager constructs (manually or automatically) a high-level design, at the domain level, from the 'A' to 'Z' ends of the circuit by specifying end-point equipment, inter-domain links and involved elements.
4 The user approves the high-level design, if not automated, and requests detail domain designs (that is, 'low-level' designs).
5 The end-to-end design manager parses the high-level design and distributes the low-level design requests to the appropriate domain management systems (for example, physical inventory manager, ATM, SDH).
6 Each domain management system forwards the completed design to the end-to-end design manager, including parent facilities supporting the design within the domain.
7. The end-to-end design manager forwards the completed service design to the logical tree manager to be included in the hierarchy and flagged as pending.
8. The end-to-end design manager notifies the order manager of design completion.

Service implementation flow steps (Figure 6):
1 The order manager sends an implementation request to the implementation manager. Data includes service identifier (that is, circuit identification) for the approved design.
2 The implementation manager retrieves the detail design from the view manager.
3 The implementation manager parses the design and schedules the distribution of the low-level implementation requests to the appropriate domain management systems.
4 Each domain management system sends an implementation completion message to the implementation manager.

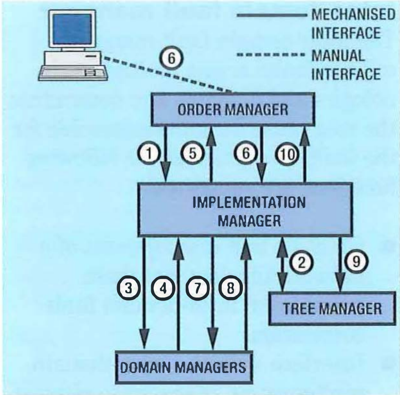


Figure 6—Provisioning flow: implementation

5 Upon completion of all individual domain implementations, the implementation manager notifies the order manager.
6 The order manager sends an 'in-effect' request to the implementation manager after all testing activities have been completed.
7 The implementation manager parses the design and schedules the distribution of the low-level 'in-effect' requests to the appropriate domain management systems.
8 Each domain management system sends 'in-effect' completion messages to the implementation manager.
9 Upon completion by all domain management systems, the implementation manager notifies the logical tree manager to update the trail identification status to 'in-effect' and the physical inventory manager to update the physical inventory used to 'in-effect'.
10 The implementation manager notifies the order manager of 'in-effect' completion.

Conclusion

This paper highlights the operational and business requirements imposed on operators that force them to consider and implement tools allowing service assurance and provisioning capabilities across technology domains. Investments made at operations (ERP) and accounting domains need to rely on technologies in the network management domains that are domain independent. Inter-domain is a powerful concept that is designed to 'hide' the technology dependency to the above software layer and provide the necessary level of abstraction to be able to isolate the investments made at different levels.

The fast-changing environment in the emerging data technologies constitutes a challenge for the emergence of the IDM concept. However, it makes no doubt that such a concept will have to be implemented to cope with the complexity of the network infrastructure. It will require all the power and flexibility of the current software technologies to be successfully implemented in a short time.

Bibliography

1. BAGGA, YUDHVEER S.; and CHAUDHURY, ARATI. Communications Software Management Systems (CSMS) Inter-Domain Solution Architecture. Internal Lucent publication, 28 May 1999.
2. MILLER, PHILIP. Communications Software Transport Management System Solution—Market Requirements. Internal Lucent publication, 17 April 1999.

Biography

Baudouin Boreux
Lucent Technologies

Baudouin Boreux is Business Director in the Communications Software division, Lucent Technologies, for the Europe, Middle East and Africa region. His organisation is responsible for the bid management, product marketing and delivery support of network management products. He has worked in the telecommunications industry since 1987, when he graduated from Louvain-la-Neuve university with a Bachelor's Degree in Telecommunications. While working for Lucent, he also earned a Master in Business and Administration from the same university, In the early 1980s, he was involved in the architecture and systems engineering activities for major network management projects in the EMEA region. He was involved in the specification of complex network management structures involving switching and transmission management systems for both PDH and SDH technologies.

Dionisis X. Adamopoulos, George Pavlou, Constantine A. Papandreou and Emmanuel Manolessos

Distributed Object Platforms in Telecommunications: A Comparison Between DCOM and CORBA

Continuous advances in telecommunications technology coupled with the development of powerful desktop workstations, and increased user demands for complex service provision chains, integrated service offerings, and new sophisticated telecommunications services with multimedia characteristics are fueling the growth of object-oriented distributed computing. Recognising the growing importance of distributed object platforms in telecommunications service engineering, this paper attempts to compare the two most important of them; namely Microsoft's COM/DCOM and OMG's CORBA. After a brief overview of both architectures, a decision framework is proposed by identifying a set of core and service engineering related properties and examining the way that DCOM and CORBA supports these properties. Based on the proposed decision framework, some conclusions are drawn about the suitability of each platform for different service development requirements.

Introduction

There are many driving forces which have compelled telecommunications operators and vendors to seek new solutions in telecommunications service engineering. Among them, growing competition and the progressive convergence of information and telecommunications technologies has led to an increased focus on how a great variety of advanced multimedia telecommunications services (telematic services) with enhanced functionality can be efficiently and effectively developed and deployed in shorter time-frames taking advantage of different network technologies, end-systems, communications protocols, operating systems, and programming language environments. Distributed object platforms have been recognised as a key technology solution to this problem, mainly due to recent developments in object orientation and distributed computing.

Upon these platforms, which are actually object-oriented distributed processing environments (DPEs), telecommunications services are isolated by the underlying computing and networking technology and are realised by a (possibly large) set of interacting service objects/components, which are distributed across different network elements. Currently the two most important available distributed object platforms are Microsoft's (Distributed) Component Object Model (COM/DCOM) and the Common Object Request Broker Architecture (CORBA), which is supported by the Object Management Group (OMG).

This paper examines DCOM and CORBA as the current key general-purpose distributed object-oriented environments. A comparison between them is attempted based on a set of carefully selected core and service engineering related properties, and as a result a decision framework is proposed with the objective to guide service designers/developers during the selection process.

Distributed Objects in Telecommunications

As the telecommunications environment is gradually changing its face towards an open market of information services, it is becoming apparent that major private and public networks are actually large distributed object systems. These systems are populated by a dispersed set of objects that can request services from one another through a communications mechanism, using interfaces defined in a consistent interface definition language (IDL). The most important benefits offered by distrib-

Dionisis X. Adamopoulos:
Centre for Communication Systems Research
School of Elect. Eng. and Mathematics
University of Surrey, England
D.Adamopoulos@ee.surrey.ac.uk
Prof. George Pavlou:
Centre for Communication Systems Research
School of Elect. Eng. and Mathematics
University of Surrey, England
D.Adamopoulos@ee.surrey.ac.uk
Dr. Constantine A. Papandreou:
Hellenic Telecommunications Organisation (OTE)
Athens,
Greece
kospap@org.ote.gr
Emmanuel Manolessos:
IBM Hellas
274 Kifissias Avenue
Athens,
Greece
manoleso@gr.ibm.com

uted object technology to telecommunications systems are ease of development and maintenance, abstraction, modularity, reusability, and granularity flexibility [8, 9, 14].

Distributed object technology is already in use in the telecommunications world[2, 14]. A characteristic example is the telecommunications management network (TMN) development (especially its information model), and an example where this influence is maximised is the telecommunications information networking architecture (TINA), standardised by the TINA Consortium (TINA-C). The main objective of TINA-C is to define and validate an innovative architectural framework (a long-term service architecture) that will address in an integrated manner service control and service management. TINA-C services are considered as a set of distributed computational objects that operate on a distributed object platform[13].

Microsoft's COM/DCOM

COM constitutes the foundation of Microsoft object services and has been assigned to the Open Group for standardisation[10]. Distributed COM (DCOM) is the distributed extension to COM that builds an object remote procedure call (ORPC) layer on top of DCE RPC to support remote objects. DCOM makes COM objects location-independent, and adds security and multithreading to COM[3].

A DCOM server can create object instances of multiple object classes. Each DCOM server object can support multiple interfaces, each representing a different view or behaviour of the object. An interface consists of a set of functionally related methods. A DCOM client interacts with a DCOM object by acquiring a pointer to one of the server object's interfaces. Thus, it invokes the server object's exposed methods through the acquired interface pointer as if the server object resided in the client's address space.

As long as a platform supports COM services, DCOM can be used on that platform. DCOM is extensively supported on the Windows platform. Companies like Software AG provide DCOM service implementations through their EntireX product for Unix, Linux and mainframe platforms, Digital for the Open VMS platform, and Microsoft for the Solaris platform.

OMG's CORBA

CORBA is supported by the Object Management Group (OMG) as part of an initiative to develop a comprehensive object management architecture (OMA) for object-oriented computing[11]. CORBA adopts an object-oriented approach. Object interfaces are described in terms of an implementation language-neutral IDL. CORBA has a special component, called the *Object Request Broker* (ORB), which is responsible for making object distribution transparent and providing a mechanism for trading, enabling object requests to be carried out in a heterogeneous distributed environment.

Besides CORBA, OMA also defines certain key object interfaces. These can be divided into the lower-level CORBA services and the higher-level CORBA facilities.

Table 1 Comparing DCOM and CORBA: Basic characteristics

Basic Characteristics	DCOM	CORBA
Inheritance of a base interface	Every object implements *IUnknown*	Every interface inherits from *CORBA.object*
Unique identification of a remote server object	Through its interface pointer	Through an object reference (objref)
Unique identification of an interface	Using the concept of interface IDs (IIDs)	Using the interface name
Unique identification of a named implementation of a server object	Using the concept of Class IDs (CLSIDs) the mapping of which is found in the registry	By the mapping to a name in the implementation repository
Reference generation of the remote server object	Performed on the wire protocol by the Object Exporter	Performed on the wire protocol by the Object Adapter
Handling of common tasks like object registration, skeleton instantiation, etc.	Either explicitly performed by the server program or handled dynamically by the DCOM run-time system	Performed implicitly by the constructor
Underlying remoting protocol	Object Remote Procedure Call (ORPC)	Internet Inter-ORB Protocol (IIOP)
Activation of a server object	Mainly by using *CoCreateInstance()*	Mainly by binding to a naming or a trader service
Mapping of object name to its implementation	Handled by the registry	Handled by the implementation repository
Storage of type information	Type library	Interface repository
Client side stub	Called a proxy	Called a proxy or stub
Server side stub	Called a stub	Called a skeleton
Definition of parameters passed between the client and server objects	Defined in the interface at the interface definition file. Depending on what the IDL specifies, parameters are passed either by value or by reference	All interface types are passed by reference. All other objects are passed by value including highly complex data types
Definition of complex types	Complex types that will cross interface boundaries must be declared in the IDL	Complex types that will cross interface boundaries must be declared in the IDL
Support of distributed garbage collection	On the wire by a pinging mechanism which garbage collects remote object references and encapsulates them in the *IOXIDResolver* interface	No
Platform support	Any platform as long as there is a COM service implementation for that platform	Any platform as long as there is a CORBA ORB implementation for that platform
Programming language support	Since the specification is at the binary level, diverse programming languages can be used	Since it is just a specification, diverse programming languages can be used, as long as there are ORB libraries suitable for coding in a specific language

CORBA services define such object interfaces as naming, life cycle, persistence, transaction, concurrency control, relationship, time, and security. CORBA facilities provide horizontal and vertical application frameworks, by defining collections of facilities that processes may use through CORBA objects, such as compound documents, user interfaces, and system management[12].

Comparing DCOM and CORBA

Taking into account the basic characteristics of DCOM and CORBA, as were presented in the previous sections, Table 1 represents an initial attempt at comparing the two technologies. This table reveals the wide scope and the richness of both platforms, and it is believed to be much more concise and informative than other comparison attempts found in the literature, which are based extensively on code examples[4, 6]. These code examples, while reasonable and correct, are extremely limited and involve choices among a variety of possible approaches. Furthermore, it is very difficult to keep the two implementations exactly equivalent. Therefore, comparisons based heavily on code examples can be used only as a means to become familiar with DCOM and CORBA, and not as a basis for general conclusions about either technology.

In order to derive such desired conclusions, a set of core properties that should characterise every distributed object platform are identified. Table 2 summarises the way that DCOM and CORBA supports each of these properties, and offers an insight on the capabilities of the platforms pertaining to their use in practical situations. It is explained in more detail in the subsection on 'Core Properties'.

Furthermore, a set of service engineering related properties is also identified, and their support by DCOM and CORBA is summarised in Table 3. This table focuses on how DCOM and CORBA provide a solution for developing effective (possibly large-scale) telematic services, and how they assist in the deployment of these telematic services across the Internet, within an intranet, over an extranet, or simply with a web front end.

During this comparison attempt, the various value-added services provided by DCOM and CORBA are considered. These include, for DCOM, the Microsoft Transaction Server (MTS), the MicroSoft Message Queue server (MSMQ), the Microsoft Cluster Server (MCS), and the Microsoft Management Control (MMC), and for CORBA (2.0), the naming, events, life cycle, persistence, relationship, externalisation, transaction, concurrency, property, licensing, time, trader, and security services (CORBA services)[7, 12]. Table 3 is explained in more detail in the subsection on 'Service Engineering Related Properties'.

Core properties

Object locator

A mechanism by which objects can be located and subsequently activated is necessary. DCOM allows for a locally maintained object locator on the server machine using object names, while CORBA and MTS centralise the locator on a single (or perhaps a few) domain machines that can identify object servers in the domain. More specifically, the most well-known object locator in DCOM is the registry. It maps a CLSID (or a readable name called *ProgID*) to the path name of the server executable that supports the CLSID. However, the registry is consulted only after the Service Control Manager (SCM) has failed to locate any running object instance. In CORBA, an object can locate another object in a system, by using either the naming service or the trader service.

Server activation

A DCOM object server is not necessary to be running when a client request is made to instantiate an object. DCOM locates the server code through the registry, and will start the server using SCM. DCOM also

Table 2 Comparing DCOM and CORBA: Core Properties

Basic Characteristics	COM / DCOM	CORBA
Object Locator	Locally maintained	Centralised in the domain
Server Activation	Yes (Service Control Manager)	Yes (Basic Object Adapter)
Data Typing	Strong and predetermined (vtable method)	Strong and predetermined (Static Interface Invocation, SII)
Dynamic Invocation	Dispatch interface	Dynamic Interface Invocation (DII)
Communication Type	Synchronous, Asynchronous (callback support)	Synchronous, Asynchronous, Deferred Synchronous
Inheritance	Interface, Implementation (containment, aggregation)	Implementation
State Persistence	Yes (2 models)	Yes (Persistence Service)
Load Balancing	No	No
Exceptions Handling	Not directly (error reporting)	Yes (CORBA IDL)
Multithreading	Yes (2 models)	Yes

Table 3 Comparing DCOM and CORBA: Service Engineering Related Properties

Service Engineering Related Properties	COM / DCOM	CORBA
Scalability	MTS, Active Directory Service Interface (Win NT 5.0)	Naming service, Trader service
Reliability	MTS, MCS, MSMQ	Transaction service
Security	Built-in: NT LAN Manager, MTS, MS Crypto API, Authenticode SDK	Platform dependent: 3 security levels (0, 1, 2)
Manageability	MMC	Vendor specific tools, Transaction service
Support for Web-based Telematic Services	ActiveX, MS Active Server Page Technology	JavaScript/Java
Support for Internet/Extranets	Two-factor authentication, Remote Data Service (RDS)	Two-factor authentication, Secure Socket Layer (SSL)
Support for Intranets	Desktop tools, ActiveX, Active Data Objects, MSMQ	Desktop tools (via a bridge), Event service, Persistence service

allows access to servers that are already running when the client request is made, as running objects are registered with the Running Object Table (ROT). In CORBA, server activation is handled by the Basic Object Adapter (BOA). If a client makes a request for an object that is not running, then the BOA finds the server and launches it to create the object. In this process it uses the implementation repository, which holds information about the location of every server object.

Data typing
Once an object has been located and activated, the client will need to be able to communicate with it. Strong data typing is supported by both DCOM and CORBA through the use of interfaces. In the static method of invoking operations on DCOM objects, the MIDL compiler, based on the IDL definition of the object and its interfaces, creates the corresponding proxy and stub code. Due to the way that the static invocation is implemented, this is often referred to as the *vtable* method for invoking objects. In CORBA, in the case of the static interface invocation (SII), all methods are specified in advance and are known to the client and server through the stubs and skeletons that are produced by the IDL compiler.

Dynamic invocation
Although strong and predetermined data typing is extremely important for the creation of complex and robust code, sometimes the flexibility of slightly looser typing, similar to the kind which is important to interpreted scripting languages, is necessary. This can be allowed by providing dynamic querying of objects for the functionality that they support. DCOM provides this facility through its IDispatch interface, and CORBA through its DII mechanism. In essence, predetermined typed interfaces are used that allow a dynamic interface to be queried[15].

Communication type
The communication between objects can be either synchronous or asynchronous. DCOM is mainly synchronous. However, it allows for flexible callback mechanisms, such as connection points, to be implemented. In CORBA, a client can invoke a method, either synchronously or post it asynchronously. Posting means that the calling object is not blocked waiting for the reply. Instead, it can specify which of its methods the response should invoke. It has to be noted that the receiver cannot tell the difference between a synchronous or an asynchronous call.

Inheritance
DCOM allows interface inheritance, whereas CORBA allows implementation inheritance. In interface inheritance when one interface is derived from another, the derived interface must supply an implementation for the methods of the base interface; all it inherits is the responsibility to supply the interface. In implementation inheritance a derived interface inherits the interface and an implementation. DCOM provides a similar mechanism using containment or aggregation[7].

State persistence
Objects represent both functionality and data. A client wishing to access an object would typically create the object, access its services, and then destroy it. The object server needs to be able to associate a client connection with a particular object, since each client has some assumption about the state of the object when it last accessed it.

Both DCOM and CORBA use the notion of saving object state for later reactivation. DCOM has two persistence models. The original model requires that objects implement an interface that supports persistence using one of several known storage media (file, stream or storage). A more recent persistence model in MTS provides server-managed storage. In CORBA, the persistence mechanism is completely transparent (persistence service). The client has no legitimate means of determining where or how an object is stored (unless some object with knowledge of the storage details provides an interface with a method that divulges the information). The implementation is exclusively responsible for managing persistence[12].

Load balancing
A server machine may provide several object servers, and each of these may provide several object types. Thus, the server machine may become a bottleneck in the distribution of objects, and this leads to the need for load balancing. This facility is not offered currently, neither in DCOM, nor in CORBA. Generally, load balancing is an area that has little available support in the mainstream distribution framework at present, but intensive development is currently underway[15].

Exceptions handling
DCOM has a standard way of handling error data through the return of a 32 bit error code, called an *HRESULT*, by all methods. At the language/tool level, a set of conventions and system provided services (the IErrorInfo object) allows failure HRESULTs to be converted into exceptions in a way natural to the programming language. On the other hand, CORBA specifies an extensible exception capability that maps naturally into languages that have native exceptions, like C++ and Java, and that maps into exception data in languages that do not. It is based on user-defined exception types declared in CORBA IDL.

Multithreading
DCOM supports multithreaded server objects, but it requires that the DCOM libraries be initialised in the threads that use them. There are two main models (the *apartment* model and the *free threading* model). A third model (still to be released) is called the *rental* model. In this model, which will be used by the MTS, one thread 'helps' another, in a fashion that still behaves as if the object is single threaded. CORBA object servers can also be multithreaded. Issues such as (for example) whether an object is created in a new process or in a new thread are handled by the ORB through the object adapter.

Service engineering related properties

Scalability
MTS provides a set of DCOM interfaces and libraries that allow telematic services to easily scale as the number of users and user data increase. With the forthcoming Windows NT 5.0, DCOM will obtain the Active Directory Service Interface (ADSI), which will allow components to seamlessly use a variety of existing naming services, such as NetWare Directory Service (NDS), Lightweight Directory Access Protocol (LDAP) or even the Windows registry. In that way, a telematic service will be able to handle an increasing number of geographically dispersed users.

In CORBA, the Object Activation Daemon (OAD) and the implementation repository allow efficient use of resources by only instantiating objects when required. The centralised naming service provides location independence for applications and their users, while the trader service allows more sophisticated component searches. Static load-balancing among replicas of an application is available via the naming service.

Reliability
Distributed objects should offer transparency to a client, and part of this transparency is the guarantee that the object connection will be reliable throughout the client's use of an object. In the optimum case, a remote object must be as reliable as a local object.

Reliability can be achieved by using a transaction monitor, like MTS or the OMG transaction service. MTS allows telematic services to use distributed transactions to reliably update data across disparate data stores, while the OMG transaction service supports an Open Group Distributed Transaction Processing (DTP)-compliant model for distributed transactions. As far as DCOM is concerned MCS and MSMQ also increase reliability.

Security
DCOM has been designed with security built in, while CORBA objects can implement their own security mechanisms for the platform on which they are implemented. More specifically in DCOM, the NT LAN Manager (NTLM) and the MTS authenticate users and authorise checking via Access Control Lists (ACLs). Additionally, the MS Crypto API provides data encryption and integrity to prevent eavesdropping and tampering, while the Authenticode SDK uses digital signatures to provide non-repudiation. The challenge for the future is to integrate all these value-added services in a single solution.

On the other hand, CORBA defines two security levels; Level 1 and Level 2. Level 1 allows a telematic service that is unaware of security to participate in a secure domain. It provides user authentication, authorisation via ACLs, data encryption and integrity, and optional non-repudiation. Level 2 requires telematic services to be security-aware, thus enforcing stronger versions of the security policies. Some CORBA vendors, such as Iona and Inprise, have provided a Secure Socket Layer (SSL) implementation of IIOP, called *Level 0*, that allows user authentication and data encryption.

Manageability
Microsoft Management Console (MMC) provides a unified GUI for managing MTS and MSMQ based components. Features include centralised configuration and administration, as well as remote deployment of components. As far as CORBA is concerned, both Inprise and Iona have sophisticated tools for centrally configuring and administering CORBA applications. Iona also allows CORBA applications to be centrally managed from any SNMP-compliant system management console (for example, OpenView).

Support for web-based telematic services
A telematic service is web-based when its front end (or presentation layer) is a web browser, and it does not necessarily mean that the telematic service is deployed over the Internet[5]. In this case, DCOM front ends in the form of ActiveX controls can execute within Internet Explorer and, via a plug-in, within Netscape Navigator. Furthermore, Microsoft's Active Server Page technology allows the seamless integration of both HTML and ActiveX clients with DCOM servers. It also allows DCOM services such as MTS and MSMQ to be used with web-based telematic services. On the other hand, Java-based front ends to CORBA telematic services can execute on all major browsers and platforms, and the Netscape Enterprise Server provides the Web Application Interface, which allows HTTP-based clients to communicate with CORBA servers.

Support for Internet/extranets
Telematic services that need to operate across the public Internet or the semi-public extranet are typically deployed across great distances and often through several firewalls. Such telematic services, when they involve transactions executed over the Internet, require additional security measures to ensure accuracy, confidentiality, and credibility.

DCOM provides two-factor authentication (through public certificate and smart cards), and Remote Data Service (RDS) support for Internet/extranet applications. Currently, there is limited support for SSL and its integration with NTLM security. However, this will be expanded to include SSL-to-Kerberos integration in Windows NT 5.0. On the other hand, several CORBA vendors support both SSL and two-factor authentication in their implementation, although these features are still immature.

Support for intranets
Telematic services which are limited inside an intranet are usually optimised for use within an organisation, have higher network bandwidth, and little or no firewall restrictions. Therefore, they can be built with more sophisticated front ends, both in terms of user interface and functionality.

DCOM-based telematic services can be built with sophisticated user interface and functional features, as every major development environment on Windows supports the rapid development of graphical DCOM applications. Furthermore, Active Data Objects allows the support of persistence, while MSMQ provides publish/subscribe capabilities. On the other hand, CORBA-based telematic services can be integrated with desktop tools using a DCOM/CORBA bridge available from several CORBA vendors. CORBA services such as event and persistence can also be used to add publish/subscribe and persistence features.

Comparison remarks
From Table 1, Table 2, and Table 3, which collectively constitute a decision framework supporting the selection between DCOM and CORBA, it is evident that DCOM and CORBA have similar architectures as both provide the infrastructure for supporting remote object activation and remote method invocation in a client-transparent way. They adopt a client/server based programming style and agree on the most fundamental aspects of their object models.

As far as the support of service engineering related properties is concerned, DCOM and CORBA differ in many respects. Most significantly, while DCOM provides a rich set of tools and technologies, it is essentially a Windows-only solution. Even though DCOM is available on other operating systems, key pieces such as MTS, MSMQ and MCS are not currently offered. Additionally, many of DCOM's value-added services are very new and are still maturing. On the other hand, CORBA's main strength, which is its availability from different vendors, is also its biggest weakness. Since no vendor has a complete solution, integration issues are usually introduced when CORBA is used to build telematic services. For this reason, neither technology provides a complete solution for service engineering activities. However, both provide a solid infrastructure, and there are specific scenarios in which each excels over the other.

Another important remark that has to be stressed is that DCOM and CORBA have a comparable performance. However, DCOM's performance can be improved in certain circumstances by extending its remoting architecture (that is, the entire infrastructure that connects clients to server objects). This is possible because DCOM's remoting architecture has built-in extensibility. By supporting a mechanism called *custom marshalling*, DCOM allows a server object to bypass the standard remoting architecture and construct a custom one, optimised for a particular situation, without requiring source code modifications to the former[7].

In general, the proposed decision framework makes it evident that DCOM and CORBA have much in common and continue to converge in several aspects. However, each architecture has different origins, with consequent strengths and weaknesses.

Interoperability Between DCOM and CORBA

Because both DCOM and CORBA are being used in practice with considerable success, and because of the economic implications that result from this fact, it is unlikely that one platform will soon overwhelm the other. Therefore, interoperability between DCOM and CORBA is crucial[5].

Since CORBA 2.1, the interoperability of DCOM and CORBA is part of the CORBA specification. More specifically, bridges receive object invocations from a CORBA application, translate them into equivalent data structures for DCOM, and have the function call executed in the DCOM application. In a similar manner, DCOM clients can access CORBA objects through bridges[14].

Conclusions

Advances in distributed object platforms have been rapid in the past few years. These advances have been largely driven by increasing demand for efficient object creation, interaction, management, and distribution. Both DCOM and CORBA address these issues, and are increasingly being used to develop new telecommunications services as distributed object applications. However, further progress is expected, and as both technologies are still evolving, it is likely that in the near future they will converge in more areas.

DCOM is built on a proven desktop component architecture. DCOM-based applications are robust and perform well, while DCOM's integration into development languages and tools greatly simplifies application development. Furthermore, as Windows-based desktop systems exist in nearly all organisations today, these organisations will probably choose to use DCOM. Additionally, Microsoft services (MTS, MSMQ, and other mainframe integration tools) make DCOM an attractive infrastructure even for large organisations in enterprise-wide applications. However, DCOM is not a well-partitioned architecture and relies on a key optimisation for a single platform.

In contrast, CORBA has a more complete and well-defined architecture and provides a better solution for heterogeneous environments. It offers advantages in (value-added) services, platform and tool support, maturity, and overall architectural integrity. Furthermore, OMG IDL ensures an extensible architecture and support for both new and legacy applications. The disadvantages of CORBA are its complexity and variation in vendor implementations.

Therefore, DCOM is an effective solution for the development of telematic services in Windows-based environments, particularly by small organisations and departments. On the other side, a requirement for multi-platform support or for a choice with the least technological risk will drive an organisation towards a CORBA solution. However, such a decision will (should) be highly influenced by more general factors, such as the available information technology (IT) resources and skills, the IT structural characteristics and its relation to business units of the organisation, the desired level of standardisation, and the capability to adopt new technologies. It has to be noted that Java is also a candidate for the development of telematic services, but not the most prominent as a lot of the transparencies built on the core object models of OMA and COM/DCOM are yet to be defined for Java. The significance of Java is expected to rise rapidly as its standardisation proceeds[1].

It is envisaged that in the near future CORBA and COM/DCOM will interoperate via a standardised single two-way gateway specification (a bridge) between them[5]. However, with dissimilar object models, components will not collaborate as effectively across the gateway between the two environments as they can within each of them. For this reason, the service designer/developer will always consider a choice between CORBA and OLE/COM on the client, and between CORBA and COM/DCOM on the server. The decision framework proposed in this paper is expected to significantly assist the selection process.

Developing new telecommunications services using distributed object technology presents many challenges and alternatives. The correct choice is never at either end of the spectrum, but falls somewhere in the continuum that lies in between. Where it falls depends on both the user/customer (business) requirements that have to be satisfied, and the technical problems that have to be solved.

References

1 Adamopoulos, D. X.; and Papandreou, C. A. Distributed Processing Support for New Telecommunications Services. Proceedings of the IEEE/IEE International Conference on Telecommunications (ICT '98), Chalkidiki, Greece, 1998, Vol. III, pp. 306–310.

2 Aidarous, S.; and Plevyak, T. (Eds.), Telecommunications Network Management. IEEE Press, 1998.

3 Brown, N.; and Kindel, C. Distributed Component Object Model Protocol – DCOM/1.0. Microsoft Corporation, Nov. 1996.

4 Chung, P. E.; Huang, Y.; and Yajnik, S. DCOM and CORBA Side by Side, Step by Step, and Layer by Layer. Bell Laboratories, Murray Hill, NJ, 1998. http://www.bell-labs.com/~emerald/dcom_corba/Paper.html

5 Dolgicer, M.; and Fischer, P. DCOM, Active X and CORBA Must Live Together. Application

Development Trends, April 1997, pp. 38–52.
6 GOPALAN, S. R. A Detailed Comparison of CORBA, DCOM and Java/RMI. Sept. 1998. http://www.execpc.com/~gopalan/misc/compare.html
7 GRIMES, R. Professional DCOM Programming. Wrox Press, 1997.
8 KRIEGER, D.; and ADLER, M. The Emergence of Distributed Component Platforms. IEEE Computer, **31**(3), March 1998, pp. 43–53.
9 LEWANDOWSKI, S. M. Frameworks for Component-Based Client/Server Computing. *ACM Computing Surveys*, March 1998, **30**(1), March 1998, pp. 1–27.
10 Microsoft Corporation and Digital Equipment Corporation. The Component Object Model Specification. Draft Version 0.9, Oct. 1995.
11 Object Management Group. The Common Object Request Broker: Architecture and Specification. Revision 2.0 July 1995 http://www.omg.org/library/o2indx.html
12 ORFALI, R.; HARKEY, D.; and EDWARDS, J. Instant CORBA. John Wiley & Sons, 1997.
13 PROZELLER, P. TINA and the Software Infrastructure of the Telecom Network of the Future. *Journal of Networks and Systems Management*, Dec. 1997, **5**, pp. 393–410.
14 REDLICH, J.-P.; SUZUKI, M.; and WEINSTEIN, S. Distributed Object Technology for Networking. *IEEE Communications Magazine*, Oct. 1998, **36**(10), pp. 100–111.
15 VINOSKI, S. CORBA: Integrating Diverse Applications Within Distributed Heterogeneous Environments. *IEEE Communications*, Feb. 1997, **14**(2), pp. 46–55.

Biographies

Dionisis X. Adamopoulos
University of Surrey

Dionisis X. Adamopoulos holds a degree in Computer Science from the Athens University of Economics and Business, and a Masters degree in Telematics with distinction from the department of Electronic and Electrical Engineering of the University of Surrey. Currently, he is involved in Ph.D. research at the Centre for Communication Systems Research (CCSR) of the University of Surrey. His research interests include service engineering, distributed multimedia, object-oriented analysis and design, telematic services, distributed object platforms, groupware, and telecommunications management.

Dr. Constantine A. Papandreou
Hellenic Telecommunications Organisation (OTE)

Dr. Constantine A. Papandreou holds a Dipl. Ing. degree as well as a postgraduate degree in Engineering Economics both from the Technical University of Munich. He also holds a Doctor degree in Telematics from the University of Munich. He has served since 1970 in the Hellenic Telecommunications Organisation (OTE) as an expert in telecommunications and informatics in various positions. Since 1992, Dr. Papandreou has been a Director in OTE. In parallel to these activities, for several years he has been teaching teleinformatics and information technology at the Athens University of Economics and Business, the University of Piraeus, and the Higher School of Telecommunications of OTE. He is the author of over 60 scientific papers pertaining to his research activity in the fields of telematics, information systems, telecommunications policy, education and training, multimedia, office automation, service engineering, etc.

Prof. George Pavlou
University of Surrey

Prof. George Pavlou received his Diploma in Electrical and Mechanical Engineering from the National Technical University of Athens, and his M.Sc. and Ph.D. in Computer Science, both from University College London. Over the past 12 years he has been undertaking and directing research in the areas of data communications and telecommunications with emphasis on performance evaluation, network and service control and management and the use of distributed object-oriented technologies in new telecommunications architectures. He has contributed to ISO, ITU-T, NMF/TMF, OMG and TINA standardisation work. He has published around 50 papers in international refereed conferences and journals and he is the co-author of two books. He is currently professor of information networking at the University of Surrey, School of Electrical Engineering and Information Technology, where he leads the activities of the networks research group.

Emmanuel Manolessos
IBM Hellas

Emmanuel Manolessos holds a degree in Computer Science from the Athens University of Economics and Business, and a Masters degree in Advanced Information Technology from the department of Computer Science of the Imperial College, University of London. Currently, he is a SAP consultant at IBM Hellas. His research interests include distributed-processing environments, interactive teletraining, enterprise resource planning, and management information systems.

Vittorio Gelmi

Preventive Maintenance: Using Data Mining Systems

The collection of alarms from a network is beneficial when a failure in the system is permanent and it is therefore possible to identify the failure itself. However, often temporary failures are reported by the alarms, but the network soon comes back to working conditions. These chronic failures can be understood only by analysing possible correlation with other events, on the basis of synchronicity, logical sequence, etc. It is impossible to define a priori how to search properly for this correlation, but an artificial intelligence system can be used methodically to help the operator. This paper describes how an intelligent data mining system can be designed and its main features.

As many systems control different elements of the network and different databases can hold part of the overall information needed to identify the correlation, the intelligent data mining system must be equipped to interact with these external systems. To optimise the process (permitting parallel activities on different platforms) this interaction is carried out by mobile agents that are sent to the different systems, locally collect and elaborate the necessary information and come back to the intelligent data mining system bringing the required results.

Introduction

Fault analysis in large telecommunications networks usually involves handling a very large amount of alarms data from different subsystems. When a fault becomes stable, usually a correlation among detected alarms can be detected and repair of the fault is feasible. If a fault appears just for a moment, disappears later and (often after hours or days) comes back again, the fault analysis subsystem often cannot isolate the proper cause of this *chronic* fault[3]. A large number of alarm records crowds the database which are not usable. The operator tries unsuccesfully to resolve the problem, analysing alarms in many ways: istograms, time correlation etc. However, because of the large number and variety of subsystems providing data, a systematic analysis is quite difficult.

Looking at telecommunications applications in network management and fault monitoring, some common requirements can be identified.

Vittorio Gelmi:
Sirti S.p.A. Systems Division Cassina De Pecchi (MI 20060 ITALY).
E-mail: V.Gelmi@sirti.it

Generally applications related to fault analysis must:

- interface heterogeneous sources of data (databases for network configuration, alarms, SNMP agent, custom applications etc.),
- collect data from remote hosts with different procedures,
- drive the user (not just 'help') to perform a complex task, and
- coordinate many synchronous actions/tasks on databases, to obtain reports and information.

To support efficiently the localisation of equipment subject to chronic failure and to isolate area(s) of degradation (to enable preventive maintenance to be carried out), we tested a prototype system which provides:

- intelligent assistance to the operator in searching for the cause of problems, and
- a datamining facility to dig into alarm subsystems to correlate and isolate cause(s) of faults.

Elements of the Problem

The analysis of chronic faults is based on some elements of classification of fault type as described in the following sections.

Fault classification

Starting from the basics, a fault in a telecommunications network causes alarms to rise. Faults can be classified in three types:

- 'Steady'—fault is persistent, and its effects do not change with time. A relatively small number of alarm records is collected (indeed only one 'start time' and 'end time' record). For example, a break in a (not backed-up) power-supply subsystem of a transmission system causes (usually) a steady fault. A steady fault is often related to a hardware fault (and generally can be automatically isolated).
- 'Transient'— fault produces symptoms which change with time. A number of causes can produce transient alarms in telecommunications networks; for example, thunderbolts (if not actually causing a fault to equipment) can temporarly blind some links or increase trasmission line errors dramatically. Usually (but not necessarily) a steady fault is announced by a transient phase during which a large number of alarm records are collected (sometime expressively defined *ripples*). After this phase, all is silent and steady state is reached. Hardware faults are (often) sources of this kind of transient faults.
- 'Chronic'—fault initially appears to have symptoms similar to a transient fault; after some time symptoms disappear too. Later (and usually without visible correlation with the previous event) alarms appear again and continue in an impredictable (usually not periodic) sequence. For example, if a ground connection of a transmission equipment's shield has high resistance, the equipment may be

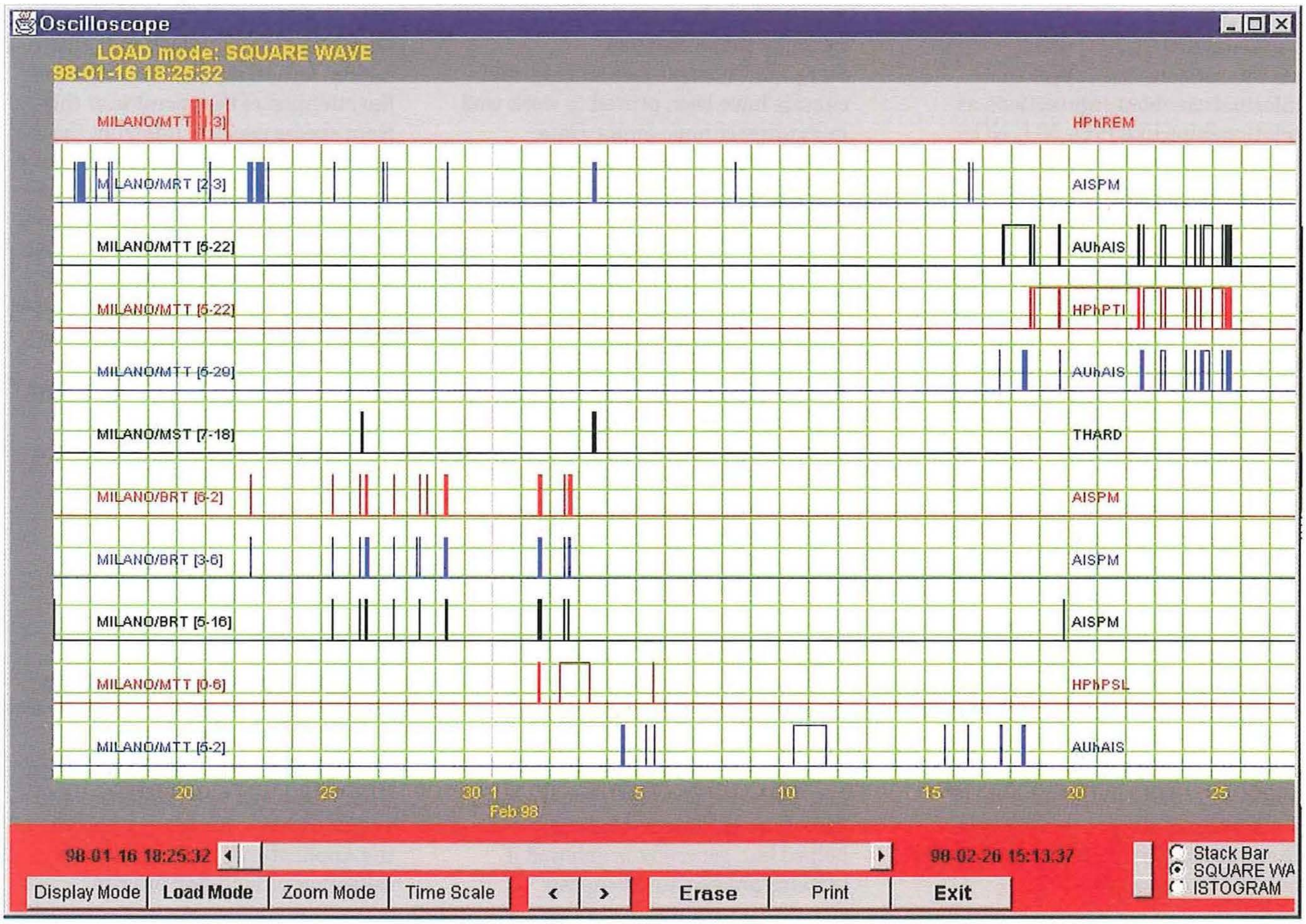


Figure 1 – Isochronous alarm patterns in PDH and SDH links. Some isochronism is justified by interconnections

exposed to random electromagnetic influences from other equipments close to it. In this situation alarms are created as the electromagnetic influence exceeds (randomly) some level. Another example is if some hardware is (improperly) working at its thermal limit, as the external temperature grows some reversible malfunctions can appear/disappear creating alarms in the transmission stream.

Alarms classifications

Alarms too can be divided into two categories:

- *source* or *primary* originated by the faulty equipment; for example, a power supply fault creates primary alarms; and
- *derived* or *secondary* risen as a consequence of electrical connections among the elements of a telecommunications system; for example, a transmitter fault in a transmission-chain propagates derived signal alarms along the entire chain.

Maintenance issues

Chronic faults generally are not properly isolated and removed. This may cause progressive degradation of some areas of the telecommunications network thus increasing unavailability time.

The target should be to isolate chronic faults, properly investigate their cause, locate equipment and repair it before faults can reappear. Removing causes is more than repairing a faulty device. Often causes are not only damaged network objects (NOs) (in the sense that they are in network databases or management information bases (MIBs)). Causes of chronic alarms may not be associated with the telecommunications equipment†; for example, poor installations, etc. Experience has demonstrated that the idea of finding in databases or MIBs one or more device(s) responsible for the problem is erroneous. It is not adequate to limit the search only to alarms databases, layout databases and in general all that describes the network: often the search will stop without a clear identification of causes, or a (possibly small) set of NOs is indicated. The human operator (or a set of heuristics about possible causes or classes of causes derived from experience) will complete the search including proper 'external' causes or relations.

Chronic Faults Search Methodology

A prototype has been built to investigate plesiochronous digital hierarchy (PDH) and synchronous digital hierarchy (SDH) circuits and links. For these NOs a way to investigate chronic faults is to analyse signal alarms and some quality indicators*. (See Figure 1.) Signal-alarm records are explored to identify equipment with sufficient grade of isochronicity in alarm duration and start time. Usually alarms patterns do not exactly match: an algorithm used to compare each pattern with others needs, to work well, to be improved with some heuristics.

Derived alarms and (unknown) primary alarms are expected to exhibit very similar patterns. During analysis, network objects which report isochronous alarms are grouped in disjointed *isochronous sets*.

† For example, in a real case signals from a radar installation (not telecommunications related) some miles distant from the telecommunications office has been found to be the cause of chronic faults.

* For example: unavailable seconds (UAS), severely errored seconds (SES) etc. counted hourly.

An NO_A can generally propagate a derived alarm to NO_B due to reciprocal interactions: we collect this information about interactions as relationships like $Pr(A,NO_A,NO_B)$ where A is the alarm type. Elements which support Pr() are grouped into *connected* sets. Pr() relationships can be discovered from the layout of circuits, interconnections, databases and rules specific for each NO. Note that for this analysis we really do not need to know which is the primary alarmed NO in the isochronous set.

Each isochronous set is then decomposed into a union of connected sets. Intersections among connected sets in an isochronous set are NO investigated for chronicity[3]. Indeed field experiments demonstrate that often the network operator can easily locate the faulty device at this point. Some aspects (not described here) make the problem more complex; for example, some external events like a substitution of a piece of equipment, the removal of the cause of a chronic alarm internally to the interval of analysis (not always recognisable from telecommunications databases) need to be considered.

Analysing link and circuit quality parameters

Some quality measures on links and circuits have been proved to work well in locating chronic faults. Data collected on an hourly base in SDH (155 Mbit/s) links or PDH (140 Mbit/s) links or circuits about SES, ES etc. can be treated almost in the same way as alarms. Alarms are merely on-off events (that is, square waves), but SES, ES measures appear as istograms placed along the time axis. Some differences occur in recognising isochronous sets but basically the idea is the same. Despite the poor temporal resolution (data are collected usually on an hourly basis) analysis appears to be much more sensitive in warning about degraded network areas and NOs subject to chronic fault (Figure 2).

Assistance to the Operator

The process for discovering chronic fault in long-distance circuits is a sequence of analysis mainly guided by heuristics and operator's knowledge of the network. The operator needs to be helped but, generally, is annoyed if systems proceed autonomously impeding any deviation from his/her reasoning scheme. Sometimes an event (alarm, layout, relation etc.) captures the attention of the operator, at this point she/he needs to interrupt the system to follow its reasoning sequence.

The operator is very annoyed too when requested to inform the system about (partial) conclusions of his/her analysis. Partial conclusions have been reached just interacting (by user interface with istograms, selections using mouse etc.) with the system itself: the system should 'remember' all that it supplied to the operator.

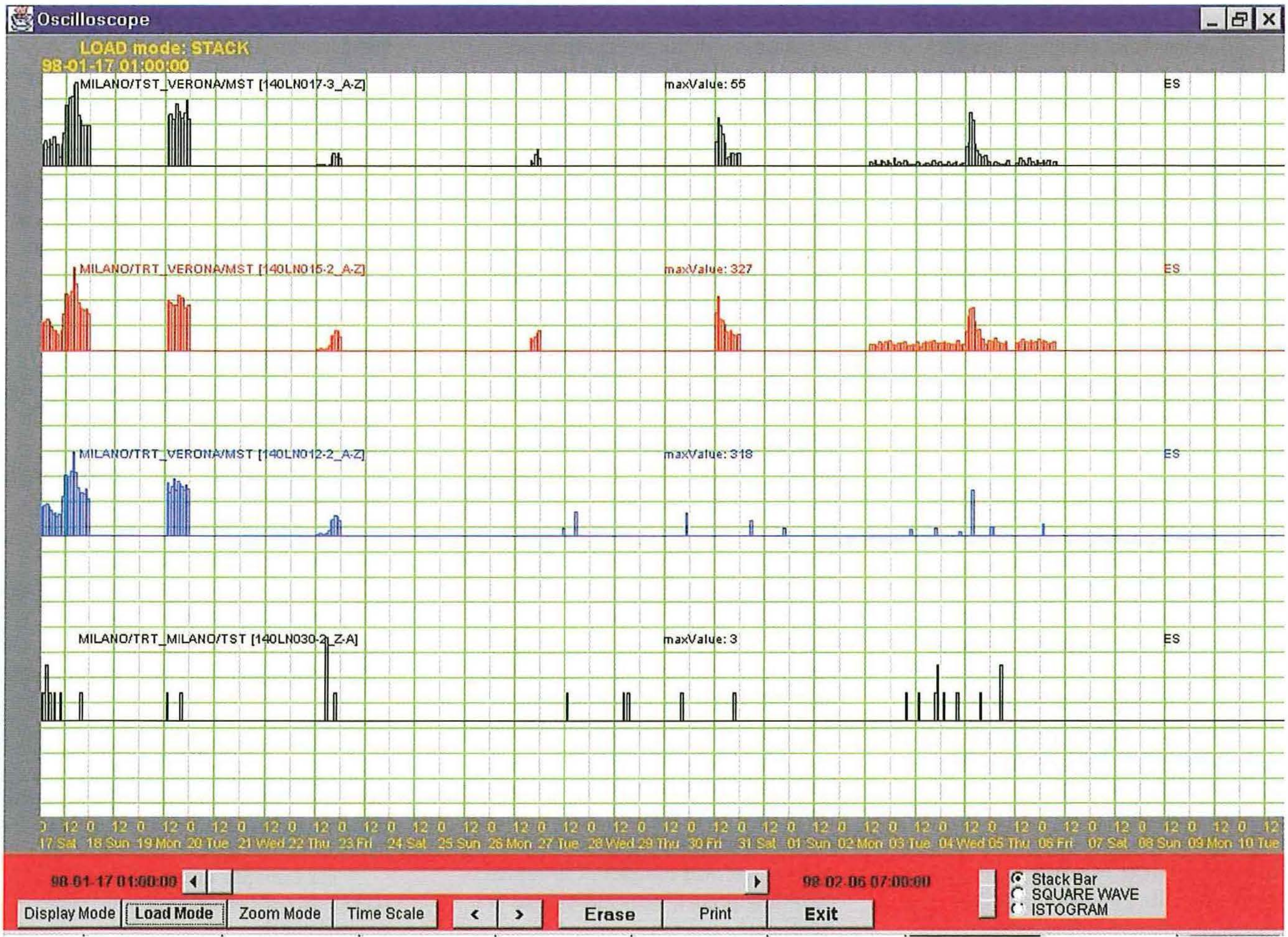
To satisfy these requirements, a three-part architecture has been used including:

- user interface (the client),
- application server, and
- an assistant.

Mining System Structure

Due to the great diversity of telecommunications subsystems from which data are to be collected and analysed, the overall architecture of the application is structured in a collection of modular plug-in

Figure 2—ES measures reveal isochronous patterns. Disturbed links are all carried by a 565 Mbit/s mux



components over a minimal background structure.

To facilitate interaction with remote applications, a mobile agent structure has been tested[5, 6, 7].

The application has some core components:

- *Mobile code server(s) (MCS)*—one or more on each host that supports applications for alarm collection, circuits or equipment interconnection, etc. This is a Java module responsible for receiving the mobile agent code and running it.
- *Mobile agents*—a module that interacts with remote systems and extracts relations, data etc. as requested.
- *Assistant*—a module including a knowledge base devoted to helping the operator or acting autonomously in a data-mining sequence.
- *User interface.*
- *Portable instruments*—modules with some defined interface usable by mobile agents or a static application to perform a specialised task. Portable instruments could be provided by MCS. Each agent has various portable instruments to easily interface local agents (that is, SNMP agents or custom agents) or the data source and to perform its task. Each portable instrument supports a common interface and can be added to the agent just at the start time and destroyed as used†. Portable instruments can be loaded at the arrival time of the incoming agent as convenient. Portable instruments can also be sent after the arrival of the agent in-loco as a consequence of a request message sent to the control centre.

Functionality

Operations are various and generally not in a fixed sequence. Generally an operator starts by investigating some suspect events. Either at the operator's request or the assistant's, agents are launched to the appropriate host to perform the requested job. As the agent returns, usually it has the answers. A search can be driven by the operator, automatically or jointly.

Mobile Agents

The system should be capable of collecting data from different sources, preprocessing different formats and processing data to support user's (or the assistant's) requirements. To perform this task we used a framework to build mobile agents.

A mobile agent is a software code that can be moved among different hosts (Figure 3). The agent's binary code is platform independent. The same code moves from one host to the other and there performs its task.

Instead of creating a monolithic structure to obtain data from remote databases, process them and extract knowledge, autonomous mobile agents are sent by the application server (or the assistant) to the remote host with a workplan to be performed in-loco. Mobile agents in our prototype are built in Java.

Some general benefits in using mobile agents

- *Efficiency savings* CPU run time is not of the application, but of the agent's host.
- *Space savings* Each agent resides on one node at a time. Mobile agents carry the functionality with them which gives a high degree of customisation.
- *Reduction of network bandwidth/network traffic requirements* The transfer of the mobile agents along the network creates less traffic than transferring the data. The agent is not connected to the application and works standalone.
- *Heterogeneous environment and applications glue* Every system can be targeted by the agent at the price of a small interfacc and no change to agent structure and code. This results in fast and cheap customising of new applications.
- *Easy software upgrades* The mobile agent can be replaced without change of remaining application's code.

Agents communication

Each agent in the network performs a message-passing facility to exchange data and requests. This structure is used to develop a protocol for interaction from the user interface to the server(s) and assistant.

Database access facility

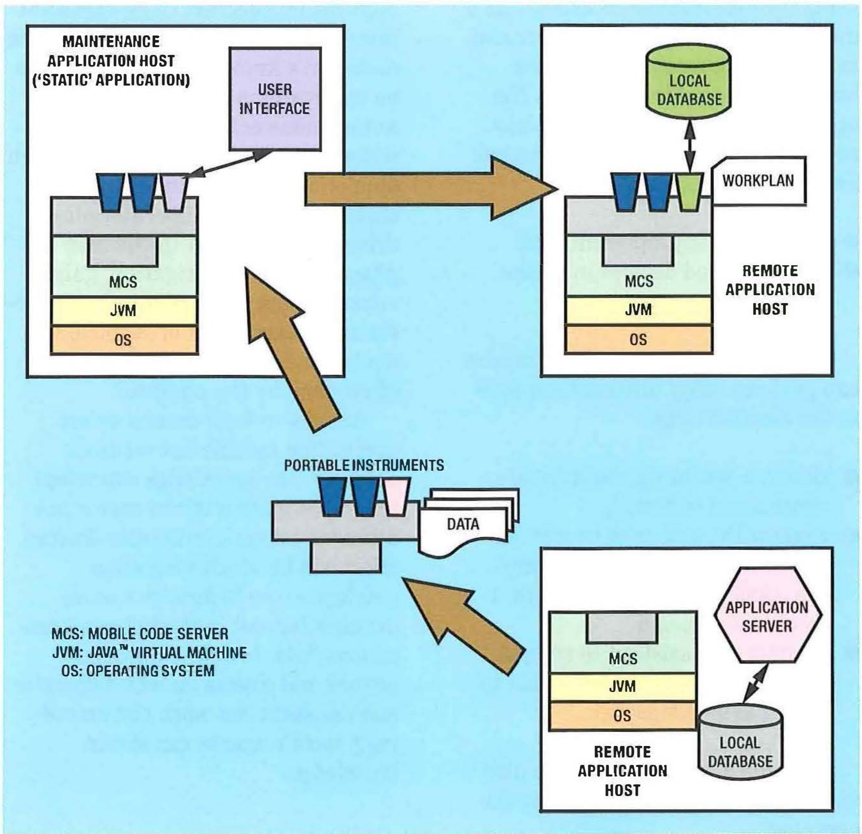
Due to the Java structure of the code of agents, accessing local databases is performed via JDBC and ODBC classes added as portable instruments.

The SNMPv1/2 stack for MIB access is added as a portable instrument too. CORBA (IIOP) is planned.

Assistant Integration

The user interface for the operator is connected to the application server

Figure 3—Mobile agents perform the task of discovering relations and data in remote applications



† Thus reducing network bandwidth requested in migration of the agent.

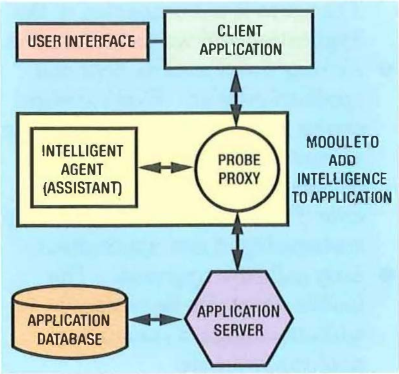


Figure 4—Architecture for assistant insertions

by means of a *probe proxy* node supporting a special facility: each message to/from the server is copied to a third party called an *assistant* (Figure 4). The assistant can send and receive messages to the client or server (without any duplication). In reality, the user interface and application servers are mobile agents 'fixed' in an MCS located in their host.

Usually the assistant does not disturb the interaction between the operator and the application server: it just looks at requests and answers and silently—using its knowledge base—obtains some conclusions. In effect, the operator and the assistant expert systems have the same goal(s) (mainly to identify chronic faults in the network). Each action made by the operator (query, selection of an NO etc.) using the user interface is encoded as a message for the application server and is a significant message too for the knowledge base of the assistant. The operator interacts with some tools to analyse the system being investigated; for example, to ask for:

- alarms density istogram, and
- most alarmed devices in a time interval etc.

As the operator needs help, he/she can perform many different requests to the assistant like:

- obtain a synthesis about facts/conclusions collected,
- request the assistant to continue autonomously the search operation (and eventually interrupt it as necessary), and
- request the assistant to suggest what should be the best action to continue investigation.

If enabled, the assistant can also spontaneously send messages to the operator.

Obtaining Knowledge about the Network

Mobile agents using specific portable instruments are used to extract from the network layout relations Pr(A,x,y) and some configuration information from resident SNMP agents. Interfaces to custom data sources are also built as portable instruments. As the assistant attempts to reach a sub-goal in its reasoning chain, either it fires locally executed rules or activates a specific mobile agent giving it a workplan for a job to perform on the remote host. A set of portable instruments is supplied to the agent to facilitate its job. The workplan thus is a piece of knowledge code supplied by the assistant. Portable instruments interface with the local application and carry out the work. The result of the action of portable instruments is understood at a level of the knowledge base as a sub-goal realised.

Experienced Result and Future Effort

Currently some tests have been perfomed using a prototype to collect signal alarms and quality measures in PDH and SDH long-distance networks used as a test-bed. Isochronism analysis has been demonstrated to be a valid method, but discovering Pr() relations is quite application specific. User–assistant interactions are well focused but the assistant's knowledge-base tends to be applications specific too. The autonomous action of the assistant unfortunately does not always reach appreciable results in locating chronic faults. The operator often drives the assistant in the first phases of the investigation, gains vantage from the assistant's synthesis until a small set of suspicious devices is located; the final step is often done by the operator.

Actions on local databases are application specific but relations obtained (the knowledge extracted) tend to be quite uniform over many different network structures. Future effort will be identifying some ontologies (see Reference 4 as an example but not in the telecommunications field) to define uniformly actions and objects on which operator and assistant can work and accordingly mobile agents can obtain knowledge.

References

1 SASISEKHARAN, RAGURAM; KAO HSU, YUNG; and SIMEN, DAVID. SCOUT: An Approach to Automating Diagnoses of Faults in Large Scale Networks. Proc. IEEE Globecom 93, IEEE Com. Soc., Pisctaway, N. J., 1993, pp. 212–216.

2 CORN, P. *et al.* An Autonomous Distributed Expert System for Switched Network Maintenance. Proc. IEEE Globecom 88, IEEE Comm. Soc. Pisctaway N. J., 1988, pp. 1530–1537.

3 SASISEKHARAN, RAGURAM; SESHADRI, V.; and WEISS, SHOLOM M. Data Mining and Forecasting in Large-Scale Telecommunication Networks. *IEEE Expert Intelligent Systems and their Applications*, Feb. 1996, pp. 37–43.

4 FERNANDO LOPEZ, MARIAN; GOMEZ-PEREZ, ASUNCION, *et al.* Building a Chemical Onthology Using Methontology and the Ontology Design Environment. *IEEE Expert Intelligent Systems and their Applications*, Jan./Feb. 1999, pp. 37–46.

5 SUSILO, GATOT; BIESZCZAD, ANDRZEJ; and PAGUREK, BERNARD. Infrastructure for Advanced Network Management based on Mobile Code. Systems and Computer Engineering, Carleton University, 1125 Colonel by Drive, Ottawa, Ontario, Canada K1S 5B6.

6 BIESZCZAD, ANDRZEJ. Advanced Network Management in the Network Management Perpetuum Mobile Procura Project. SCE Technical Report SCE-97-07. http://www.sce.carleton.ca/netmanage/perpetuum.shtml#Publications

7 Mobile Code Toolkit. Carleton University Ottawa. http://www.sce.carleton.ca/netmanage/mctoolkit

Biography

Vittorio Gelmi
Sirti S.p.A.

Vittorio Gelmi is currently at Sirti S.p.A. He has been working on appling techniques from artificial intelligence to problems in network planning. His interest include approaches in automatic network design, and automatic fault diagnosis.

Maff Holladay, Chris Russell, and Pete Barnsley

Problem Management in 2005

The Players, The Rules, The Games and The Winners

In the year 2005, services provided over networks will have evolved. Some will be supplied to corporate customers as part of a complete enterprise solution but many more will be provided to meet the moment for individuals and smaller enterprises. The marketplace will be a delicate balance of competition and cooperation.

This paper explores the influences the world of 2005 might have on problem management.

In 'Players' we describe the enterprises that will be competing in the network markets. In 'Rules' we explore the assumptions that have been made about problem management today and how a vision of 2005 might impact this. In 'Games' we describe new methods that might be adopted for managing problems. In 'Winners' we explore the support capabilities that are required in each 'Game' and use the findings to draw conclusions about these methods.

Players

It is tempting to see the network marketplace of the future as a world dominated by relatively few players. The growth in owned subsidiaries[1] supports this. However, as these large companies wrestle to maintain control of the market sector they like to see as their own, emerging technologies uncover new opportunities that smaller companies find easier to exploit. In 1996, the growth in SMEs (small-to-medium enterprises with less than 100 employees) in the US was seen to be twice as fast as the average[1].

To explore this future world we decompose the industry by roles and responsibilities described in an enterprise model (see figure opposite)[2, 3]. This describes four types of role detailed in the table opposite: these are roles not players; a player may take on one, a few or all the roles. This is not a business model.

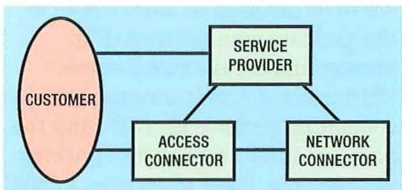
Using this model allows us to explore the effects of shifting responsibilities from one role to another.

Maff Holladay:
E-mail: maff.holladay@bt.com
Chris Russell:
E-mail: chris.m.russell@bt.com
Pete Barnsley:
E-mail: pete.barnsley@bt.com
British Telecommunications plc
Room 69/B54
BT Adastral Park
Martlesham Heath
Ipswish IP5 3RE

Rules

What then are the 'rules' of problem management? We want to establish what aspects of dealing with problems have been taken as immutable and see if this is still the case in the future. Three representative rules are listed here.

Industry enterprise model

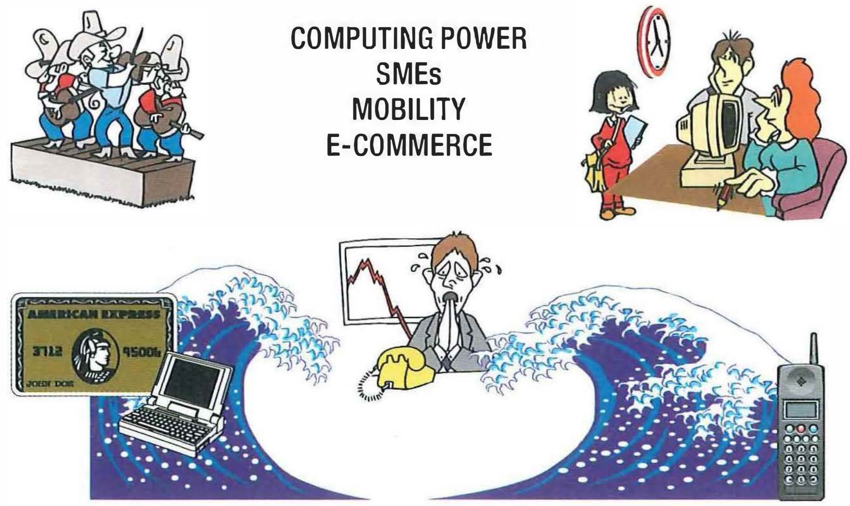


1. Problems must be resolved
When dealing with reported problems, we assume that an investigation is justified and, where causes are found, that remedial action must be taken. For network connections provided with levels of guarantee, this may remain true, but for services provided over these networks in a 'use and dispose' environment there may be no need to resolve every problem.

2. An enterprise deals with its own problems
A network operator has, until now, taken responsibility for tracking down and resolving problems in its own network. With global services provided over infrastructure owned by many players will this always be the case?

3. Problems are raised by the customer.
Operators have put in place significant customer facing infrastructure to handle the receipt of problem reports. Evolving services, however, are provided over complex value chains where the enterprise responsible for resolving a perceived problem is less clear.

Role	Prime Responsibility
Customer	To use services to contract terms
Service Provider (SP)	To provide addressed services to customers by using network and access connectors; for example, intelligent network (IN) serrvices information storage and processing, conferencing, content, etc.
Network Connector (NC)	To deliver addressed 'streams' to correct address (network connection point)
Access Connector (AC)	To provide physical connectivity between customer premises and network connection points. Logically represent the customer at the network connection point and the network at the customer's location



A vision of 2005

The year 2005

The rules above give us a flavour of the assumptions made about problem management today. We now need to travel through time to the year 2005 and see how well these rules hold. To do this we use a *vision world* of 2005 developed in the IBTE Journal[4].

Here we find that to operate a business with just a telephone as a communications tool is not enough. Some form of computing peripheral (CP) is essential. Computing tools will also be in use in the home. People will be much more mobile and use voice communications while on the move. Electronic commerce between businesses will be ubiquitous to support a dynamic service economy. Small and medium enterprises are the major driving force to the service economy. They are specialised highly-dynamic companies responding to the demand of end-customers and the global corporates whom they supply. The days of the 'one-man-band' are numbered.

Games

So in a market reacting as much to technological influences as to political, social and economic influences, two types of world could evolve:

An *Open World* is one in which cooperation is fostered. By 2005, regulation will have encouraged the openness of systems, supported by gateways or application programming interfaces (APIs), and the convergence of standards that enable these gateways to evolve. This standardisation enables uniformity between players' infrastructures. Distributed processing and IP technologies will be open thus enabling access to testing facilities in a user friendly manner. The trust required to allow and use these gateways will be founded on the trust already gained from electronic commerce between company roles such as SP/NC/AC. This will precipitate a trend towards the sharing of tools and data used to diagnose and resolve problems.

A *Closed World* is a world where trust is not prevalent. In this world, services are delivered to individuals without the confidence of e-commerce and solutions which keep player participation more segregated will find favour. This world will be supported by technologies that allow cheap replaceable items and services. This world can coexist with the open world.

Scenario: Customer enabled

The open world encourages customers to be involved in more stages of the problem management (PM) process and breaks rule 2 above. With users of services empowered by computing peripherals (CP), and the power of these technologies growing year on year, users of services in 2005 will be keen on control and want to be empowered to track, diagnose and fix their problems.

Customer enabled

A barrier to this customer enabled approach is that e-commerce is not prevalent in the individual market. Since individuals are not fully e-commerce enabled, SP/NC/AC players will hesitate to allow people access to their infrastructure to test, diagnose and fix faults. Nevertheless, these reservations will not apply in all scenarios; the following scenarios PM enable the customer without significant access to systems being opened up by the service providers:

- Things are so simple that diagnosis is not necessary;
- Self-testing may have pointed out a fault with a clear set of instructions to the user on how any fault may be fixed. This enables the customers to repair their own faults.
- The customer is in the habit of checking. This would remove the need to inform the customer when a fault is cleared.
- The customer's CP is continually checking the service, as your modem does when you are using the telephone line for data.

Scenario: Third party

The open world will foster niche PM service players. The specialisation and the competitive marketplace will encourage SME service providers to outsource non-core activities (such as dealing with problems), and so

Third-party testing

develop a PM service market. This will break rule 2 above.

The players operating in the NC/AC roles will see the market price for their services reduced to the margin and will be looking to differentiate themselves. Diversification into the PM service market using their established credentials and 'trusted' brands is a likely option. In fact the regulator may encourage separation of PM service roles since this may increase customer service levels generally and boost the market, concentrating as it does on the value parts of the service people buy (sentiments like 'I don't want mine to fail').

The fact that there will be extended personal choice will mean that people will like this approach of having independent auditors, companies mirroring the AA† who can be called upon to do problem diagnosis and fixing for them.

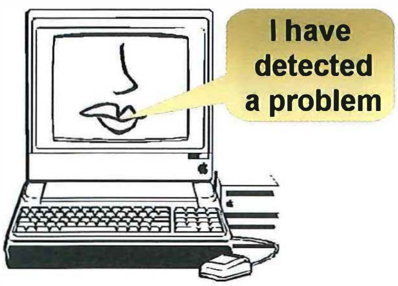
An adaptation of the third-party scenario is that the third party fixes faults that have already been diagnosed. Service providers, network connectors, or access connectors may still handle the reporting and diagnosis of their faults, but the fixing and repairs may be carried out by an external third party. A subtle addition is that players only fix faults diagnosed as belonging to them. Another variation on this theme is that the specialist with the tools, perhaps the network connector, diagnoses all faults, but service faults are passed on to the service provider to fix and access faults to the access connector.

The vision environment tends to indicate that SP role players want to concentrate on their core business and outsource problem management to another player. The NC role player is less likely to use a third party PM since they will have problem control as a key part of their core business to differentiate them in their marketplace. This also applies to the AC role player unless they are 'locked-in' to one NC and negotiate with the NC to be the third party that undertakes PM on their behalf.

Scenario: PM enabled

The open world allows the service to be more aware of PM issues. In this scenario, the service provider is still responsible for problem management

PM enabled



† The AA (Automobile Association) is a roadside assistance company operating in the UK

* The MOT (Ministry of Transport test) is required by each vehicle using the publlic roads in the UK

PM enabled

but is controlling from a distance by empowering the service to look after itself. Under these circumstances, it is the service that reports faults rather than the customer, breaking rule 3 above.

We expect the advances in technology to precipitate the introduction of 'problem aware' services. In today's office, the photocopier exhibits these characteristics; it guides the user through the diagnose and repair procedures, with instructions for opening doors, removing jammed paper and reordering originals, in order to complete a given job. The photocopier also identifies when the scope of the problem is beyond its own capabilities and tells you to contact an engineer.

The above example gives the customer (user) greater visibility of the problem. Network technology on the other hand can hide the problem from the customer by automatically fixing the problem; for example, automatic re-routing (ARR) detects a network failure and sets up another call-path (maintaining revenue). This can rely on a degree of over-provision.

With the service itself PM enabled, the service provider is going to take on responsibility for many of the steps in problem resolution. The SP is keen to accept these responsibilities because they have invested in a service that will do a lot of the work for them.

The open standards, providing appropriate exchange of information between the AC, NC and SP role players, is key to adopting this approach.

Scenario: Servicing

The style of service that best suits the closed world is one with little need for PM operational support. One way to minimise this requirement is to pre-empt problems by scheduled servicing.

This servicing analogy can be extended to include the concept of an

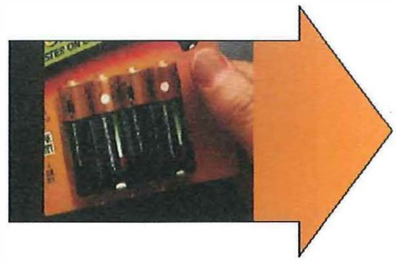
Servicing

MOT*—a 'worthy for service' badge that has been earned by the role player. The SP role players can adopt this approach to develop trust from the marketplace for their service and allows them to differentiate the same service with tailored packages. The NC/AC role players can adopt the same model with their suppliers, allowing them to check compatibility of their infrastructure.

Where the lifetime of the service is too short to justify this, the scenario is unlikely to be adopted by SPs. It does, however, provide an attractive proposition for roles with much of the 'value-add' depending on owned infrastructure, as is the case for NCs and ACs.

Scenario: Replace

One step further than minimised PM operational support is no support at all, breaking rule 1 above. Why invest in complex diagnosis and fault-fixing tools and procedures when it is more cost-effective to



Replace

simply replace a faulty component (or even the entire service)?

We all take for granted that when the fault is due to a depleted battery, or a broken bulb, that it is up to us to replace it. Within the 'best-effort' philosophy of the Internet world, this can apply to the whole service: if a search engine returns no matches, we are not surprised, we try a different search or use a different search engine.

The SP role players may adopt this for disposable services of low value but the NC and AC role players have time constants in development of their infrastructure and its payback that preclude them adopting this. Margins mean efficiency and over provision is not an option.

The Winners

We can see from the collection of scenarios portrayed above that different approaches can be adopted for problem management depending upon factors that include: the role adopted (SP/NC/AC), the market environment and the service offered.

These factors will also influence the capabilities we would expect of our support systems. Taking seven generic capabilities that might be present in the networked service area, described in the table, the perceived importance of each capability in the open and closed worlds can be explored.

Capability	Short Description
Records	Keep the collated problem records
Diagnosis	Diagnostic aid, logic
Transfer	Manage information transfer
Legacy	Integrate legacy
Notification	Notify customer facing
Routing	Where to send alerts and reports
Monitor	Compare expected and actual performance

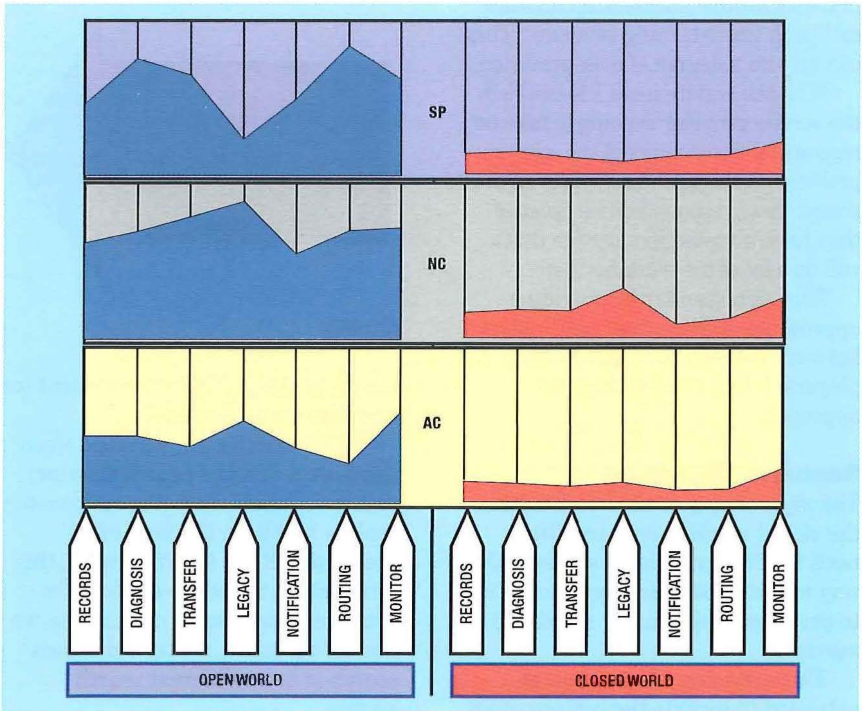
The aggregated results for each of the three roles are shown in the diagram below, which gives us 'the shape of problem management' in the open and closed worlds.

For the service provider, diagnosis and routing are the most significant capability in the open world. This reflects the level of dependency the SP has on other roles. In the closed world, the level of overall operational support is less, reflecting the drive to pre-empt or replace without diagnosis, and therefore the extent to which one capability dominates is not so clear cut, but 'monitor' appears to be the most significant.

For the network connector, 'legacy' is most significant with 'monitor' also important, in both the open world and the closed world. This supports the hypothesis that the NC should concentrate on pre-emptive management of the network regardless of the services it transports; that is, it should look after the business of network connectivity.

For the access connector, 'monitor' is most significant in the both worlds, with 'legacy' running a close second. This reflects the synergy between the AC and NC roles, although the levels of operational support would seem to be less for the access connector and the relative significance is less pronounced.

The shape of PM



Conclusions

The winners, then, will be the enterprises that build the appropriate set of problem management capabilities into their support systems. The reverse of this is that those enterprises launching services in market arenas to which their problem management systems are not suited will be likely to fail.

For any player, recognising the complexion of their PM support enables them to assess the chance of success in a new market. For a player fighting for a share in many markets, the shape of the PM support required for a new service can help decide whether to adapt an existing support system or use a new one.

References

1 Colchester, Nico. The World we are in. Prospect, Nov. 1996. http://www.prospect-magazine.co.uk/.

2 Dobson, J.; and Martin, M. (The University of Newcastle upon Tyne). The Ontology of Enterprises and Information Systems. Association for Information Systems, 1997 Americas Conference, Indianapolis, Indiana, Aug. 1997. http://hsb.baylor.edu/ramsower/ais.ac.97/papers/dobson.htm

3 Russell, C.; Holladay, M.; Barnsley, P.; Martin, M.; and Dobson, J. Enterprise Modelling the Telecommunications Industry. *BT Technol. J.*, **17**, April 1999.

4 Russell, Chris; Holladay, Maff; and Barnsley, Pete. The Future World of Telecommunications, 2005. *Br. Telecommun. Eng.*, Oct. 1998, **17**, p. 163.

Biographies

Maff Holladay
British Telecommunications plc

Maff Holladay joined the UK Post Office in 1979 and has held a variety of posts in PO Telecommunications and BT. These have covered migration from paper to computer records and from analogue to digital exchanges. Prior to undertaking research work he was managing the development of interfaces to the BT charging data. His current work within Advanced Communications Research (ACR) is analysing the changing requirements of BT's operational support systems (OSS) as we enter the 21st century.

Chris Russell
British Telecommunications plc

Chris Russell is a senior member of the OSS research team within Applied Communications Research (ACR). His current research is concerned with enterprise modelling and how enterprise analysis can be used to inform the architecture and design of IT systems. He gained an B.Sc.(Hons.) in Physics and Computing from the University of Hull and joined BT in 1987. Before joining ACR Chris worked as a systems engineer within Systems Engineering. He gained a Diploma in Management Studies (DMS) in 1993 and is a member of the Institute of Management.

Pete Barnsley
British Telecommunications plc

Pete Barnsley joined BT in 1987 after completing an M.Sc. in Optoelctronics and Lasers at St Andrews University. While investigating developments in high-capacity optical transmission systems he published more than 40 research papers, generated 6 granted patents and completed his Ph.D. In 1997, he joined a team researching flexible service platforms for fast service creation, based on ideas from TINA-C. In early 1998, the team conducted the world's first operational TINA-like public Internet-based platform (called *ESP*). His present research interests are the study of systems that support services in a dynamic competitive and regulated market and the tools and methods to analyse problems associated with building them.

Romolo Pietroiusti, Alessandro Farnesi and Enzo Garetti

Network and Service Interoperability in the Infocom Era

The changing market of telecommunication services, characterised by the fast growth of the Internet and mobility services, emphasises the need for true interoperability between different networks and services/ applications provided on heterogeneous networks. The requirements for interoperability apply not only to the network transport infrastructures, but also to the control and intelligence layer, to provide advanced services across different network boundaries. These needs are strongly related as well to the perspectives of voice-data-fixed-mobile convergence processes fostered by Internet paradigms.

Romolo Pietroiusti: Telecom Italia
Viale Europa, 190
00144 Rome (Italy)
Tel: +39 063 687 6580
Fax: +39 063 687 5137
E-mail:
romolo.pietroiusti@telecomitalia.it
Alessandro Farnesi: CSELT
Via Reiss Romoli, 274
10148 Torino (Italy)
Tel: +39 011 228 7378
Fax: +39 011 228 5685
E-mail: alessandro.farnesi@cselt.it
Enzo Garetti: CSELT
Via Reiss Romoli, 274
10148 Torino (Italy)
Tel: +39 011 228 6981
Telex: 043 220539 CSELT
Fax: +39 011 228 6909
E-mail: enzo.garetti@cselt.it

Introduction

The market of telecommunication services is rapidly changing due to many factors such as technological advancements, convergence processes (for example, convergence of information, telecommunications and entertainment technologies, convergence of fixed and mobile services and networks), liberalisation, competition, changing end-user requirements etc. There is in general a growing expectation, especially from business customers, for the establishment of globally interoperable multimedia communications. More and more, enterprise mission-critical applications are relying on telecommunication services and networks capabilities to meet increasing needs for security, seamless availability, one-stop-shopping integrated offers and high quality of service. The fast growths of Internet and mobility services are impacting dramatically the way customers access and use telecommunication services. Customers will want to use existing and new services independently of their actual position and whatever customer premises equipment they use. At the same time, service providers and network operators want their services to be available as widely as possible. In order to make this possible, services and networks need to be interoperable in a multi-operator and multi-network context.

True interoperability between networks and services is one of the most important requirements in an information and communication environment that is rapidly evolving toward complexity and heterogeneity. Competition in the standards arena and in the markets, the speed of technological evolution and the changing requirements of customers are bringing faster deployments of new networks and services, often based on proprietary or pre-standard and mono-vendor solutions. This way of deploying telecommunication networks and services generates high risks in terms of scalability, evolution and interoperability of the solutions.

So there is an increasing need of guarantees for proper interworking among different networks and services, owned either by a single operator or by multiple operators. This is of particular relevance also for the convergence processes among telephone, data, fixed, and mobile services and networks.

This paper examines some aspects of these issues, outlining the role of Internet technology in the migration path from traditional vertically integrated systems towards horizontally interoperable networks which are increasingly independent of the transport technologies and open to externally provided services and applications.

Interoperability Layering and Scenarios

Interoperability layering

Interoperability has to be examined taking into account the different needs and implications involved at the various layers which may be identified in the telecommunication industry, as shown schematically in Figure 1.

Interworking and interoperability concern the transport layer, the intelligence and services layer, the applications layer, and the management features of the networks and services.

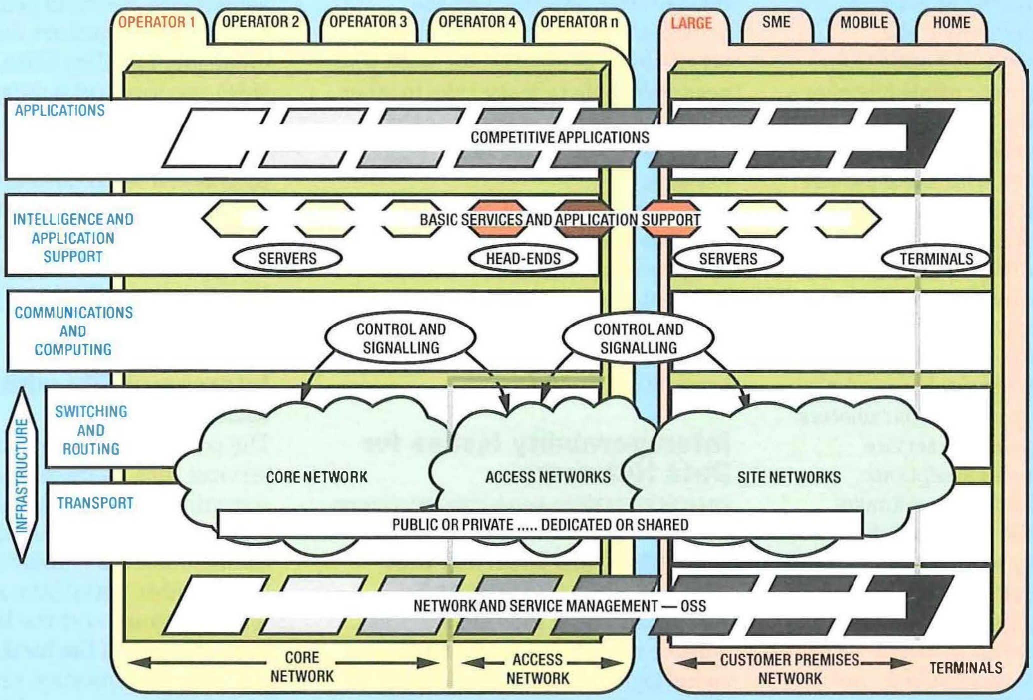


Figure 1—Layering model for interworking and interoperability

In the transport layer networks and sub-networks, interconnectivity has to be assured independently, whether they constitute the access or the backbone part of an end-to-end connection, or they are dedicated or shared, public or private. Besides the physical and logical connections of the networks, a proper interoperability among the control and signalling features of the networks is required to provide seamless networking of the customers' information.

In the intelligence and services plane, interoperability among basic services is needed as well as interoperability among 'intelligent' functions to provide advanced network services across boundary network dominions, such as intelligent network (IN) and mobility services, personal number, global virtual private networks (VPNs), and to support integrated fixed-mobile services (for example, VPN with wireline and wireless extensions). In the future, such functionality can be incorporated within terminals, network elements (service control points (SCPs) for INs, service nodes (SNs), operational support systems (OSSs), etc.) and servers and head ends. The interworking at this layer is challenging, because traditional telecommunication services have been deployed on the basis of proprietary platforms embedded in the central switches and/or in centralised intelligent nodes (for example, SCP, Internet protocol (IP), SN), with a model for network control and service provision mostly vertically integrated. Critical issues concern security and lack of adequate international standards.

Interoperability at the applications layer ensures the correct operation of customer business and/or enterprise processes. In the competitive environment problems can arise in the provision of interoperability among applications developed in competing networks because of possible commercial impacts. Only if adequate incentives exist for service providers and operators to organise interoperability to their mutual benefit, can interoperability at applications layer be pursued.

Competition will indeed develop mostly on the market plane, hence at the applications layer, while on the other planes an evolution can be foreseen towards a mix of competition and cooperation among the various actors, aiming at a cooperative development of the technological and intelligence capabilities of the networks. This could increase the global market of the services and of the applications to the benefit of all the interested actors.

As far as the management plane is concerned, needs for OSS interconnections are increasing owing to network-interconnection requirements, number and service portability among different providers and network operators, and network unbundling mandates from regulators.

Interoperability between the management systems allows the management functions to be carried out effectively (configurations, faults, traffic, accounting) in a competitive environment to ensure an adequate level of quality of the end-to-end service.

OSS interconnection among competitors is a complex issue for a number of reasons especially when there is the need to exchange information on customer data and services. Appropriate international and industry-wide standards are essential to facilitate any type of OSS interconnection.

Interworking scenarios

Interworking and interoperability is becoming an increasing challenging task for the various actors operating in the forthcoming infocom era. Many problems may arise because of complex environments involving multiple media, multiple protocols and interconnection of networks outside any single organisation's dominion control.

Today the pace and magnitude of changes, the diversity of network elements and hierarchies, the diversity of services and options for carrying them and multiple business imperatives make the network engineering function much more complex compared with the past. The internetworking scenario is further complicated by a growing uncertainty in the mix of traffic types.

As shown in Figure 2, a broad range of new technologies is deployed, each characterised by its own features and standards. For example each network has its own parameters related to the quality of service (QoS), policy for call acceptance, engineering rules for the management of the traffic and congestion relief. Each of these networks is being used for quick provision of new services and applications; moreover a clear trend is emerging where increasingly the same services are provided with different network technologies (for example, voice over data networks).

At the same time the proliferation of service providers and network operators implies that an increasing number of them are involved in end-to-end connections. The customers' terminals and equipment are increasing in terms of typologies and features, rising further interworking issues; meanwhile the same services are being provided on different types of terminals (for example, texts on TV, video on PC).

Market and regulatory factors will tend to allow customers to use new services irrespective of whether they are customers of the operator that has launched the service, and of the customer premises equipment (CPE) they use. Operators will want to make their services available as widely as possible. To make this possible, services will need to be interoperable across competing networks. Interoperability of services between different networks, and between the networks and the equipment used by the customer to access them, is essential to users who need to be able to make calls to other users irrespective of whether they are directly connected to the same network.

An overall network interworking scenario must envisage the interworking among a number of different public and private networks, with a wide variety of terminals and CPEs and with a broad range of service centres.

Interoperability Issues for Data Networks

Impact of the Internet and voice-data convergence

The expansion of Internet services and the advancement of related IP technologies are quickly changing the shaping of the telecommunications sector.

Internet traffic is growing very rapidly; investments on data networking equipment and in particular on technologies that foster the convergence of voice and data are growing rapidly. This evolution is happening much faster than originally envisioned. Taking into account the above considerations, voice-data convergence is one of the major issues in the evolving telecommunications marketplace. Voice-data convergence will have different benefits in the short-term, mid-term and long-term time frame. In the short-term the main benefit is related to lower costs, mainly due to the current regulatory environment, which makes it cheaper to handle telephone calls over the Internet (or corporate intranets) than by using the public telephone network. In the mid-term the main benefits are related to the development of enhanced services based on voice-data service integration. In the longer-term the main benefits rely on reduced infrastructure costs by commonly handling voice, data, and video services over a data packet network.

This requires proper inter-operability between the IP network and the PSTN, including intelligent network (IN) services and the deployment of interoperable mechanisms to control IP network resources and to guarantee high QoS levels.

Interoperability with PSTN/ISDN

The possibility of providing advanced services such as freephone, call screening, call waiting, call forwarding, mobile roaming, credit card calling, voice mail, caller identification, number portability, and directory assistance over the IP network represent one of the hottest issue in the telecommunication sector. This requires the interoperability between the IP network and the public switched telephone network (PSTN) not only at the transport layer but also at the signalling and control layer. The issue is of particular relevance because present data networks include generally only the transport infrastructure, without a clearly identifiable signalling and control layer, as shown schematically in Figure 3. The definition and the deployment of an intelligent control layer for the data network and its interoperability with the Signalling System No. 7 (SS7) and the IN of the PSTN is currently being analysed by the telecommunication industry and in standardisation fora.

To deliver advanced services across hybrid (PSTN and IP) networks, proper interworking must be built between the protocols used within the data network and the SS7 functionality that is inherent in the PSTN. This will provide service equivalency and seamless service with the PSTN, will enable advanced

Figure 2—Interworking scenarios

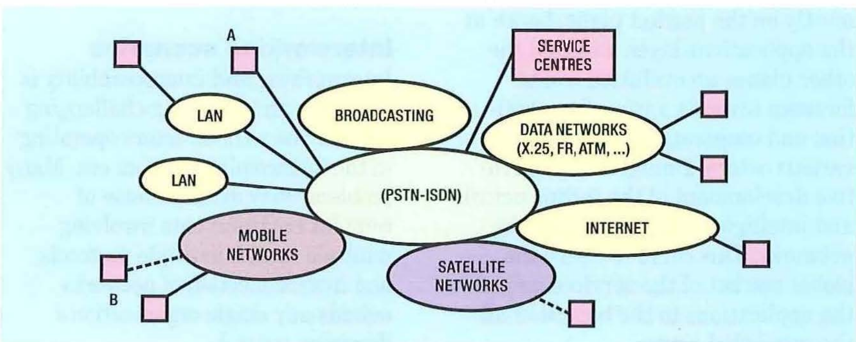
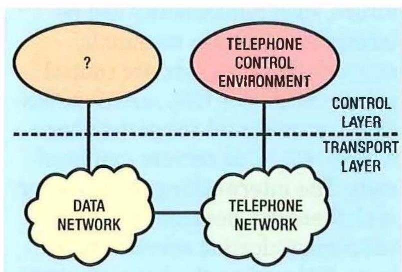


Figure 3—The missing components for a converged voice/data environment



services across boundary network dominions and will avoid multistage dialling (presently a voice-over-IP (VoIP) user must first dial the gateway, then enter a PIN number for authentication, and then dial the destination number). However, reaching agreement on IN/IP integration is a difficult process, because of the proliferation of standards; for example, H.323, SIP, IPDC, MGCP. Moreover such types of standards are primarily 'on-net' call management schemes, in which the call-signalling information shares the same call paths used to transport the VoIP services. Information for call control—such as credit authorisation, physical IP address, calling name or transport route assignments—must be contained within the IP network used to transport the user information.

This limits a service provider's ability to offer network features such as dynamic call routing, prepaid cards, subscriber roaming, 800 services and other services that require access to nationwide, or 'off-net', databases.

The next phase of standardisation is much more complicated aiming at adding signalling transfer point (STP) functionality on the IP network as well as access to SCPs to achieve a true integration of IN and a common IP-PSTN service infrastructure as shown schematically in Figure 4. All this needs to be centrally coordinated and controlled in order to allow carriers to handle billing, administration and global route control.

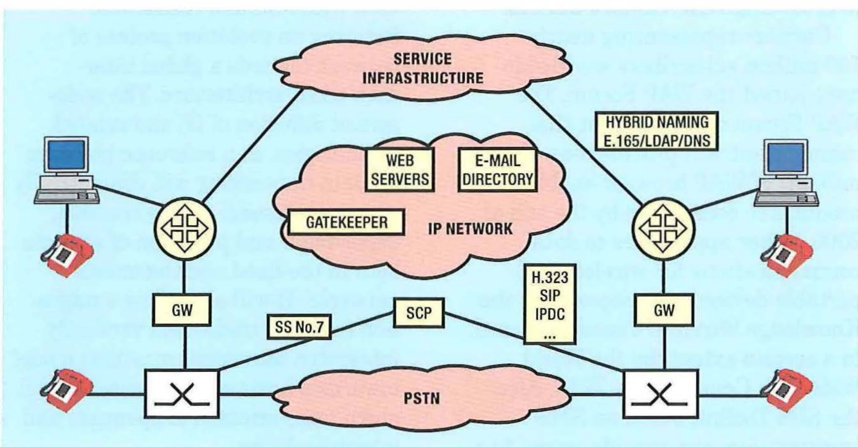
Control of QoS and network resources

As regards the control of QoS and resources within multiservice IP networks, a major issue is the quality of voice and of other real-time services. Obtaining acceptable quality of voice transmissions over an IP network is challenging because of the connectionless character of this network which implies there is no explicit mechanism for ensuring that a customer obtains a specific level of quality. For voice transmissions, a key problem is the time delay the network creates as a voice packet travels from sender to receiver. In the Internet the best latency that can typically be achieved is about 260 ms but, with a large volume of traffic, the performance of voice calls can easily be unacceptable. Fax is much less sensitive than voice to Internet delays, and that is why fax over the Internet is likely to be a successful service before voice.

Other important issues are the network engineering mechanisms, such as the management of resources, the control of traffic, the relief of network congestion and the protection against faults. While the PSTN has been standardised and engineered during the past few decades with high levels of service availability and network resource optimisation, connectionless data networks are less mature and many of the above mechanisms have yet to be found to achieve an overall reliability equivalent to that of the voice networks.

IP and asynchronous transfer mode (ATM) are presently considered as the transport technologies of future networks. IP has exploded as an industry standard with the growth of the Internet, and it is almost universally recognised as the best platform support for multimedia services. However, IP, because of its connectionless nature, cannot provide the guarantees of QoS levels required by real-time services. On the other hand, ATM can simultaneously transport multiple types of traffic including real-time traffic, so that many techniques of IP-ATM integration have been developed to benefit from the best characteristics of the two technologies. Many of these techniques are proprietary and do not interwork each other; others have been standardised (for example, LANE and MPOA) but do not scale for wide-scale public network applications. In parallel, other techniques have been developed to achieve adequate QoS levels, introducing forms of service class prioritisation (for example, DiffServ) with differentiated treatment of traffic in relation to different classes of IP services, or introducing new protocols for explicit control of both QoS and network resources, similar to that of connection-oriented networks (for example, resource reservation protocol (RSVP) and multi-protocol label switching (MPLS)).

Figure 4—PSTN-IP-IN interworking



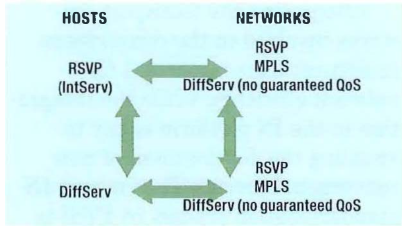


Figure 5—Interoperability issues in multiservice IP networks

These techniques may be implemented at the edge of the IP network or directly in the hosts and in the user terminals so that the various applications can indicate explicitly their requirements. Many of these techniques will be implemented, so that a rather complex set of interworking configurations can arise as shown schematically in Figure 5. In particular, the provision of end-to-end QoS levels across multiple networks dominions appears very challenging and requires careful attention in the standardisation fora.

Fixed-Mobile Interoperability

Fixed-mobile convergence

Interest in convergent solutions for the provision of fixed and mobile services has grown recently because of the new liberalisation conditions and the induced levels of competition. A service paradigm based on joint mobile and fixed service offerings can create business opportunities and benefits on both commercial and operations sides.

Integrating the transport networks involved in the convergence configuration is important for network efficiency, while the integration in the IN platform is key to enabling the development of new convergent services. The current IN standardisation process in ETSI is aimed at the definition of a capability set based on core INAP CS3 which should apply for both fixed and mobile networks. One of the most interesting services that can be provided, especially for the business market segment, sharing the same IN, is the VPN, which allows the integration of fixed and mobile users on a single service platform.

However, the above scenario cannot be realised in the short/medium term because, in most cases, two separate IN platforms are already working in the two reference networks. A good level of consistency between independent IN platforms can be recovered by adopting the concept of virtual home environment. A virtual home environment feature enables the user to maintain his/her service profile and terminal usage mode through different networks and can offer the same service to fixed and mobile users. This feature is reached through network interoperability, realising a complete alignment of the user data between the involved networks.

IP multiservices fixed-mobile network

Recent studies predict that IP multimedia services will soon contribute significantly to the overall mobile market, which by the year 2005 could range between 80 and 140 million users. Present estimates show that access flexibility and the possibility of being reached everywhere are considered to be of extreme importance by people usually working with a PC.

One of the main technological drivers to allow the provision of data communications to mobile users is the wireless application protocol (WAP), standardised by WAP Forum. WAP is a de facto open global standard suite of protocols which allows access to the Internet (and in general to World Wide Web (WWW) information) by a mobile telephone implementing a *micro browser*. In fact, WAP is derived from the hypertext transfer protocol (HTTP) and transmission control protocol/Internet protocol (TCP-IP), simplified in order to cope with mobile data services (low bandwidth, high latency, and unstable connections).

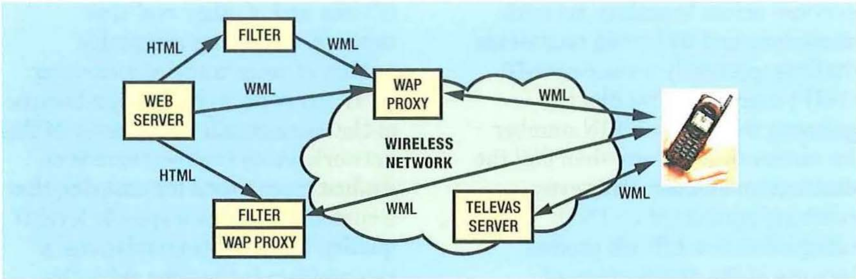


Figure 6—Configurations of WAP services

The WAP protocol is a typical example of a technical solution able to ensure the interworking among services designed originally for different kinds of networks. In particular, WAP-based technology enables the design of advanced, interactive and real-time mobile services, such as mobile banking or Internet based new services (e-mail, text-only web browsing), which can be used in digital mobile telephones or other mobile devices. A WAP proxy is located between wireless and WWW environments. It includes a gateway and a content coder/decoder, and can get, as shown in Figure 6, the content provided by a standard web server through a wireless markup language (WML) derived from the hyper-text markup language (HTML). The WAP protocol also allows the provision of telephony value-added services (TeleVAS) such as unified messaging.

The WAP specification enables applications from different equipment suppliers to work consistently on the digital networks and provide interconnection to data services and the Internet. WAP is independent from the underlying network technology (global system for mobile communications (GSM), code division multiple access (CDMA), time division multiple access (TDMA), etc.). Handset manufacturers, representing over 75% of the world market, have committed to producing WAP-enabled devices.

Carriers representing nearly 100 million subscribers worldwide have joined the WAP Forum. The WAP Forum considers that this commitment will provide tens of millions of WAP browser-enabled products to consumers by the end of 2000. Other approaches to data communications for wireless and portable devices are proposed by the Knowledge Wireless Consortium and, to a certain extent, by the World Wide Web Consortium—W3C. Also the SIM Toolkit, based on SMS transmission, can provide access to a server, mainly for information services, telephone banking and mobile commerce.

Multimedia service provision in the mobile networks will be challenging as mobile computing environments require additional mechanisms with respect to wireline network. Wireless channels are prone to burst and location-dependent error; furthermore, contention for the wireless channel is location dependent, and mobile users may move from lightly loaded cells to heavily loaded cells. As a consequence, wireless channel resources are highly dynamic; resource contracts that are made in one cell may not be valid when users move to another cell. Owing to these dynamics, the mobile computing environments require sophisticated interworking mechanisms for adaptation at customer devices and at network boundaries.

Conclusions

Interoperability among networks and services will be a major requirement in the forthcoming infocom era.

Different types of operators with different needs will rely on different technologies and will benefit from different network architectures, so that the general environment will be characterised by a multitude of heterogeneous network options.

Nevertheless services and applications need to be created, provided and used with minimal restriction, fostering an evolution process of network towards a global inter-networked architecture. The wide-spread diffusion of IP, and related technologies, as a reference platform for data networking will dramatically change the model for the creation, deployment and provision of services both in the fixed and the mobile networks. It will also allow a migration from the traditional vertically integrated telecommunications model towards a horizontally layered model much more oriented to openness and interoperability.

Biographies

Enzo Garetti
CSELT

Enzo Garetti received the degree in Electronic Engineering from the Politecnico of Turin in 1970. He was then employed in CSELT, working on telephone signalling, digital switching and broadband ATM networks. He has worked in many RACE Projects, in standardisation activities and in strategic study projects of EURESCOM. He has been responsible for switching system architectures and head of the Network Architectures Department. At present, he is Assistant to the Director of the Switching and Network Services Directorate of CSELT and is working within ACTS projects. He holds a number of patents and is the author of many technical papers.

Romolo Pietroiusti
Telecom Italia

Romolo Pietroiusti received the degree of doctor Engineer in Electronics from University of Rome in 1972. From 1974–76 he was the researcher in Ugo Bordoni Foundation on digital transmission and optical-fibre systems. In 1976, he joined Telecom Italia (formerly SIP) where he has been engaged mainly on research and development activities on network architectures and technologies. Currently he is responsible for technological strategies in the Strategies Department, working mainly on the evolution of information transport techniques, networking services and fixed-mobile convergence. He has been responsible for standardisation working groups in ETSI and in ITU. He is the author of about 70 technical papers.

Alessandro Farnesi
CSELT

Alessandro Farnesi received the degree in Electronic Engineering from the University of Rome 'La Sapienza' in 1991. He joined CSELT in 1995 working initially on the set-up and management of the Italian national host. He worked within ACTS Project AC234 (CONVAIR). At present he is involved in the coordination of CSELT's participation in international research programmes.

Dionisis X. Adamopoulos, Constantine A. Papandreou, Constantine Apostolas, George Pavlou, Christos Danopoulos and Nikos Makris

Future Trends and Perspectives in Information Networking: Towards Global Broadband Networks and Services

The telecommunications industry is currently characterised by major challenges and opportunities, and by great technological and organisational diversity, which is mainly due to the deregulation of its markets, and the progressive convergence of telecommunications and distributed computing. Consequently, there is a growing demand for a great variety of new sophisticated telecommunications services. TINA-C and personal communications services (PCS) are among the most important developments in the field of information networking, which promise to make the future open information market a reality. In this paper, after a brief analysis of important information networking aspects, both of these developments are examined focusing especially on their future perspective. Therefore, possible enhancements of TINA-C through the use of frameworks, design patterns, and agent technology, together with the application of programmable intelligent service infrastructures in PCS are considered. Finally, the need for technology integration, joint action, consensus, and cooperation is highlighted.

Introduction

Due to rapid technological development, market growth, and deregulation the telecommunications industry is heading towards a fully liberalised global market, where regulatory changes are breaking down the traditional barriers between public and private domains. The result of this liberalisation is a fragmented market with a multiplicity of competing and/or cooperating providers of telecommunications services. Increasing customer needs press for the creation, operation, and management of many types of services, ranging from simple telephony services to new advanced multimedia telecommunications services (telematic services).

In this highly-dynamic continuously-changing open environment, which is shaped by the convergence of information and telecommunication technologies, a steadily growing number of increasingly powerful communication and information services is being offered. In this context information networking is gradually gaining momentum, formulating an open market of new telecommunications services where the vision is 'information any time, at any place, in any form'[16]. Within this electronic market, which is an important precondition for the emerging information society, information resources are available to everybody, without any practical restrictions, because, in an all-digital environment, 'there is no difference

Dionisis X. Adamopoulos:
Centre for Communication Systems Research,
School of Electrical Engineering, IT and Mathematics,
University of Surrey, England.
D. Adamopoulos@ee.surrey.ac.uk
Dr. Constantine Apostolas:
Department of Informatics,
University of Athens,
Athens, Greece.
cap@di.uoa.gr
Christos Danopoulos:
Ericsson Hellas S.A.,
Zepou 33 & Vouliagmenis Ave.,
Athens, Greece.
christos.danopoulos@etg.ericsson.se
Dr. Constantine A. Papandreou:
Hellenic Telecommunications Organisation (OTE),
Athens, Greece.
kospap@org.ote.gr
Prof. George Pavlou:
Centre for Communication Systems Research,
School of Electrical Engineering,
IT and Mathematics,
University of Surrey, England.
G.Pavlou@ee.surrey.ac.uk
Nikos Makris:
Motorola Hungary,
Lajos Utca 48-66,
Budapest, Hungary.
nmakris@ibm.net

between bits which represent text, audio or video, since bits are bits'[20].

This paper examines the most important trends in the field of information networking, focusing especially on the perspectives of the framework defined by the Telecommunications Information Networking Architecture Consortium (TINA-C), and personal communications services (PCS). The application of programmable intelligent service infrastructures, like TINA-C, in third-generation mobile systems, will result in the efficient support of a variety of service engineering activities, involving both fixed and mobile services, in an environment open to competition, and open to changes in market and technology.

Information Networking: Services and Infrastructure

The innovative developments that are taking place in today's diversified telecommunications world are simultaneously driven by progress in network/system architectures and software (technology push), and by the need for more universal and personal communication and information capabilities independent of location and time (market pull)[3, 6].

To pave the way to the information society, it is necessary to define and develop a telecommunications service infrastructure (an information network), above the bearer network infrastructure, which will control and manage the distribution of information, in its various media manifestations, between geographically distributed user entities. This is the task of information networking, which more specifically is required to[2, 6, 17]:

- offer a rich variety of telecommunications services, including conventional telephony and data transmission services, as well as advanced multimedia services;
- provide and support new communication concepts, such as conference services, distribution services, and personal mobility;
- hide the complexity and heterogeneity of the underlying computing and networking technology;
- meet the scalability, reliability, and performance requirements of the customers;
- permit management by different stakeholders, such as service providers, network operators, and customers;
- allow user access to personalised interfaces from any point in the network;
- allow roaming, such as that found in cellular telephony networks;
- support the transmission of many media types, such as voice, text, still, and moving images; and
- provide openness to all types of end-users, to all types of services and applications, and to changes, in time or space, of service software and hardware.

Telematic services have a central role in information networking. A telematic service is considered to be a geographically distributed entity (actually an encapsulation of a number of cooperating entities distributed over a geographical environment) providing a number of people (users, subscribers) a predefined carefully-selected set of facilities regarding the integrated coverage of a (possibly) wide range of information and communication needs, utilising the resources of (existing and future) telecommunication networks[1].

Among the virtually limitless variety of existing and future telematic services, those with multimedia characteristics enjoy currently a great demand. People recognise that multimedia telematic services with enhanced content can offer them powerful support for communication, information, education, entertainment, and cooperation. Therefore, this kind of service is rapidly emerging, and is seen as an important and strongly increasing market in the office, at home, and on the move[3].

The efficient provision of multimedia telematic services involves a number of new technologies. These enable the necessary bandwidths, connectivity, flexibility, support of simultaneous and time-staggered sessions, improved negotiation and renegotiation procedures, as well as appropriate manageability. Furthermore, advanced transfer protocols are required to cope with the differentiated, changing, and unpredictable requirements concerning connections, bit rates, and qualities for the various information types involved in specific services.

A promising technology to support new telecommunications services in the office, at home, and on the move, including services with multimedia characteristics is asynchronous transfer mode (ATM). Beginning with data applications, ATM networks will increasingly address audio and video needs, too. ATM can be employed 'edge to edge' for public and private networks and 'end to end' including servers and terminals. Furthermore, ATM facilitates the smooth integration of networks from the local to the wide area, and is an especially promising technology for the local loop (wireless ATM)[23].

Another development that is of growing importance and should not be underestimated is the dramatic expansion of the Internet, and its corporate incarnations (intranets, extranets), together with the popularity of the World Wide Web. Graphical user interfaces provided by web browsers give the user a convenient, visual, easily manipulated interface to the vast repository of (any type of) information available at Internet sites throughout the world. The Web has therefore become the first worldwide multimedia communication system. Albeit (relatively) slow and sometimes unreliable, the Internet is able to support voice telephony and multimedia telematic services, such as videoconferencing. For fast responses and a much higher reliability level several solutions are emerging, like the IPng (IP next generation), TCP/IP over ATM, and cable-based Internet systems.

It is conceivable that the variety of existing and evolving networks will partially converge to establish full-service and application-ready networks. A common ATM core network could interlink the diverse access networks (including cordless and cellular mobile radio), integrated services digital network (ISDN), broadband ISDN, Internet/intranet and other data/corporate networks, as well as cable TV systems[3, 22]. This would be an important step towards the information infrastructures of tomorrow (for example, the European and global information infrastructure, EII-GII) that will enable people and systems to communicate securely with each other, any time, and anywhere with acceptable quality and at reasonable cost[4, 21].

The TINA-C Paradigm

The growing demand for a great diversity of broadband, multimedia, and multi-party telecommunications services, in the framework of information networking, requires the adoption of rapid, intelligent, and flexible service provisioning approaches. Therefore, an open universal service platform which supports both telecommunications and management applications is required. Such a platform will result in the 'virtualisation' of the (computing and communications) infrastructure, increasing significantly the proliferation of telematic services. A virtual infrastructure (VI) is the outcome of the adoption of a software perspective that embraces a much more dynamic and heterogeneous approach to the representation of information, the configuration of hardware and software systems that process it, and the binding of resources to its processing and distribution[24].

In full agreement with this perspective some of the world's leading telecommunications and information technology companies established TINA-C in 1993. The main objective of TINA-C is to define and validate an open, innovative, and coherent architectural framework (a long-term service architecture) that would address in an integrated manner service control and service management. This framework encompasses the long-term objectives of both IN and TMN, applies open distributed processing (ODP) standards and object-oriented design principles, facilitates the design and provision of services in a heterogeneous system and network environment with different domains of ownership, and ensures the introduction of new and enhanced services and their management, much faster and more efficiently than with current approaches[6, 12].

The TINA-C overall architecture encompasses basic design and modelling rules, which are applicable to a wide range of services, and decomposes the complexity of the problem space into a number of complementary (sub)architectures (computing, service, networking)[25]. TINA-C services are considered as software-based applications that operate on a distributed processing environment (DPE). This environment hides from services the underlying technologies and distribution concerns, and supports in this way the portability and interoperability of the service code. Therefore, a service is realised by a set of interacting service components (that is, computational objects interacting via their computational interfaces), which are distributed across different network elements. Service components can be characterised as service dependent or service independent. The former support features that are specific to a service, while the latter support features that may be common to a wide range of services[10].

To enforce the approach taken in TINA-C and to facilitate the corresponding service creation process the concepts of frameworks and design patterns are introduced and applied. A framework can be defined as a set of prefabricated services together with some architectural concepts that define the constraints to put these services together. The architecture includes the rules that can be used to integrate the single services and to define possible flows of control between them[19]. Design patterns represent abstract solutions for specific problem classes. They capture the static and dynamic structure, and collaboration of a group of objects. They can be defined in an abstract language-independent way[11]. While design patterns can be considered as a horizontal structure over a set of computational objects, frameworks can be considered as vertical, domain specific (for example, telecommunications specific) configurations of components.

In the case of TINA-C, design patterns can be defined by identifying groups of interworking service objects, where every group is characterised by a micro-architecture that determines the way the objects interact to provide a solution for the specific aspects of a subproblem that arises during the development of a telematic service. Furthermore, a framework can be defined as the overall architecture, which specifies how the identified configurations of service objects can collaborate to implement a solution for the whole problem. Thus, a framework is a kind of construction kit for complete or semi-complete telematic services. It has to be complemented and customised using inheritance techniques. As an example, the TINA-C service architecture can be defined as a framework. On the contrary, the access to the usage of a service is an example of a design pattern (access pattern). It specifies the object group of the access session.

The introduction of design patterns and frameworks in TINA-C implies the establishment of a common vocabulary and the definition of common design structures. They assist in reducing the scope of the problem-solving process in the case of service creation, because they support the identification of similar problems and similar solutions. However, design patterns and frameworks are abstract concepts. There is no guarantee that their usage will lead to design reusability, design portability, and abstract customisability. Furthermore, good design patterns and frameworks, like good inheritance hierarchies, cannot be invented in an easy way. They have to be chosen and designed very carefully[14, 19]. Otherwise, the introduction of insufficient and wrong chosen patterns and frameworks in the service creation process may hinder or even prevent the design and implementation of successful telematic services.

Besides design patterns and frameworks, another possible enhancement of TINA-C originates from the fact that in new telecommunications services, service personalisation and service interworking are of fundamental importance, because in that way customers are enabled to define when, where, for whom, and how they will be reachable or not. TINA-C is based on the concept of DPEs, but since DPEs operate according to the centralised client/server paradigm, they are relatively inflexible, and do not adapt well to rapidly changing, or customer specific requirements. A possible solution to this problem is the application of intelligent agent technology to TINA-C.

An intelligent agent is a self-contained software element responsible for performing some set of operations on behalf of a user or another program with some degree of independence or autonomy. This means that an agent performs specific tasks delegated to it, and therefore contains some level of intelligence. Additionally, agents operate asynchronously (they are often event or time triggered) and may communicate with the user, system resources, and other agents as required to perform their tasks[16].

Moreover, more advanced agents may cooperate with other agents to carry out tasks beyond the capability of a single agent. Finally, as transportable or even active objects, they move from one system to another to access remote resources or even meet other agents and cooperate with them. These agents are called *mobile* agents in contradistinction with fixed agents, which remain in a single location throughout their execution. Agent mobility is probably the most challenging property, which provides intelligent agents with the potential to influence the traditional way of communications and service realisation[10].

In TINA-C there is the explicit notion of *user agents*, which represent, both the end users within the system, and the contact points for users in order to gain access to services. Although user agents are fixed, they can interact with other user agents (via session managers) for service session establishment. In addition, there is also the notion of *terminal agents*, which should adapt communication services to the used terminal capabilities of the users.

For this reason, the transformation of TINA-C to a static agent-based telecommunications architecture (Figure 1) is relatively easy and straightforward. In such an architecture, the user agent is aware of the communication preferences (in respect to time, space, medium, cost, security, quality, accessibility, and privacy) of its user, and controls on behalf of that user all incoming and outgoing communications in respect to intelligent routing, information filtering, and service interworking. It also negotiates and cooperates with other user agents for establishing a communication session between the respective users. Furthermore, due to its service architecture, TINA-C seems to be an appropriate candidate for the incorporation of mobile agents, both for asynchronous information exchange and for the establishment of real-time information exchange[13].

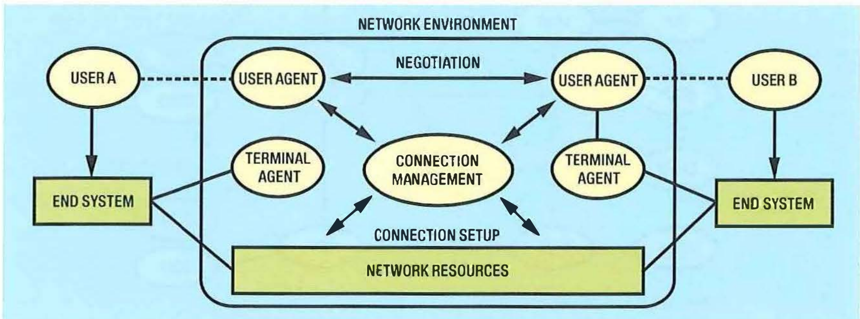
However, although intelligent agents and particularly mobile agents have a great potential to influence the design of future telecommunications services, they do not represent the ultimate solution for any kind of problem. Whether mobile agents provide for better performance than traditional client/server computing strongly depends on the interaction patterns and the 'size' (code plus state information) of the agent to be transferred[16]. Therefore, intelligent agents should be considered only as an enhancement or an 'add on' to TINA-C for the realisation of services under certain conditions.

Because of the flexibility in adapting new approaches and technologies, and because of its wide scope, TINA-C can be considered as a valuable complement to GII activities. Furthermore, it provides an architectural underpinning for the GII, as it is directed to the provisioning of any kind of services, running on a global scale, on different network technologies, allowing the combination of any kind of media and any kind of connectivity, and facilitating third-party connection setup and broadcasting as well as multiparty involvement. However, it has to be noted that the value and influence of TINA-C has been and will be measured by the extent to which it provides for industry-implementable results to accelerate the availability of TINA-like products. For this reason, the second phase of TINA-C, which started in 1998, focuses considerably on the market-driven adoption of the TINA-C architecture.

Personal Communications Services

The 'telecommunications services any time, any place' paradigm supported by rapid technological advances, propels the growth of wireless systems and networks. In full agreement with this paradigm, personal communications services (PCS) aims to provide personalised data, voice, image, and video communications services that can be accessed regardless of location, network, and time.

Figure 1—Static agent-based telecommunications architecture



The enabling factor to achieve this aim is mobility, which can be regarded in three dimensions[7,9]:

- *Terminal mobility*, which refers to the ability of wireless terminals to access services as they move across different access networks. Moreover, these networks should be able to identify and locate a terminal and notify it for the initiation of new services.
- *Personal mobility*, which refers to the ability of a user to access telecommunications services on any terminal (wireless or wireline), according to a service or user profile, and based on a unique personal number dedicated to that user.
- *Service mobility* (called also *service portability*), which refers to the ability of the network to provide subscribed services to users irrespective of the type of the used terminal, location or access network. Service mobility requires both personal and terminal mobility.

PCS provide operations to service providers, subscribers, and customers/users through mobility, customisation, service profile management, communication, and security procedures[5]. As a result, personal communications services are characterised by the following key elements:

- *Ubiquitous connectivity:* A user is always able to connect to a network accessing the subscribed services, even if the access network or the used equipment are changed. Moreover, it could be possible for a user to have different active connections using different access networks simultaneously.
- *Single session abstraction:* The service remains the same throughout the session (or call) even if the access network or terminal changes (resulting in different quality of service (QoS) experienced by the user).
- *Application transfer:* Restoring an active application (or call) from one device to another as the user changes his/her location (for example, from office to car) or the used access network.
- *Application adaptation:* The session characteristics should change in order to keep an application active as the user

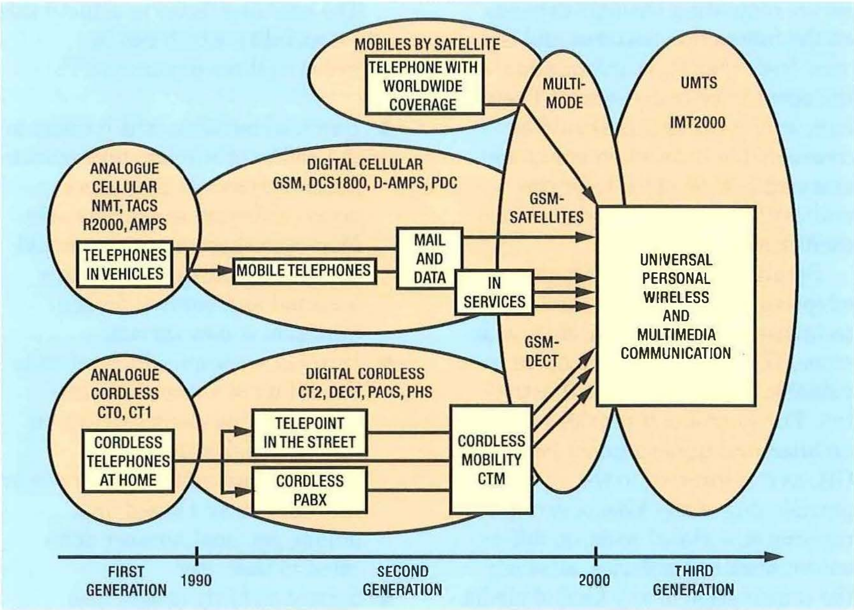


Figure 2 – Evolution of personal communications services and systems[9]

changes device or access network. Adaptation could mean changing information content or changing media format. Changes in the information content can be triggered by a device change, which influences the available network bandwidth, the processing power, the memory capacity, and the display capabilities (for example, a less capable display as the user moves from a laptop to a mobile telephone could result in a drop of the video connection for a teleconference session). Furthermore, changes in the media format are imposed by media incompatibility (for example, moving from laptop to a conventional PSTN telephone).

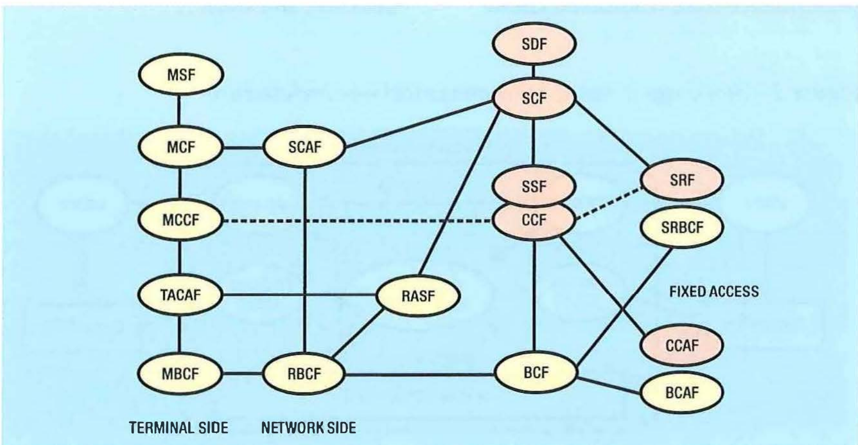
In the last decade, many PCS systems supporting a variety of services have evolved to the existing wireless networks (second generation) as Figure 2 shows. These networks are characterised by relatively narrow bandwidth and considerable limitations regarding their interworking. In order to overcome these problems and to establish a universal personal wireless and multimedia communication system, the third generation of mobile telecommunications systems is currently under development.

The universal mobile telecommunication system (UMTS), developed by the European Telecommunications Standards Institute (ETSI), and International Mobile Telecommunications by the year 2000 (IMT2000), developed by the International Telecommunications Union, which are the main representatives of this generation, are scheduled to start service around 2000–2005 and provide a range of telecommunications services to mobile and stationary users in a variety of environments. UMTS/IMT2000 will support terminal, personal and service mobility, and thus will allow person-to-person calling independent of location, terminal used, means of transmission (wired or wireless), and choice of technology. In UMTS and IMT2000, PCS (including high bit-rate data and audiovisual services) will be based on a combination of fixed and wireless/mobile services to create seamless end-to-end services for the user.

UMTS/IMT2000 will be coupled with IN to develop and deploy PCS, following the paradigm of the customised applications for mobile network enhanced logic (CAMEL) environment, which has been introduced into the GSM system to utilise IN technology and support services not covered by standardised GSM. In that sense, new functional entities have been added to the IN functional model, as can be seen in Figure 3, to deal with specific UMTS requirements. These are the mobile storage function (MSF), mobile control function (MCF), mobile call control function (MCCF), terminal access control agent function (TACAF), mobile bearer control function (MBCF), service control agent function (SCAF), radio bearer control function (RBCF), radio access control function (RACF), bearer control function (BCF), specialised resource bearer control function (SRBCF), and bearer control agent function (BCAF). These functional entities are added to provide the necessary mobility and interworking functions, and to carry out the service related processing.

This enriched UMTS/IMT2000 system will eventually lead to a virtual home environment (VHE), which is defined as a system concept for personalised service portability across network boundaries and between terminals[8]. VHE is provided by a service provider according to a subscription (that could involve more than one user) over different terminals and network operators. However, VHE will enable users to receive services with the same feel and look as they roam across different networks. In general, VHE aims to provide users with personalised services, a personalised user interface, a consistent set of services irrespective of access (fixed, wireless), and global service availability when roaming. The achievement of this goal depends on the cooperation and relative capabilities of home and

Figure 3 – Extended IN functional model for UMTS



visited networks, as well as the relative capabilities and compatibility of the used terminals[26].

Currently, most VHE approaches are heavily based on IN technology. However, IN does not provide a homogeneous programming environment to allow interworking and portability of system components, and IN infrastructure depends on the technology used by operators for implementation. To overcome such IN limitations, and to facilitate the implementation of the VHE concept in a more flexible, efficient, and effective way, the application of TINA-C is proposed[7].

This would require a modification of the TINA-C architecture to fully support mobility functions. Such an effort has already been undertaken by several European Union research projects. Among them, DOLMEN has provided a number of TINA-C extensions in the area of mobility, validated these extensions through an international field trial, and highlighted a number of important open issues that will be the target of further research in this area[27]. However, irrespective of the outcome of future TINA-C specific research regarding the VHE, integration with the VHE standardisation efforts that are currently on going is necessary.

Conclusions

Existing telecommunications systems are gradually converging into a ubiquitous information infrastructure inside an open deregulated multi-provider telecommunications marketplace. Additionally, the demand for telecommunications services is increasing and will increase rapidly in the years to come. Based on these assumptions, this paper has outlined and examined what appears to be some major factors and technologies to be considered in deriving a viable paradigm for the broadband information highway of the future.

The TINA-C architectural framework is gradually maturing and gaining acceptance at a constant rate, through validation activities and the examination of interworking concepts and migration strategies from existing technologies. For this reason, it is envisaged that TINA-C will provide universal programmable connectivity between mobile users supporting a variety of mobile services (cordless, cellular, paging, mobile satellite) in the framework of the universal broadband mobile telecommunication system (UBMTS) that will emerge as the result of the activities pertaining to the fourth generation of mobile systems. In that way fixed and mobile networks will eventually converge enabling the realisation of a VHE[7, 27].

Furthermore, the state of the Internet and its evolution is also a crucial matter for the future, whether it is going to be considered as the universal paradigm for the truly global information infrastructure, or as something destined to evolve naturally from its present state to be one of the elements of the overall telecommunications scenario.

To effectively answer the technological challenges of future telecommunications environments and to provide both fair competition and regulated liberalisation, standardisation is necessary[15]. The essential objectives of standardisation relate to interconnectivity and interoperability. Joint activities have to address suitable reference architectures, functional models, well-conceived building blocks, open interfaces and protocols, and appropriate migration scenarios, without restricting the needs and creativity of the different parties/countries involved. Balance is also required between innovation and coordination taking into consideration dynamic developments in regulation, markets, and technologies.

For the transfer towards the information society, many political and legal problems have still to be solved both on a national and international basis, in addition to technical and economical aspects. These problems relate to setting up an appropriate and stable regulation for competition, providing faster progress, greater choice, higher performance, and lower costs and to institute a suitable legislation for the content, addressing among others, intellectual property rights, security and privacy, and the timely and accurate publication of interface specifications[24]. The process of facing these issues requires common vision, harmonisation, joint action, and above all a careful balance between cooperation and competition.

References

1 Adamopoulos, D. X.; Haramis, G.; and Papandreou, C. A. Rapid Prototyping of New Telecommunications Services: A Procedural Approach. *Computer Communications*, 1998, **21**, pp. 211–219.

2 Adamopoulos, D. X.; Papandreou, C. A. An Integrated Object-Oriented Approach to Telecommunications Service Engineering. Proceedings of the 8th IFAC/IFOR/IMACS/IFIP Symposium on Large Scale Systems: Theory and Applications, University of Patras, Rio, Greece, July 1998, pp. 834–839.

3 Armbrüster, H. Information Infrastructures and Multimedia Communications: Different Approaches of the Telephone, Data, and Radio/TV Worlds. *IEEE Communications Magazine*, Sept. 1997, **35**(9), pp. 92–101.

4 Bangemann, M. Europe and the Global Information Society. Recommendations to the European Council, Brussels, 1994.

5 Chan, M. C.; and Woo, Y. C. T. Next Generation Wireless Data Services: Architecture and Experience. *IEEE Personal Communications*, Feb. 1999, **6**(1), pp. 20–33.

6 Chapman, M.; Berndt, H.; and Gatti, N. Software Architecture for the Future Information Market. *Computer Communications*, 1995, **18**(11), pp. 825–837.

7 Daoud, F. Universal Broadband Mobility Target for Post-TINA Infrastructure: Toward the Fourth Generation, *IEEE Network*, July/Aug. 1998, **12**(4), pp. 28–38.

8 ETSI. Universal Mobile Telecommunications System (UMTS); Service Aspects; Virtual Home Environment (VHE). Technical Report, 1998.

9 Faggion, N.; and Hua, T. Personal Communications Services Through the Evolution of Fixed and Mobile Communications and the Intelligent Network Concept. *IEEE Network*, July/Aug. 1998, **12**(4), pp. 11–18.

10 Hagen, L.; Breugst, M.; and Magedanz, T. Impacts of Mobile Agent Technology on Mobile Communication System Evolution. *IEEE Personal Communications*, Aug. 1998, **5**(4), pp. 56–69.

11 Hansen, T. Development of Successful Object-Oriented Frameworks. Proceedings of the 1997 ACM SIGPLAN conference on Object-Oriented Programming, Systems, Languages, and Applications (OOPSLA '97), Atlanta, USA, Oct. 1997, pp. 115–119.

12 Inoue, Y.; Guha, D.; and Berndt, H. The TINA Consortium. *IEEE Communications Magazine*, Sept. 1998, **36**(9), pp. 130–136.

13 KIND, A.; ANAGNOSTOU, M. E.; LOUTA, M. D.; NICKLISCH, J. and TOUS, J. Towards the Seamless Integration of Mobile Agents into Service Creation Practice. Paving the Way for an Open Service Market—IS&N '99, LNCS, **1597**, Springer-Verlag, 1999, pp. 328–341.
14 KOERNER, E. Patterns for Constructing CSCW Applications in TINA. Interactive Distributed Multimedia Systems and Telecommunication Services—IDMS '97, Lecture Notes in Computer Science, **1309**, Springer-Verlag, 1997, pp. 322–329.
15 LU, H.-L.; TOUBASSI, A.; LUCAS, F.; and RENON, F. Network Evolution in the Context of the Global Information Infrastructure. *IEEE Communications Magazine*, **36**(8), Aug. 1998, pp. 98–102.
16 MAGEDANZ, T. On the Impacts of Intelligent Agent Concepts on Future Telecommunication Environment. Bringing Telecommunication Services to the People—IS&N '95, Lecture Notes in Computer Science, **998**, Springer-Verlag, 1995, pp. 396–414.
17 MARSHALL, I. W.; and BAGLEY, M. The Information Services Supermarket—an Information Network Prototype. *BT Technol. J.*, April 1995, **13**(2), April 1995, pp. 132–142.
18 MCKINNEY, R. D.; MONTGOMERY, W. A.; OUIBRAHIM, H.; SIJBEN, P.; and STANAWAY, J. J. Service-Centric Networks. *Bell Labs Technical Journal*. July–Sept. 1998, pp. 98–115.
19 MONROE, R. T.; KOMPANEK, A.; MELTON, R.; and GARLAN, D. Architectural Styles, Design Patterns, and Objects. *IEEE Software*, Jan. 1997, pp. 43–52.
20 NEGROPONTE, N. Being Digital. Alfred A. Knopf, New York, and Hodder & Stoughton, London, 1995.
21 RALPH, BONNIE; and REID, ANDY, Convergence—Synthesising a New Industry. *Br. Telecommun. Eng.*, Jan. 1996, **14**(4), pp. 274–280.
22 RAYCHAUDHURI, D.; and WILSON, N. D. ATM-Based Transport Architecture for Multiservices Wireless Personal Communication Networks. *IEEE JSAC*, Oct. 1994. **12**(8), pp. 1401–1414.
23 SFIKAS, G.; APOSTOLAS, C.; and TAFAZOLLI, R. The UK LINK PCP Approach to the Wireless ATM System. *IEEE Communications Magazine*, Nov. 1997, **35**(11), pp. 60–70.
24 TENNENHOUSE, D.; LAMPSON, B.; GILLETT, S. E.; and KLEIN, J. S. Virtual Infrastructure: Putting Information Infrastructure on the Technology Curve. *Computer Networks and ISDN Systems*, 1996, **28**, pp. 1769–1790.
25 TINA-C. Overall Architecture. TB_MDC.018_1.0_94, Feb. 1995.
26 TORABI, M.; and BUHRKE, R. Third Generation Mobile Telecommunications and Virtual Home Environment: A Prioritization Analysis. *Bell Labs Technical Journal*, July-Sept. 1998, pp. 50–62.
27 TRIGILA, S.; RAATIKAINEN, K.; WIND, B.; and REYNOLDS, P. Mobility in Long-Term Service Architectures and Distributed Platforms. *IEEE Personal Communications*, Aug. 1998, **5**(4), pp. 44–55.

Biographies

Dionisis X. Adamopoulos
University of Surrey

Dionisis X. Adamopoulos holds a degree in Computer Science from the Athens University of Economics and Business, and a Masters degree in Telematics with distinction from the department of Electronic and Electrical Engineering of the University of Surrey. He is involved in Ph.D. research at the Centre for Communication Systems Research (CCSR) of the University of Surrey. His research interests include service engineering, distributed multimedia, object-oriented analysis and design, telematic services, distributed object platforms, groupware, and telecommunications management.

Dr. Constantine A. Papandreou
Hellenic Telecommunications Organisation

Dr. Constantine A. Papandreou holds a Dipl. Ing. degree as well as a postgraduate degree in Engineering Economics both from the Technical University of Munich. He also holds a Doctor degree in Telematics from the University of Munich. He has served since 1970 in the Hellenic Telecommunications Organisation (OTE) as an expert in telecommunications and informatics in various positions. Since 1992, he has been a Director in OTE. In parallel to these activities he has been teaching for several years teleinformatics and information technology at the Athens University of Economics and Business, the University of Piraeus, and the Higher School of Telecommunications of OTE. He is the author of over 60 scientific papers pertaining to his research activity in the fields of telematics, information systems, telecommunications policy, education and training, multimedia, office automation, service engineering, etc.

Dr. Constantine Apostolas
University of Athens

Dr. Constantine Apostolas holds an Engineering Diploma in Computer Engineering and Informatics from University of Patras, Greece, an M.Sc. in Telecommunications from the School of Electrical and Electronic Engineering, University of Surrey, England, and a Ph.D. in the area of multiple access control and mobility management for ATM wireless systems, from University of Surrey, England. From 1995 until 1998 he was a research fellow at the Centre for Communications Systems Research, University of Surrey, England. Currently he is with the Department of Informatics and Telecommunications, University of Athens, Greece. His main research interests are multiple access control, mobility management, and service provision in wireless systems for multimedia.

Prof. George Pavlou
University of Surrey

Prof. George Pavlou received his Diploma in Electrical and Mechanical Engineering from the National Technical University of Athens, and his M.Sc. and Ph.D. in Computer Science, both from University College London. Over the past 12 years he has been undertaking and directing research in the areas of data communications and telecommunications with emphasis on performance evaluation, network and service control and management and the use of distributed object-oriented technologies in new telecommunications architectures. He has contributed to ISO, ITU-T, NMF/TMF, OMG and TINA standardisation work. He has published around 50 papers in international refereed conferences and journals and he is the co-author of two books. He is currently professor of information networking at the University of Surrey, School of Electrical Engineering and Information Technology, where he leads the activities of the networks research group.

Christos Danopoulos
Ericsson Hellas S. A.

Christos Danopoulos holds a degree in Electrical Engineering from the department of Electrical and Electronic Engineering of the University of Patras, and a Masters degree in Telematics from the department of Electronic and Electrical Engineering of the University of Surrey. Since 1994, he has been a senior product engineer (GSM systems) in Ericsson Hellas.

Nikos Makris
Motorola Hungary

Nikos Makris holds a degree in Electrical Engineering from the department of Electrical and Electronic Engineering of the University of Patras, and a Masters degree in Telematics with distinction from the department of Electronic and Electrical Engineering of the University of Surrey. He worked as a cellular systems engineer in Panafon S.A. and Nokia in Greece, Germany, Austria, and the Netherlands for several years. Currently, he is with Motorola Hungary designing and developing wireless local loop networks.

Claudio Chiarenza and Eugenio Stefanotti

Network Evolution Towards the Optical Network

There are differences and commonalties among today's networks (telephone, data and mobile); competition and the need to save on operational costs are driving convergence.

The trends clearly indicate a preference towards integrated networks.

In the medium-to-long term, wavelength-division multiplexing (WDM) and high-density wavelength-division multiplexing (HD-WDM) will become operational and this transition will create a new networking layer; the wavelength channels will be routed by all-optical equipment, just as in the same way switches handle digital channels in today's networks.

Claudio Chiarenza:
Director of Business Development & Planning
Fixed Networks Area,
Italtel Spa,
Castelletto di Settimo Milanese,
20019 Milan, Italy.
Tel: +39 02 4388 7964
Fax: +39 02 4388 7464
E-mail: claudio.chiarenza@italtel.it
Eugenio Stefanotti:
Director of Business Development & Planning
Architecture & Services Innovation
Fixed Networks Area
Italtel Spa
Castelletto di Settimo Milanese
20019 Milan, Italy
Tel: +39 02 4388 7986
Fax: +39 02 3351 0150
E-mail: eugenio.stefanotti@italtel.it

Introduction

The rapidly changing environment and requirements for telecommunication networks, as well as insecurity about which concepts and services really will become predominant, are presenting difficult challenges both to operators and suppliers.

The challenge to the operators is how they should modify or change their current networks (Figure 1) to be able to provide their customers, at the right point in time, with competitive and business-oriented services, in a flexible way.

The challenge to the suppliers is to provide the operators with products based on a future oriented and flexible architecture in order to support their business in an optimum way.

Another key question to be answered by the operators is whether to keep different services such as narrowband, broadband, data, Internet protocol (IP) and mobile in separate overlay networks, as built today, to support single applications and services, or in an integrated one.

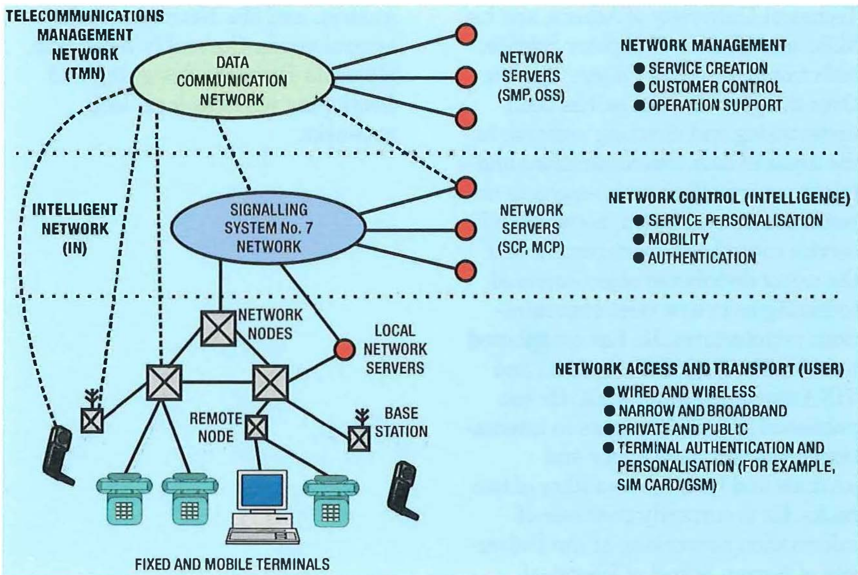
A Vision of the Current Telephone Network

It is very important to describe the two types of scenarios that can be seen now in the telecommunications market.

The high-end market segment includes about 70% of the value and the other segment about 30%. The split into two different market segments is becoming more evident as we approach the year 2000. The first is distinguished by total competition among operators leading to a vertical market approach and extreme orientation towards end customers and services. From the network perspective there will be a strong push to change the network structure towards voice and data integration.

The second segment is distinguished by only limited competition due to political influence and classical telecommunication-services such as plain-old telephone service (POTS) and TV. From the network perspective the trend is towards no change to the network structure.

Figure 1—Telecommunications network architecture



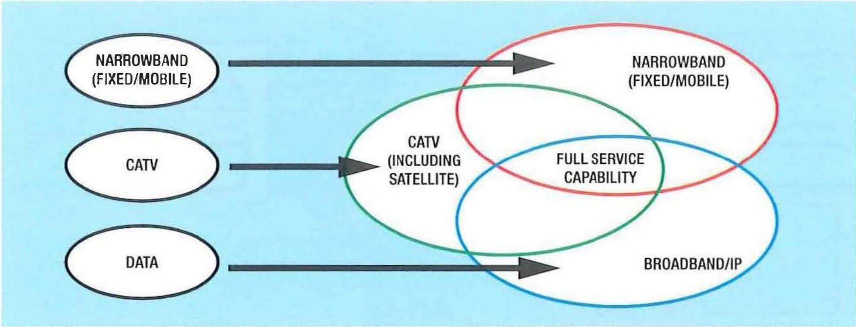


Figure 2—Network evolution. Besides the narrowband and TV broadcast networks a broadband/IP network is being established. The different networks are becoming interlinked and will have 'full service' capabilities

The evolution of the network (Figure 2) is mainly driven by the high-end market; in this market, pumped mainly by Internet and intranet growth, the sheer volume of data and mixed-media traffic has surged on both packet-and circuit-switched networks. This trend continues and will be amplified by the expansion of higher-bit-rate access options and by improved coding and applications for Internet telephony, audio, and video. Not a moment too soon, packet technology has reached the point where it surely can play a role in giving network operators—private and public—the flexibility, economy, reliability, and simplicity they require. A network operator is therefore forced to make a decision about the network architecture because the traditional one can no longer satisfy the needs—integration is the mandatory key word.

A Vision on the Current Mobile Network

Currently the architectures of the mobile networks (Figure 3) are still based on the same functional model. Some operators are introducing new services; for example, in Italy, a connection to the Internet can be offered to subscribers, as a special service, to provide a cheaper international telephone service, still with a reasonable quality.

Some operators are making trials with IP-backbone and others are making plans for incorporating the general packet radio service (GPRS) in their network, even if that can be considered an intermediate solution waiting for the universal mobile telecommunication system (UMTS), at least in Europe.

The third generation of mobile systems (TGMS) is based on a new air interface capable of supporting 2 Mbit/s data services but following essentially two different views:

- European Telecommunications Standards Institute (ETSI): UMTS—one new air interface as an evolution from the global system for mobile communications (GSM); and
- International Telecommunications Union (ITU): future public land mobile telecommunications system (FPLMTS)—'family' of air interfaces including:

—Europe: UMTS;
—Japan: probably a new air interface based on wideband code-division multiple access (WB-CDMA) technology; and
—USA: evolution of existing personal communications services (PCS) standards.

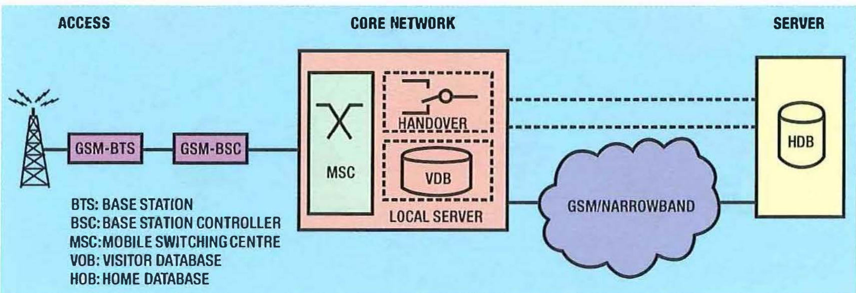
Currently FPLMTS standards activities have a more rapid time frame than UMTS activities. A failure in aligning the standards (by acceleration of the UMTS time frame) will result in a collection of incompatible FPLMTS standards rather than a single standard.

Mass explosion of UMTS will be achieved through parallel evolution in different technology domains such as information technology, transmission, terminals, service creation, mobility management, radio access and service control.

In addition to improving coverage, increasing capacity is a primary focus for mobile operators. Only UMTS offers suitable capacity for mobile voice and data and then it is a real candidate as a problem solver for the mobile network.

Another important open issue that affects operators with mobile and fixed networks is convergence. This can be obtained in three ways: convergence of services, convergence of functions (that is, billing centre, customer care centre, provisioning centre), convergence of network/architecture. It is possible to use one or more of these ways. The driver is to make savings on the network and administration, operation and maintenance by means of common use of resources. This allows lower prices to be offered to the customers.

Figure 3—Mobile communications networks architecture



A Vision on the Current Internet-Like Networks

The backbone of the Internet was originally a series of high-speed links between major supercomputer sites and educational and research institutions within the US and throughout the world. A major part of it was the NSFNet, managed by the US National Science Foundation.

In 1995, large commercial Internet service providers (ISPs), such as MCI, Sprint and UUNET, took responsibility for the Internet backbones and have considerably enhanced their capacities. Regional ISPs link into these backbones to provide lines for their subscribers, and smaller ISPs hook either directly into the national backbones or into the regional ISP.

Internet computers use the TCP/IP (transmission control protocol/Internet protocol) communications protocol. There are more than 20 million hosts on the Internet, a host being a mainframe, mini, workstation or high-end PC that is on-line via TCP/IP. The Internet is also connected to all types of computer networks worldwide through gateways that convert TCP/IP into other protocols.

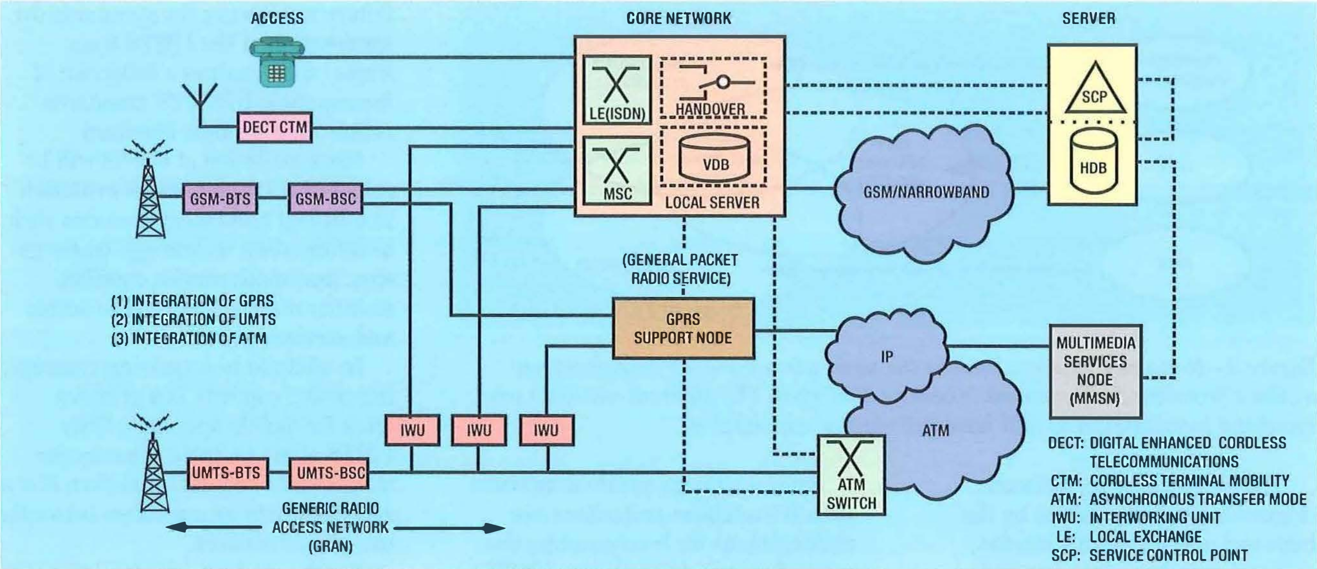


Figure 4—Convergence of communications networks

Ironically, some of the original academic and scientific users of the Internet are developing their own networks once again. The Internet is so jammed these days that they no longer enjoy the quick access they were used to.

The Internet is also used by some telco operators in order to provide cheaper international and oversea connections for international and overseas telephone traffic. Based on that experience some new operators are planning to deploy networks that can connect USA and Europe using Internet-like architecture. This means the use of a backbone, based on routers (or better Gigarouters) for the switching network and normally using optical fibre for transmission, supporting voice and data over IP.

This kind of network normally uses voice gateway for voice coding/decoding and a call agent to manage the call. The call agent normally converts the protocol from the ISDN user part (ISUP) to the media gateway control protocol (MGCP) to control the voice gateway.

These networks are based on connectionless (packet) switches while traditional networks are based on connection-oriented switches. The packets are routed in the network along several 'hops' starting from the origin to the destination; the path used for the call can change depending on the availability of the routers and links.

Such a network is well suited for carrying voice and data traffic. This means that it is suitable to support the merging of voice and data traffic—the latter is increasing rapidly worldwide while the former is increasing at a constant rate of only around a few percentage points per annum.

Steps Towards Integrating the Various Networks

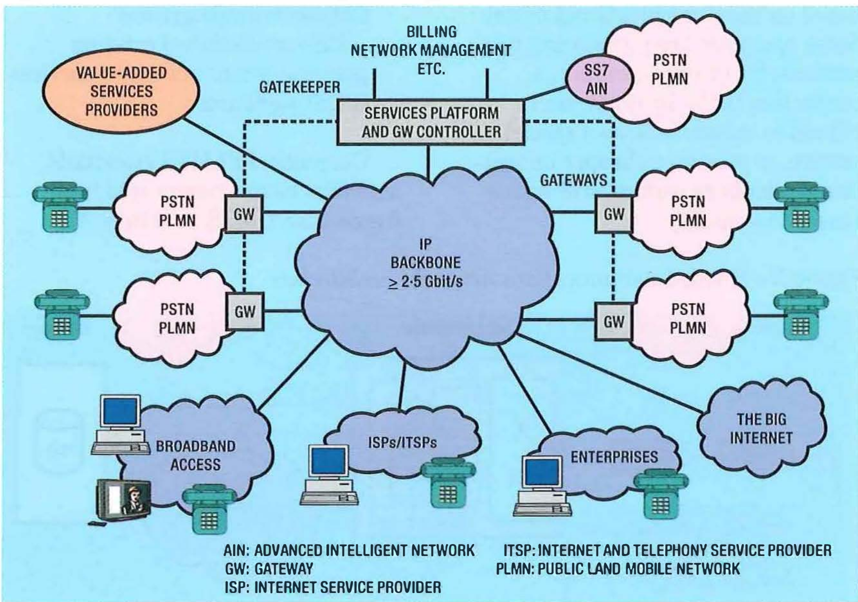
Following the theme of integration, mobile and fixed networks will converge, then data and voice networks will merge and ultimately they will collapse all into one (probably the IP network) (Figure 4).

From the network point of view, the trends clearly indicate:

- A big interest in an integrated network; this new concept and related technology brings new opportunities and ways to gain benefits and enables a complete range of services to be offered to the customers.
- Orientation towards networks based on asynchronous transfer mode (ATM), especially at the backbone level; this will come primarily from demands of the edge data networks including IP. The deployment of the ATM infrastructure will force narrowband/broadband integration and penetration of ATM in the local/access networks.
- A solution using ATM at the edge and IP inside the backbone; that could provide a reasonable compromise (integrating narrowband and broadband traffic) or an intermediate step before the full IP network can be realised (Figure 5).

At any rate, the convergence process of the telephone networks

Figure 5—Network architecture for next-generation telcos



(including both fixed and mobile ones) has to take into account one of the methods mentioned above for integrating voice and data traffic; as a consequence data networks and telephone networks will be integrated, at the same time.

Steps Towards the Optical Network

Customers are demanding more services and options and are carrying more and different types of data traffic. To provide full end-to-end connectivity, a new paradigm was necessary to meet all the needs for high capacity and various forms of traffic.

Optical networks provide the required bandwidth and flexibility to enable end-to-end wavelength services. Optical network began with wavelength division multiplexing (WDM), which arose to provide additional capacity on existing fibres. The optical network will be based on wavelengths as its basic building blocks rather than using a defined bit-rate and frame structure. The components of the optical network will be defined according to how the wavelengths are transmitted, groomed or implemented in the network.

Viewing the network from a layered approach, the optical network requires the addition of an *optical layer*. The steps needed for the complete optical network are dependent on the following basic components (Figure 6):

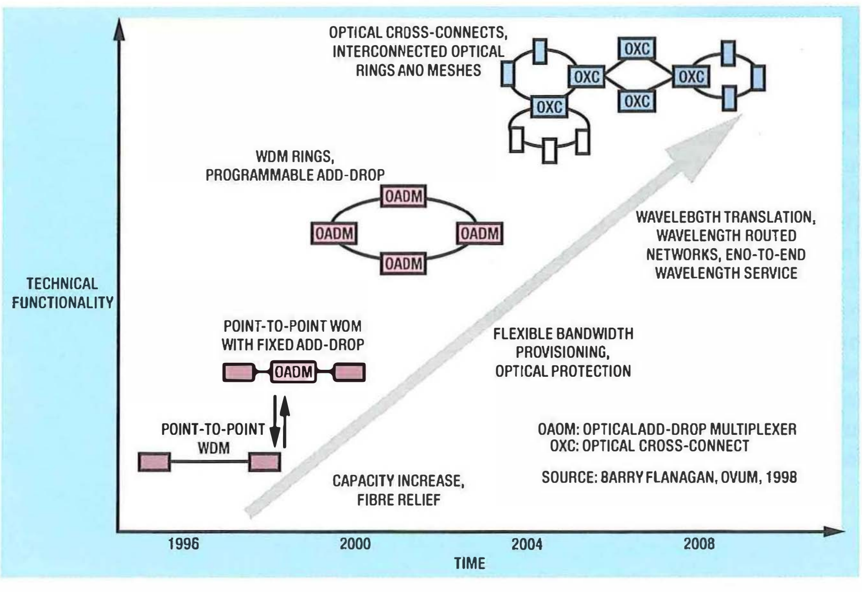
- wavelength division multiplexing (WDM, dense WDM (DWDM), and high-density WDM (HD-WDM)),
- optical add-drop multiplexer (OADM), and
- optical cross-connect (OXC).

Experience has shown that the public carrier market is an evolutionary, not a revolutionary, environment. Optical networking technology must be closely integrated with the rest of the network, but it is here that much work remains to be done. It is clear therefore that the emergence of the all-optical network and all its benefits is not going to occur overnight.

The definition of a standard will become increasingly important and essential as operators seek to interface their optical systems and evolve to more complex optical network architectures. Basic standards exists for WDM point-to-point links, but there is much work to be done in the areas of optical networking and vendor interoperability. The standard bodies are addressing many of the issues, but it is uncertain whether they can keep pace with rapidly changing market developments.

WDM and optical components will form the foundation of broadband networking in the next decade. Interesting times lie ahead in the optical transmission arena.

Figure 6—The evoltion of optical networking



WDM: A New Technology on the Scene

Wavelength-division multiplexing is a technique for transmitting traffic over fibre in multiple channels. Traditional optical fibre transmission uses light of a single wavelength (that is, color); in contrast, WDM combines multiple wavelengths of light into a single multiplexed signal for transmission along the fibre.

Each channel utilises the full capacity of the fibre being used; WDM then creates a virtual fibre for carrying traffic over; for example, four WDM wavelength channels boosts the capacity by four times.

WDM is now being used worldwide to boost dramatically the capacity of installed optical fibre cables.

However, while the basic principle of the technology is very simple, the effect WDM will have on the telecommunications industry is much more complex, extending far beyond simple capacity increases in networks. WDM is emerging as fundamental to enabling networks to cope with the telecommunication demands of the next century. The future growth of the Internet, the development of broadband networking, the creation of new high-bandwidth applications and the economics of the bandwidth market itself are all inextricably linked with this simple technology.

Outside the US, deployment has been small, although the market is now starting to take off. In Europe in particular, WDM was not a concern before 1997; network traffic levels are lower than in North America, and the network has different characteristics. Whereas North America has a large number of high-capacity long-haul routes, European backbones networks tend to consist of a number of relatively short haul (30–100 km) high-capacity links meshed together. This suggests different network economics for WDM and alternative solutions.

As traffic levels rise however, and prices come down, a host of deployments in Europe and elsewhere are emerging. BT for example, has been conducting trials on two WDM links in its national network, with plans to roll out more of the technology over the next year. It also intends to use WDM on its new pan-European network. In August, the first of a series of WDM systems went live in the network of Deutsche Telekom. Worldcom deploys WDM in its pan-European network project, Ulysses, and Hermes Europe Railtel has purchased 40-channel WDM systems for its network linking over 20 cities in Europe. These announcements and others indicate the increasing importance of WDM in carrier networks.

The Emerging Optical Layer

In the medium-to-long term, WDM will be used to route traffic on individual wavelengths in all levels of the network, significantly increasing flexibility, leading to *optical networking*.

This transition will create an optical layer (a new networking layer) in which wavelength channels are processed and routed by all-optical equipment, just as electronic multiplexers, cross-connects and switches handle semi-permanent digital channels in the SDH/SONET (synchronous digital hierarchy/ synchronous optical network) and ATM layers of today's networks.

This will involve the deployment of optical add-drop multiplexers, enabling WDM ring architectures to be constructed. In the longer term, this will also require the deployment of optical cross-connects to reconfigure and reroute individual wavelength channels in the network.

It is evident then that the broadband networks will evolve in the next decade to include a number of different layers. There will be no solution based on a single architecture for all applications, rather a range of diverse interfaces will exist in the network between ATM, IP, SDH/SONET and WDM equipment.

The main challenge for operators and vendors is how best to integrate and interwork electronic-layer and optical-layer technologies, and ensure that both can be managed in the network effectively.

The Vision of an 'All-optical Network'

The vision of an all-optical network will provide operators with flexibility and cost savings by allowing them to simplify network architectures and achieve more efficient provision and management of bandwidth resources in the network.

The optical layer will be transparent to bit rates and protocols, potentially providing a common transport mechanism independent of the type of service carried.

Different technologies will access the optical layer via an optical interface, which will translate the output of multiplexers, routers and data switches into optical signals following the standard.

This will be particularly important for access networks where a wide range of different types of traffic and line rates are being generated.

Eventually, optical networking will enable operators to provide wavelength services—that is, the ability to resell bandwidth, rather than fibre—on different dedicated wavelengths for different customers, making more efficient use of the network capacity.

Biographies

Claudio Chiarenzo
Italtel

Claudio Chiarenza was born in 1950 and joined Italtel in 1975 after graduating from Politecnico of Milan. He began working as a hardware systems engineer and then he headed Switching Systems Integration Tests. In 1987, he became Director of the Marketing Division where he coordinated several start-ups of switching and radio products. Since 1989, he has been Director of Business Development and Planning in the Fixed Networks Area; he has been programme manager for several Italtel projects such as: Linea UT, intelligent network, ISDN and DECT network.

Eugenio Stefanotti
Italtel

Eugenio Stefanotti was born in 1944 and is currently working in Italtel, which he joined in 1964. He is now Director of Architecture and Service Innovation, in the Business Development and Planning department. He has worked on the Italian intelligent network and then on DECT trials and the development of cordless mobility services for Telecom Italia and for cordless mobility applications for other countries.

Mario Pagani

The Evolution of the Optical Transport Network

Traffic density growth, hunger for bandwidth and fibre congestion are just some of the main issues addressed by telecommunications operators in pursuing one of the greatest challenges they have: the development of future-proof multi-service platforms.

Telecommunications manufacturers are currently fully committed to identifying advanced and reliable network solutions to support operators in paving the path towards the network of the future.

The optical transport network is one of the cornerstones for the new network architectures and all-optical networking is its natural evolution.

Even if the basis for this future network environment is already a reality in terms of the availability of hardware and software product, a few hurdles are still present in this telecommunications race.

Technological constraints, definition of standards, network management, performance monitoring at the optical level, network protection and survivability are the most important barriers to be overcome.

Nevertheless, reliable outlooks by manufacturers give us a glimpse of consistent solutions, offering stability to operators' network enhancement plans.

The Current Environment

In recent years transport networks have been subject worldwide to profound renewal, passing from a pure plesiochronous digital hierarchy (PDH) configuration, achieved through a series of point-to-point connections, to synchronous network configurations. These solutions have been based either on US standards, that is, SONET (synchronous optical network), or on European Telecommunications Standards Institute (ETSI) standards, namely synchronous digital hierarchy (SDH).

These synchronous hierarchy networks were, at first, devoted only to leased lines applications and initially very simple network management was employed, mainly for fault monitoring and for performance checking. Recently, SONET and SDH networks, constituting the grounds for the future optical networking evolution, have become increasingly fully managed, extending the management features to centralising the network control, routing the information and generating new services.

Today, additional demand for network bandwidth is being generated by the need to provide new broadband services and improved quality of service. Telecommunications operators have to face a new challenge, since parts of their optical transmission networks are beginning to show signs of saturation. Overcoming fibre congestion problems has been one of the main factors driving the deployment of wavelength-division multiplexing (WDM) systems in the recent past. The adoption of this technique allows, as a first step, the transmission capacity of backbones and metropolitan rings to be increased by re-using the existing cables. Transmission systems based upon the consolidated technique of WDM are now a commercial reality. A large number of point-to-point systems have already been widely deployed in the US since 1996 and in the European market since 1997.

A real optical layer is increasingly taking shape, over the SDH/SONET and services layers, effectively structuring the transport network of the future (see Figure 1). A sensible design is to construct an architecture in which this optical layer is independent of the needs of the services/technologies it supports Internet protocol (IP), video, asynchronous transfer mode (ATM), SDH, SONET, PDH, etc.), thereby providing maximum flexibility.

Network Evolution

Considering the services to be delivered to customers as the source of requirements on which transmission network evolution is to be based, three main factors are driving the services market:

- Internet data traffic growth,
- the shift of some voice traffic from fixed to mobile, and
- video channels.

In terms of subscribers, the Internet (IP) boom is characterised by a compounded annual growth rate (CAGR) of more than 60%, compared with 35% for mobile and only 7% for fixed telephony.

To effectively follow the traffic evolution, the potentiality (already available or attainable by expandability) of the equipment supporting the transmission network is an essential condition, in satisfying the hunger for bandwidth, which is continuously increasing. The bandwidth need is already appearing at local operator level (see the case of local exchange carriers (LECs) and competitive LECs in the US) and is expected to become increasingly vital in the near future.

A much discussed question is the actual opportunity to maintain a layered transmission network, as shown in Figure 1, or to eliminate

Mario Pagani:
Alcatel
Transmission Systems Division
Via Trento, 30
20059 Vimercate (Mi)
Italy
Tel: +39 39 6864700
Fax: +39 39 6864849
E-mail: Mario.Pagani@netit.alcatel.it

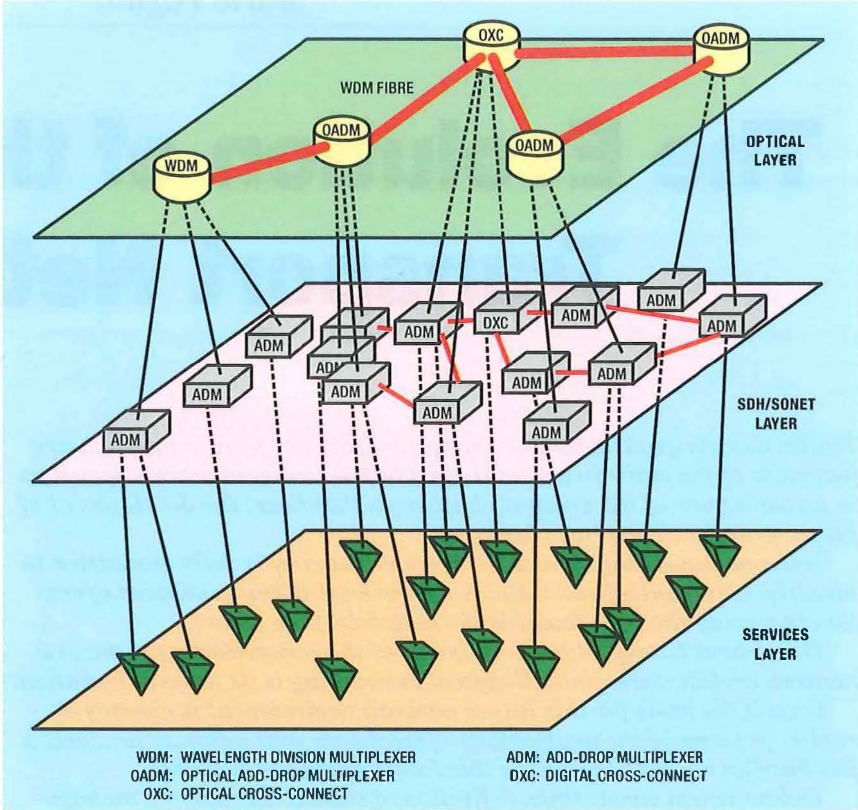


Figure 1 – Network layers

one or more layers (various alternatives are shown in Figure 2).

In particular, direct IP transport over the WDM or optical layer is still being studied in order to solve problems such as performance monitoring and restoration at the optical level.

To achieve this, optical add-drop multiplexers (OADMs) and optical cross-connects (OXCs) are needed, together with the appropriate network management. However, it still takes several years to get fully industrialised products extensively employed in the transport networks. In contrast, at the network periphery, where IP and voice coexist, the electrical layer will inevitably be required, with grooming and forwarding features for both IP and ATM, to assure, for example, all the functionalities of the SDH and SONET layer. Furthermore, another trend to incorporate WDM in routers has the very significant disadvantage of a lack of flexibility. On this subject, history teaches that the insertion of PDH in the switching exchange was a dismal failure.

Telecommunications manufacturers are continuously subject to a frenetic activity devoted to designing increasingly sophisticated network solutions, capable of adapting dynamically to the different traffic configurations required by enhanced as well as completely new services.

The future transport networks will have to possess the capability not only to allow the transit of these new services, in the most transparent way, at local, national or intercontinental level, but also to easily and economically manage them.

With this objective, a smart solution is to incorporate, besides the canonical transmission features, some key functionalities, such as IP or ATM, in the transport equipment, enabling the transport optimisation of the new services (see Figure 3).

To meet the evolutionary transport network challenges, the new optical layer will also provide the following capabilities: high-level restoration, dynamic transport reconfiguration and optical protection. Facilitating high-level restoration by aggregating bandwidth into larger components (optical channels) creates both faster and more cost-effective restoration and the optical layer is instrumental in achieving this level of efficiency.

The related network management will have to be strictly correlated to the modularity through which network and equipment are evolving. In addition, interworking with IP service management will be required.

Figure 3 – IP and ATM in transport equipment

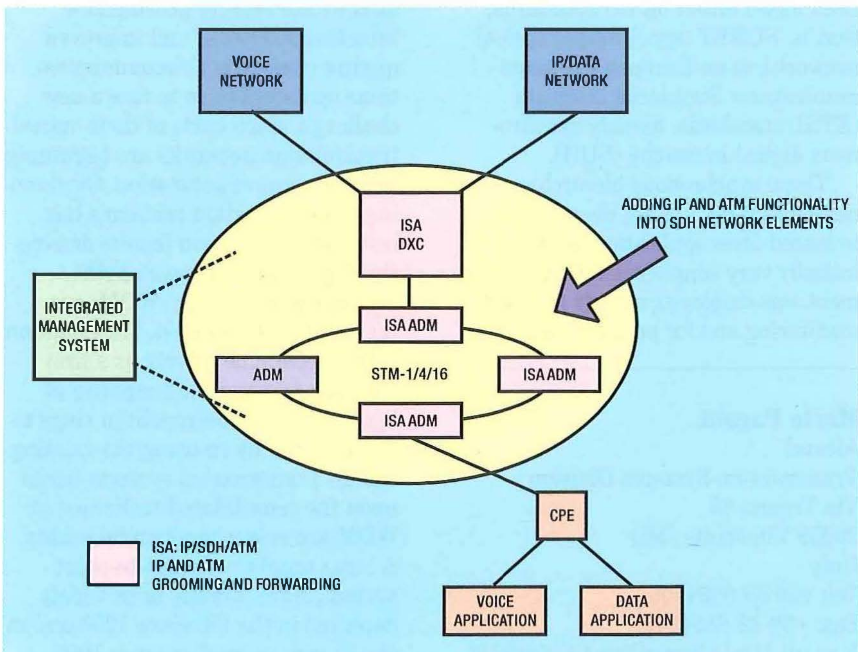
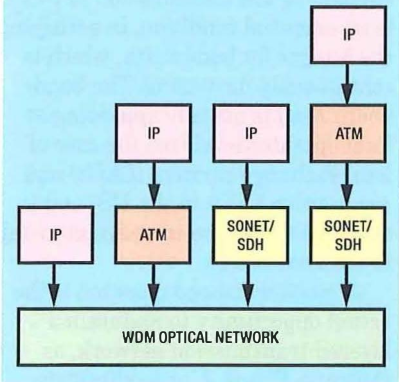


Figure 2 – Transport alternatives



Optical Layer Equipment

Consolidated WDM or dense WDM (DWDM) techniques, together with optical amplifiers, used in point-to-point applications, allow a total capacity per individual fibre of 400 Gbit/s and more. Cost-effective versions of DWDM are, on the other hand, employed in the metropolitan enterprise access market, where bit rate and protocol transparency, together with cost and performance optimised for short-haul applications, are necessary.

Next to these product families, the market also requires fully reconfigurable systems, opening up the prospect of efficiently performing networking functions in the optical domain, similar and complementary to those already implemented in the electrical SONET and SDH domains.

Passing from point-to-point connections to the multipoint-to-multipoint ones (into the optical layer), more sophisticated full optical equipment has to be employed such as reconfigurable OADMs and OXCs. An example of this network configuration is shown in Figure 4.

These systems provide protection, restoration and performance monitoring mechanisms performed at the optical layer level, therefore assuring the survivability of the optical network, arguably one of the most important factors in the design of optical-fibre telecommunications networks.

Technology Drive

In actual fact, to design the equipment used in the future optical networks most people are still working at the limits of technology. How these innovative technologies can be applied at the industrial level is yet to be assessed.

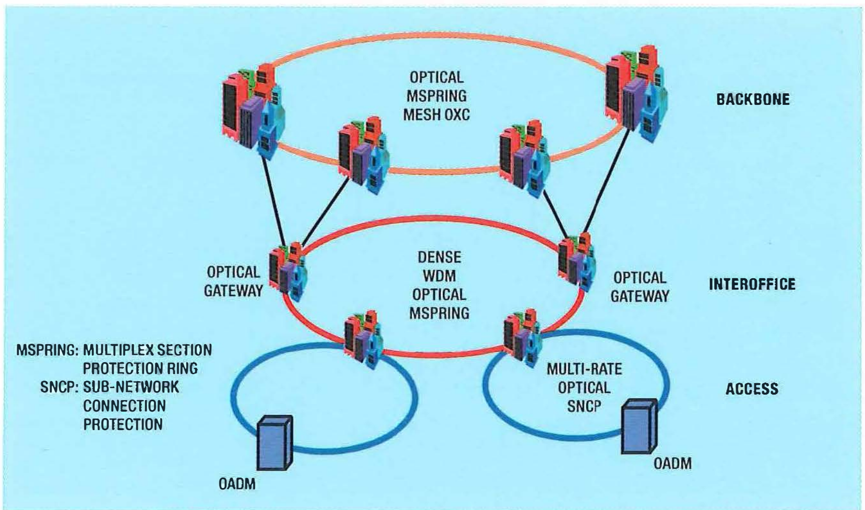
Essentially, for the optical transmitters, increasingly stable laser modules are required. The focus is on planar optics and optical micro machines as the technologies for the optical components to be used in future optical transmitters.

There is also an important technological impact on the receivers, even if not as critical as the case of transmitters. Narrow-band filtering and optical pre-amplification, at the optical receivers' front-end, will be used extensively.

In the case of line regenerators, investigations on the fully optical 3R (reamplifying-reshaping-retiming) regenerator are currently being carried out in research laboratories.

Another key technology, at present apparently given less attention, is the dispersion fibre compensator employed in compensating the chromatic dispersion of conventional G.652 installed fibres. Today the best compensator is almost unanimously recognised to be the dispersion compensating fibre (DCF) type. One of the driving points for this ranking is its easy integration with the erbium-doped fibre amplifiers (EDFAs).

Figure 4—Optical network configuration



Network Management

Extending the use of network management products already well proven for SDH and SONET to manage all optical layer products such as WDM, DWDM, OADM and OXC, etc., is the most effective choice for operators, thereby introducing an additional transmission layer on top of the 'electrical' layer.

Operators will choose the appropriate network management solution for their operational procedures. Going for combined management, the network operator enjoys the benefits of the network resource usage optimisation for traffic and protection as well as fast fault localisation and common service management throughout the layers. Going for separated management of the layers, the operator will take advantage of the rich set of network management features; for example, state-of-the-art alarm management and enhanced security mechanisms, such as network access domain (NAD) management.

Independently of the solution chosen, all the benefits mentioned above apply in full.

One of the main issues in technology driven industries, in general, is to design systems to cope with future technology changes and shifts in applications and demands. Therefore, the transmission management system has to be suitable to manage the very first synchronous networks, as well as the most updated networking technologies such as IP and ATM, WDM, DWDM and optical networking.

A reliable network management system also needs to provide evolution of transmission services as well as continuity of management services throughout upgrade or migration procedures. This allows simple introduction of new network features by upgrading the existing network or adding new network elements without affecting previous functions and services. This protects network investments by allowing the existing infrastructure to be upgraded to exploit benefits from state-of-the-art technology.

Another important feature of the network management system is its 'openness'. The network management system must be suitable for integration into legacy systems and be able to facilitate the integration of legacy systems. To achieve this, a range of open interfaces, flexible and customisable, have to be designed, ranging from an alarm export interface to an external alarm collector up to a bi-directional service level interface including retrieval accounting data sets and quality of service data.

Conclusions

Here we have considered only a few aspects (however the more important ones) impacting the evolution of the transmission network. These are the main criteria operators should use in deciding how to build up their future networks or enhance their installed base.

Technological evolution, availability of industrial products, safeguarding legacy networks investments all make up the operator business environment which will characterise the next few years.

Even if a computer data networks approach has some beneficial impact on the telecommunications world, a full IT mentality is not the key to success for telecos in pursuing their targets and in evolving or building up their networks.

Radical restructuring in upgrading the network is not the solution. Only evolution by gradual steps is the course to be followed to generate, in an aggressively competitive market environment, the required service quality, security and reliability.

Biography

Mario Pagani
Alcatel

Mario Pagani was awarded a degree in Physics from the University of Milan in 1970. He joined Telettra laboratories where he was involved in the design and standardisation of 8/34/140 Mbit/s PDH multiplexers. After some experience in the sales department, he became product manager for optical-fibre digital transmission systems and subsequently responsible for the strategic plan of FIAT, Telecommunication Sector. In 1990, Telettra joined the Alcatel group and, since January 1996, he has been appointed Director, Marketing and Business Development for the Transmission Systems Division. In this position he is responsible for strategies, strategic marketing, standardisation, intellectual property and marketing communication.

Dave Newman and Hill Stewart

Regulating the Next Generation of Communications Services

A View from the ACTS Programme

ACTS is the Advanced Communications Technologies and Services programme of the European Union, and has produced a significant range of results that identify the future direction of the converging telecommunication, broadcasting and IT sectors. Several results impact the development of future regulatory policies, such as the blurring of the boundaries between core and access networks, and new flexibilities for charging related to quality of service, are identified, and the challenges they present to regulators discussed.

Introduction

The ACTS (Advanced Communications Technologies and Services) programme is the focus of the European Union's research effort in advanced communications and forms part of the European Commission's Fourth Framework Programme of research and technical development. It has made significant contributions to the development of information and communication technologies (ICT) in Europe. A major feature of the ACTS programme has been the testing and validation of advanced technology in trials and experiments, many involving real users. The impact on business, on societal and regulatory issues has also been analysed and contributions made to the economic, socio-political and regulatory debates.

Leading-edge users, service providers, network operators, broadcasters, manufacturers and academics have pooled their knowledge and resources in pursuit of specific research and trials objectives defined by the ACTS workplan. ACTS has brought together over 1000 such bodies from all over Europe, sharing a European Commission contribution of around 680 million ECU. A typical project receives 50% funding from the Commission and is carried out by a consortium of partners based in several different countries. The programme is directed by the Directorate DGXIII of the European Commission.

ACTS projects are grouped into six technical domains:

- Interactive Digital Multimedia Services (including interactive digital television),
- Photonic Networks,
- High Speed Networking (including ATM-IP integration),
- Mobility and Personal Communications Networks,
- Service Engineering, Security and Communications Management, and
- Horizontal Actions and Supporting Projects.

Projects also take part in *chains*. Chains are groups of projects from different parts of the programme and they are set up to examine issues of strategic interest. Relevant projects bring together their experiences and formulate a set of messages concerning the issue. These messages are then published as ACTS Guidelines. The 20 or so chains are organised in five chain groups covering:

- broadband access evolution (BA),
- generic access to applications (GA),
- network level interoperability (NI),
- multimedia service integration (SI), and
- cable TV (CATV).

The chains have defined 80 guidelines on issues as diverse as sustainable development and interoperability of management systems. About 60 of these are now publicly available.

All the ACTS projects meet together regularly in a process known as *concertation* to review progress and pool ideas. The resulting consensus is valuable for Europe as a whole. Interoperability is an important aspect of ACTS and many projects are actively involved in developing or validating standards and open interface specifications.

The ACTSLINE project promotes the results of the ACTS programme to decision makers in business sectors, to public policy makers in politics, regulation and to administrations who have a role to play in the early launch or uptake of advanced communications services. Regulators will play one of the most influential roles in the evolution of electronic communications because they have to balance the consumers' interest in affordable services with

Dave Newman: Telscom AG, Switzerland.
dave.newman@btinternet.com
Hill Stewart: Central Research Laboratories, Italtel spa, Italy.
hill@btinternet.com

the suppliers' interest in a reasonable return on their investments.

Main Achievements of the ACTS Programme

The ACTS Programme has made significant progress in the development of both hardware- and software-oriented technologies. In some fields, Europe is leading or is taking a driving seat in the developments and in the setting of standards. A general rationale has been the development of open architectures and open interfaces, which are paramount for successful competition.

680 Million ECU buy a lot of research—the list of results is quite impressive. Highlights include:

- standards for digital broadcasting have been validated so effectively that commercial services are now up and running in several countries;
- millimetre-wave broadband radio systems are now a serious alternative to fibre or cable for broadband access;
- a range of new optical components has been produced, from wavelength multiplexing devices to high-speed transmitters and receivers;
- commercially viable strategies for delivering broadband access have been defined;
- a large number of interactive multimedia applications have been tested, often with real business or residential users;
- an architecture has been delivered for secure electronic commerce;
- ways have been identified for interworking between ATM and IP;
- management systems can now work across organisational and technical boundaries;
- an open services architecture has been defined; and
- standards for third-generation mobile services are now almost complete.

The Communications Environment

The communications environment has changed significantly over the past decade. A notable milestone was the advent of full competition at network and service level in most European countries at the beginning of 1998. This has resulted in the appearance of many new network operators and service providers. In addition, the market has reacted positively on both the supply and demand side. New products (services) and new pricing approaches are emerging and the expectations and demands of consumers are changing.

Many countries regard the Information Society as a critical issue and are developing detailed policies for supporting its evolution. These policies include regulatory, governmental and administrative frameworks, financial prerequisites, education, social and societal issues, micro- and macro-economic aspects, methods of work and employment, and the support and focus of directed research and development (R&D). Communications are often seen as the driving force in the evolving Information Society.

Background to Developing Regulation

Long-established paradigms are being superseded by new ones. For example, the previously distinct businesses of telecommunications, broadcasting and IT are converging with the advent of interactive multimedia services. Different types of distribution networks (for example, fixed and wireless switched networks, broadcast networks) offer various degrees of interactivity, so that the same, or similar, services can be delivered by a variety of platforms and infrastructure types. The capabilities of both the final delivery link to the customer and the core networks themselves, and innovation in the service provider area, are proceeding at different rates. This could lead to the development of new delivery bottlenecks. Globalisation of network infrastructures and the services provided over these networks is increasing. In these circumstances, regulation and standards development needs to be more uniform in effect and the technological neutrality of regulation generally maintained, in order to guard against artificial bottlenecks, or other forms of control, developing and being exploited.

The Internet is introducing different paradigms from those of the traditional telecommunications and broadcasting worlds, in terms of charges, quality and the ability to search extensively for material and information.

The fast-growing mobile communications sector is having an impact on contactability, and creating demand from users for convergence in the facilities and services provided by fixed and mobile connections. These developments are helping to drive forward innovative service offerings and lower prices to users.

Market developments are causing the value chains of the different communication sector vertical markets to converge. The combined chain covers content origination, content and service packaging, services provision (including navigation/brokering), infrastructure (distribution network) provision, and terminal vending. Major players with market power have the ability (on their own or via partnerships) to acquire bottleneck control of the interface between any two elements of the value chain. Critical areas are access to the distribution networks and access to the customer via the capabilities embedded in set-top boxes. Regulators are increasingly taking action to promote market entry and to limit the ability of existing players to abuse a dominant position.

The increasing scope of services on offer and the increasing numbers of suppliers in the marketplace mean that there is a continuing need to protect consumers from fraud, deception and the infringement of their personal privacy. The social needs of particular communities, the disabled and other disadvantaged individuals will also need protection in the public interest. Regulators have a role to play in ensuring that consumer protection and social needs issues continue to be dealt with in the changing and converging environment.

Regulatory Considerations

Regulators are likely to face a wide range of new issues in the evolving competitive, converging, multi-network, multi-provider, multi-supplier telecommunications, broadcasting and IT domains. These issues can be grouped under a number of headings. Those issues that have an impact on market development are related to:

- open infrastructures—enabling interconnection and interoperability across competing networks of both similar and dissimilar types;

- charging and pricing—relating to new flexibilities for charging on offer in a multimedia services environment and potential conflicts with developing regulation;
- open access to services and content—enabling users to access and communicate with service providers of their choice as the range and type of services expands; and
- numbering evolution—the long-term approach to numbering in an environment where different approaches are in use in different domains. Although this is an important area, ACTS has very little to say about this!

There are also issues more related to consumer protection and the needs of society.

Implications of ACTS

Open infrastructures

ACTS projects and trials seem to be identifying a major trend in the evolution of the access network (or local loop)—that new technology will increase its reach from a few kilometres to over a hundred kilometres. In doing so, the access network naturally splits into two fairly distinct segments; a feeder (or shared) segment that links network nodes to a flexibility point near the customer (which can be over 100 km in length) plus a distribution (or dedicated) segment that covers the last few hundred metres. This blurs the boundary between access and core networks and may change the commercial relationships between access and core network providers—indeed in smaller countries it could virtually eliminate the core network.

There is currently no formal definition of access and core networks in regulatory circles, but there is a lot of discussion about *unbundling* the local loop and this may ultimately require a clearly specified interface. ACTS has made a useful contribution by helping to define the VB5 interface—an open interface between access and core networks provided by different operators, or using different technologies. The Iu interface takes this a stage further by including mobile networks and an ACTS guideline has shown how fixed and mobile operators might usefully share the 'feeder' segment of future access networks.

Although VB5 and Iu provide clear technical specifications for the access/core interface, it is not clear that these are enough. As new technology blurs the boundary between access and core networks—and indeed between fixed and mobile infrastructure—will the balance of market power between the various players be affected? Regulators may need to consider whether effective competition will need a new market model and whether defining and policing the commercial interfaces determined by that model will require a new technical interface closer to the customer.

There are also other implications for any policies developed in relation to unbundling the *copper* local loop as new technology and its application in the access network evolves. It seems that, just at the time when a number of regulators are implementing or considering mandating unbundling of the copper loop, all sorts of new multimedia access technologies are blossoming and being implemented by various players (not just wire—or cable—but fixed radio access and cellular mobile). Regulators will need to think sufficiently through how their current policies will dovetail into future scenarios where a whole range of different technologies are deployed and various players' market power change.

Interoperability is not just about physical connectivity—it also involves interworking between management systems—an important aspect to remember in an interconnection context. ACTS is working closely with the TeleManagement Forum and EURESCOM to validate 'X' interfaces—for exchanging management information across network boundaries (regarding for example fault handling, accounting, changes to routing, adding and deleting customers).

Several guidelines have been produced on asynchronous transfer mode (ATM) and Internet protocol (IP) evolution indicating that both sets of standards will eventually provide a similar range of capabilities and allow IP applications to demand appropriate quality of service across ATM bearer networks. This means that the Internet will be able to support services normally carried over conventional telco networks, while ATM will give those telco networks the flexibility required to handle the differing quality requirements of individual services more effectively; for example, voice over the Internet, and multimedia over ATM. This again blurs the boundaries between regulatory environments—this time between the closely regulated telco world and the largely unregulated Internet.

Charging and pricing

Both ATM and new developments in IP provide for different categories of network service. This means that real-time video, where delay and latency are important, can demand better treatment than e-mail, where they do not matter. The problem is that it costs network operators almost the same to provide the 'gold plated' network service as it does to provide the 'bargain basement' one. However, efficient use of network capacity needs a mix of the different types of service. If everybody chooses the 'gold plated' service, irrespective of whether their applications need it, networks will quickly become overloaded. Additional capacity will have to be installed and paid for, even though it is not really needed. New types of charging schemes, which give customers an incentive to choose the quality of service best suited to their applications, may be needed.

ACTS projects have investigated and tested such charging schemes and the message is that they should not be cost-based, and probably not distance-based either. They define new parameters for an environment where quality is an important feature of the network service. Apart from challenging the principle of cost-based pricing, these schemes raise other questions for regulators. One is that the balance of market power between actors is very sensitive to some of these parameters.

Another is that they offer the possibility of a real-time auction of network capacity. Customers are familiar with peak and off-peak tariffs but ATM charging parameters could be adjusted second by second in response to demand for capacity to control network congestion at least at the wholesale level. It is not clear whether such an approach would be good or bad, as it might drive peak usage prices up significantly, for example. There is general recognition that ordinary retail customers should be shielded from instantaneous fluctuations in this 'spot market' for network capacity, as they would not be able accurately to predict their charges—but the possibility does exist!

Moving from network services to multimedia applications, there can be a large number of players involved in the value chain. This raises the question of who is charging whom and for what. Customers are likely to be confused if too many of the individual players charge piecemeal for their part of the service—however, it will be difficult to provide transparency if one provider puts all the charges together in a single bill.

Open access to services and content

In this area ACTS probably offers regulators comfort rather than challenges. The programme has conducted many trials of new multimedia services over a range of network infrastructures—in fact multimedia is one of the technical domains of the programme.

The message is that this work has been strongly based on the principle of open standards at all levels of the delivery chain and the ACTS work has made significant contributions to the development and validation of these standards. DVB and DAVIC are two of the fora where contributions from ACTS players have been influential and it is interesting to note that these are not the traditional 'official' standards bodies but fast-track industry groupings, which agree pre-standards in time to meet windows of market opportunity.

However, it is clear that new services may well combine content with delivery and interactivity taking place using components from broadcast, fixed and mobile networks, and using IP or ATM protocols. Hence regulators will have to ensure a degree of conformity between the various regulatory regimes in this converging services environment if the marketplace is to develop competitively without artificial bottlenecks being generated and affecting players adversely who elect to deliver particular services in different ways.

Consumer/social needs

In the area of e-commerce, ACTS has been active in exploring how to implement proof of identity, payment security, etc; it has also been addressing issues of privacy, copyright and content screening. It made significant contributions to the work of the MoU group on e-commerce for small and medium enterprises (SMEs). Regulators will need to review this, and other work in this area, in order to ensure that consumers are being protected sufficiently by the developing technical capabilities. ACTS has also looked at the socio-economic impact of advanced technologies. Among a range of topics, it has produced a paper discussing whether today's strongly-established traders are likely to dominate the cyberspace marketplace also, a topic of significant interest to regulators and competition authorities.

ACTS has also been looking at the provision of advanced services to remote regions. There are no simple answers for providing multimedia services to remote areas but broadband wireless may provide an economic solution. Here, alas, there is no magic bullet. The problem of whether to, or how to, pay for such services remains a question that individual national regulators will have to consider.

Conclusions

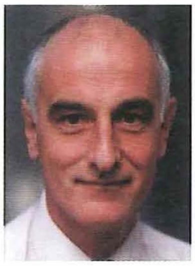
The paper has identified a number of areas where ACTS results have an impact on the development of future regulatory policies. Generally, the continuing convergence between telecommunications, broadcasting and IT technologies and services seen in the ACTS work implies a continuing need for regulators to look more widely across the electronic communications marketplace in their deliberations on policies for the future. More specifically, the blurring of the boundary between core and access networks and the new flexibilities becoming available in respect of charging are two areas where the work done within ACTS can assist evaluation of future regulatory directions.

Further information on results from ACTS, tailored to the regulatory sector, can be found by accessing the ACTSLINE web site: www.actsline.org.

Biographies

Dave Newman
Telscom AG

Dr Dave Newman joined the UK Post Office (now BT) in 1963. After carying out research on new materials and optoelectronic devices for telecommunications use, he became responsible for optoelectronic device development and, in 1984, for switching-related research and development, within BT Laboratories. Later, he managed BT Laboratories' network systems R&D; this was followed by a period of responsibility for software development for some of BT's major internal network management and customer facing systems. He joined Oftel in 1995 as Deputy Technical Director. Since 1998, he has been a telecommunications consultant, which includes working on the ACTSLINE project within the ACTS programme.



Hill Stewart
Italtel spa

Dr Hill Stewart joined the UK Post Office (now BT) in 1972. After carrying out research on optical communications technology and systems, he moved to technology forecasting and strategic planning and managed BT research programmes on architecture and standards and network and service management. His final job with BT was producing the first release of an integrated architecture for the company's networks and systems. During his time with BT he was active in supporting EU activities in collaborative R&D and, since leaving BT, has worked as a consultant on European RTD. He is currently managing the ACTSLINE project for Italtel.

Paul Nihoul

Will Electronic Commerce Change the Law?

Towards a Regulatory Perspective Based on Competition, Information and Learning Skills

I will not attempt, in this paper, to analyse the proposals currently under discussion regarding the regulation of e-commerce[N1]. My purpose is rather to examine the consequences of these proposals for the vision that we generally have about regulation; that is, the rules that are adopted by authorities to ensure compatibility of economic activities with social values. The perspective should help us understand how, and to what extend, regulation is being affected by the information society.

A Confrontation Among Actors

The basic idea is that human attitudes change with the information age. Our perspective has been influenced over a long period by an 'industrial' vision. Industrialisation is generally associated with labour and production organisation. However, industrialism also affected the law—the mechanisms which were used to ensure compliance with social values. The industrial vision of the law appears to be based on a 'macro-legal' attitude, where regulation is seen as (*i*) a set of normative principles (*ii*) imposed by an authority and (*iii*) expressing a project that members of society are to implement under a threat of penalty.

That vision has led to a systematic confrontation among economic and social actors. One example may be taken from discussions surrounding e-commerce regulation[N2]. In this debate, two positions are generally opposed[N3].

Firstly, the position adopted by businesses. As we know, businesses endeavour to minimise costs and maximise benefits. Such attitude is said to be necessary, as customers are to be attracted if undertakings want to collect resources for survival and future development. In that context, regulations are regarded as burdens. As a result of the law, undertakings are to commit resources to behaviour they would not otherwise have contemplated (for example, the protection of personal information on customers). The rules also prevent them from engaging in activities that might have proven profitable (for example, the installation of electronic devices in software in order to monitor customer behaviour and gather marketing information).

Secondly, the position adopted by customers. Cost minimising and profit maximising are not limited to businesses. Such behaviour may also be found with customers. The pattern will however be different, as an action bringing a benefit to an undertaking will often imply a cost for customers (and vice versa)[N4]. By customers, rules are generally

Paul Nihoul:
Law School,
Université Catholique de Louvain,
Rue Lannoy 15/4,
1050 Bruxelles,
Belgium.
Tel: +32 10 474635
E-mail: nihoul@cpdr.ucl.ac.be

[N1] For a global overview of the telecommunications regulation in force in Europe, including aspects of Internet regulation, see Nihoul P., Droit européen des télécommunications—L'organisation des marchés, Bruxelles, Larcier, 1999. For a similar study on US law, see Benkler Y., Rules of the Road for the Information Superhighway: Electronic Communications and the Law, Saint Paul (Minn.), West Publishing, 1996.

[N2] For an excellent presentation of the discussion related to that subject matter, see the first issue of the new review *Info—The Journal of Policy, Regulation and Strategy for Telecommunications, Information and Media*, published by The Camford Group and endorsed by the International Telecommunications Society as its membership journal.

[N3] This presentation is simplified for the sake of clarity. Groups are not as homogenous as one may think. For instance, the rules of privacy protection may be considered as a burden by consumers who might prefer a simpler—hence cheaper—treatment, contrary to what is generally suggested. Some businesses might also consider these rules as a source of profit. That is the case for the undertakings which make their business of rating the policy adopted by e-commerce companies with regard to privacy. Furthermore, and assuming that groups are rather homogenous, the status of the rule among businesses and consumers often depends on its content. For instance, many e-commerce businesses will consider the rules on privacy as constraints, but they will have a different view with respect to the rules protecting their intellectual property. Such rules will then be considered as absolutely necessary, even though they may place a constraint on the behaviour to be adopted by other economic agents—in this case the customers. Similarly, rules may be considered positive by customers with respect to the protection of privacy, while some customers will complain about the intellectual property rules which, they say, prohibit the free circulation of ideas and art.

[N4] A decrease in price will often be regarded as positive by the customers, but rarely by undertakings as it will imply a profit lower than would otherwise have been attained. By contrast, higher prices for a good or a service will require a larger share of the customer purchasing power to be affected to the operation, whereas it will mean a bigger profit for the undertaking.

regarded as remedies—not burdens. They are supposed to provide a protection against risks[N5].

As we see, different solutions may be adopted in various situations. These solutions are decided upon by authorities[N6]. The latter are supposed to occupy a central position in society. Most of them are elected, or are at least submitted to some form of (sometimes indirect) electoral control. For that reason (and probably others as well), they may be sensitive to the concerns shared by the consumers.

It has long been recognised, however, that electoral results do not only depend on the behaviour of the citizens. Money is essential in that market too. The marketing activity which is related to elections depends to a substantial extend upon the sum collected by the candidates. In many instances, the electoral money is handed over by corporations. It is thus essential for authorities and politicians in general to be sensitive as well to the ideas—and suggestions—put forward by the undertakings[N7].

The Political Process

On the basis of these remarks, a general presentation may be proposed of the regulatory process as it is analysed by many observers.

- In *economic* terms, rules may be said to provide a tool for the allocation of costs and benefits among social and economic forces. Take the discussions on the allocation of damages caused by illicit information posted on the Web. Various solutions may be contemplated by law makers. The damages can be imposed on the victim, by deciding that no reparation will be granted. A payment may also be requested from the service provider, or from other intermediaries (firm providing storing facilities, etc.). The choice in favour of one option will determine who will pay the costs associated with the activity.
- The regulatory process may be described in *political* terms as a struggle to obtain legal protection. Interest groups are aware of the prevalence that may be given by the law to their position. That prevalence means that their position will be accepted, and defended with the authority of the state, as long as other groups are not in a position to articulate a coherent counter-proposal and mobilise social forces to have that proposal adopted by the legitimate representatives.

The struggle appears clearly in many economic areas. See for instance the discussions which oppose undertakings for the adoption of technical standards[N8]. The attitude adopted by undertakings may easily be understood. The 'winner'— that is, the undertaking whose innovation will be regarded as the standard—will receive payments from intellectual property rights licences[N9]. It will also be placed in an excellent position to undertake research and development for related innovations. Finally, it will command the know-how necessary to manufacture the goods and services related to that standard.

A Locked Society?

These comments have led many observers to consider that our society is somehow locked. Parties are trying to obtain legal or economic advantages to the detriment of others, with the result that satisfactory compromises are rarely attained.

A good example of that situation may be found in the discussions that have taken place in Europe with respect to the universal service in the telecommunications industry[10]. As we know, that sector was reserved for decades to national monopolies. With the advent of competition, authorities wondered about the implementation of the political objectives which were pursued under the former regime, in particular the broad dissemination of telecommunications goods and services throughout the population and the territory[N11]. They decided that political actions should be undertaken apart from competition.

The question was however to determine who would pay the costs associated with the universal service. The new entrants were rather opposed to a substantial public intervention. They feared that the costs would be imposed on undertakings, including themselves. Their concern was that they would then be hampered in their ability to compete—enter markets dominated by former operators and innovate to develop new services/technologies.

By contrast, the former telecommunications operators supported a large version of the universal service. Due to their command of the public network, they were the only ones able to perform the relevant services. For that reason, they were sure to be entrusted with the exclusive responsibility for the programme. They would thus be protected from competition on these markets and receive a remuneration which would compensate their relative inefficiency in other fields as well.

The position advocated by the former telecommunications operators was supported to some extent by certain consumer groups. The latter have traditionally been wary of changes coupled with competition. They hope that prices will decrease as a result of a change in the market organisation. At the same time, they are concerned that undertakings might be forced to cheat in a competitive environment; that is, to grab resources from customers while providing little to them.

The debate led the Member States to discuss an amendment to the EC Treaty during the Amsterdam intergovernmental conference. The partisans of both attitudes submitted proposals supporting their stance. A

N5 Let us suppose a customer wants to underwrite securities via the Internet. Several techniques may be used to that effect. One of them is to transfer funds from a bank account to another one—that of the undertaking issuing the securities. Such operation implies risks. For instance, a third party may intercept the code used to open the electronic account and later use that code fraudulently. The undertaking issuing the securities might also behave dishonestly and, for example, not deliver value in return for the funds transferred to it. These risks may be allocated by the law among the parties involved. Thus, the legislator may decide that the loss will be compensated by a special fund created to support electronic transactions. It may also consider that as a specialised institution, the customer's bank should have made sure the code would not be intercepted.

N6 The authority has been identified with the term *legislator*, but may receive a different name in other legal systems.

N7 Businesses also have an influence on politicians via the information they are able to collect, and the intelligence they are in a position to produce and disseminate.

N8 For example, the standard for the third-generation of mobile telephones.

N9 The good or service may only be used and/or produced if the other undertakings resort to the intellectual property rights which are associated with the innovation.

N10 See Nihoul P., Les services d'intérêt général dans le traité d'Amsterdam, 1998, in Lejeune, Y. (ed.), Le traité d'Amsterdam: espoirs et déceptions, Bruxelles, Bruylant, 341 s.

N11 About the European rules on universal service, see Nihoul P., Droit européen des télécommunications—L'organisation des marchés, Bruxelles, Larcier, 1999, 273 s.

compromise was reached for a new provision to be inserted in the Treaty. That provision, however, is hardly legible. It contains a summary of the positions expressed by the parties, without clear indication as to the option which was finally chosen. All parties were thus in a position to claim victory. The confrontation was not resolved, and is likely to occur whenever the question is asked again[N12].

Rules With Various Origins

Rules created by authorities

As was said at the outset of this paper, the regulation that we experience nowadays in many sectors of the economy is inspired by an industrial vision. In many respects, industrialisation is accompanied by some sort of centralisation. With industrialisation, workers gathered from villages to urban centres. Their work was organised as the implementation by many of tasks determined by a few. Such economic and social concentration found an echo in political organisations, with large territories being administered from a single place[N13]. In that organisation, the law was no exception. As observed above, the legal system was regarded as a body of binding sentences expressing a project designed by a central authority. That project was supposed to become reality; that is, to be implemented by the members of society under the threat of sanction.

Compared with that organisation, a change is under way with the development of the information society. The latter does not imply, nor rest upon, a centralisation of power. It may rather be associated with a dispersion or dissemination of power. We are witnessing the possibility of workers performing duties away from traditional professional premises. Workers are increasingly independent, as their task implies the creation of value away from a mere implementation. The political power is also affected by the evolution: it is progressively devoluted to entities closer to the citizens, both at national and international levels.

May conclusions be drawn from that evolution with regard to regulation? May we still consider that the law—and the regulatory tools—are created principally, if not exclusively, by public authorities[N14]? Or are we in a position to consider alternative sources of regulation, which may gain momentum with the information society? That perspective is essential for the development of future regulatory tools to be used in the information society. It is also important for the democracy, as the latter implies (and presupposes) the possibility for citizens to participate in the 'norm making process'.

To me, a positive answer must be given to the question. Rules are not—and should not be—created in an exclusive manner by authorities. Research is only starting in that direction, but it is essential to demonstrate that new areas are progressively discovered in that field. In this paper, three sorts of project will be shortly introduced with an emphasis on the last one.

Rules created by communities

Rules are created by communities, apart from those which are adopted by authorities. Seminal reflections may be found in that regard in the works written by Yale Law Professor Robert C. Ellickson[N15]. These works are based on a study related to a small farming community in the United States. The objective was to identify how the members of that community resolve their conflicts; that is, to determine what regulation was used in order to manage their sometimes conflicting interests. Surprisingly, Prof. Ellickson found that most of the time parties did not resolve conflicts using the laws enacted by the local and/or the federal government. They rather referred to norms which had progressively developed in their community and which remained unaffected by the official legal production.

The works initiated by Prof. Ellickson were continued in other projects. It was essential indeed to determine whether the pattern observed in a farming community could be found in other, probably more sophisticated, circles, such as those formed by businesses in the modern economic world.

A project of that kind is currently undertaken with respect to the Internet in the Berkman Center for Law and Society, at Harvard Law School. In that project, Professor Larry Lessig regards electronic software as a sort of legal code. That code is to be used by the Internet members to reach the goal they are pursuing—realise a given operation, attract customers, etc.[N16]

Another legal scholar, New York Associate Professor Yochai Benkler, looks at the impact of technical choices on behaviour[N17]. For him, the Internet is to be regarded as a medium which allows virtually everyone to post creations on the Web. By contrast, creation is severely restricted, when it comes to television, to some entities or people—basically those which can collect funds necessary to finance televisual content. For Benkler, the choice in favour of one or the other media has an impact on democracy, as it determines the ability to participate in the creation and dissemination of ideas[N18].

Rules created by individuals

Another research direction is related to the rules which are created by individuals, or by entities. In the industrial vision, society is regarded as homogenous. On one side are public authorities, which enact rules supposed to govern society. On the other side are the members of society, which are expected to abide with instructions given by public authorities. Non-compliance implies a penalty, as the objective is to ensure

N12 On the subject matter, see a. o. Dehousse, F. and Van den Hende, L., La place des services publics dans la conférence intergouvernementale, 1997, in Les services publics et l'Europe : entre concurrence et droit des usagers, Actes du colloque organisé par l'Institut d'études européennes, le Centre de droit public et la Maîtrise en management public, 13-14 Décembre 1996, Université libre de Bruxelles, 37.

N13 That was the case in France, Germany, and the United Kingdom.

N14 The question is thus not related to the determination of the entities (businesses, authorities, etc.) which should set the rules, as is often the case in discussions concerning e-commerce regulation. It is rather to identify sources of normativity, by trying to determine who sets in effect the rules that are abided with by the actors.

N15 Ellickson, R. C. Order without Law—How Neighbors Settle Disputes, Cambridge (Mass.), Harvard University Press, 1991.

N16 A description of the work which is performed in the programme is available at cyber@law.harvard.edu.

N17 See a. o. Benkler, J. Communications Infrastructure Regulation and The Distribution of Control over Content, 1998, in Blackman, C. and Nihoul, P. (ed.), Convergence between Telecommunications and Other Media: How Should Regulation Adapt? , *Telecommunications Policy*, special issue, 183 s.

N18 The choice is realised by authorities, which may encourage the development of one or the other technique. It also depends upon the community as a whole—the innovations which are made by its members and the kind of technology which is supported by the latter.

a homogenous behaviour—that which is deemed appropriate by authorities.

That representation corresponds with the perspective adopted by the authorities. In that context, regulation is seen as a project designed by authorities for society. People—who are affected by regulation—have no existence on their own. They constitute mere tools for the realisation of the public project. Thus, the industrial vision conforms in many respect with a centralised version of the regulatory phenomenon[N19].

Yet, the information society is characterised by a movement to the contrary. Rather than concentrating, it expands organisations in space and volume. With the information society, the economy, the labour and the political process appear to move away from concentration towards decentralisation. That evolution commands for the study of regulation a perspective not restricted to authorities but rather based on the views adopted by the people.

In that regard, it is essential to set aside the classical tradition in legal analysis, where rules are analysed from the authorities' perspective. A research project is currently taking place at The Telecom Unit[N20] in that direction. Our focus is placed on the process whereby individuals and entities forge their own rules on the basis of a variety of external influences, including: social pressure, threat of sanction, etc. The scheme is based on an ordinary question: 'Who really makes the law?' Concededly, public authorities adopt propositions which pretend to govern behaviour. However, the real law maker appears to be located elsewhere. Nobody can really be forced to obey the law. If he/she does not want to, he/she will always find a mechanism which will allow him/her to bypass the law—legally or not.

In that context, the ones who really make the law are not public authorities: it is the addressee which confers obligatory character to regulation by accepting and/or deciding to modify his/her behaviour in a way compatible with the rule.

Rules created by economic interactions

More concrete examples are available in the third direction which I would like to present in this paper. That third category is related to interactions among society members—business interactions in this case. The analysis is based on an example concerning competition regulation. European states now have competition laws similar or analogous to those embodied in the European Treaty (prohibition of cartels and dominant position abuses). Prior to that stage, most of them had other rules, which were said to promote 'fair competition'. Both bodies of regulation contained in effect the same kind of prohibitions, with differences as to modalities.

The example focuses on one practice considered illegal under both sets of rules: the refusal to supply a good/service if the customer does not agree to acquire another—related or not—market (tying-in). Under the rules of unfair competition, that practice was prohibited in all circumstances. The approach is different with the rules of competition, where the prohibition is only directed against dominant firms adopting that behaviour. The reason underlying that latter approach may be described as follows.

- In normal circumstances, a customer may change supplier if he/she is not satisfied—for instance if he/she is confronted with tying-in practices. An economic pressure is then placed on the supplier to change its behaviour, or face the departure of dissatisfied customers. No specific rule appears to be necessary to bring about a change in behaviour on the part of the supplier.
- The mechanism will only be effective if the customer can change supplier, should he/she be dissatisfied. That possibility is seriously hampered in cases where the supplier holds market power. The concept describes a situation where the undertaking can behave independently of the markets. It does not have to take into account the possible reactions of dissatisfied customers nor those of competitors, who are in effect unable to attract customers.[N21]

Under both sets of regulation, a negative consequence is attached to a specific behaviour. It is hoped that this consequence will force the undertaking to change behaviour. The only difference is related to the nature of the sanction (Table 1). In one case (absence of market power under competition rules), an economic sanction is applied by business actors (change of supplier). In the other ones (fair competition and presence of market power under competition rules), the sanction is applied by an authority and takes the form of a civil and/or criminal penalty.

Table 1

	Former Rules on Competition	Present Rules of Competition
Attitude vis-à-vis tying-in	Tying-in is prohibited in all situations	Tying-in is prohibited where the firm holds market power,and only discouraged in other circumstances[N22]
Sanction	A criminal and civil penalty is imposed	A criminal and civil penalty is imposed where the firm holds market power; the firm suffers a decrease in market share in other circumstances[N23]
Reason for which a sanction is imposed	1. The hope is that tying-in will be dropped by the firm as a result of the imposition of the sanction. 2. It is even hoped that the practice will be dropped before the sanction is imposed (anticipation of the sanction on the part of the firm).	1. and 2. : *idem*

N19 In that context it comes as no surprise that the legal profession has been articulated around activities carried out by authorities. Some legally educated people work for institutions. In that capacity, they contribute to a definition of the project by public authorities. Others work in the private sphere. Their job is then aimed at trying to anticipate the decisions of authorities in order to counsel clients on how to implement their strategies without impairing the implementation of that project.

N20 The Telecom Unit is a research institute which focuses on issues related to the information society. The emphasis is placed on regulatory issues, examined in an interdisciplinary perspective.

N21 That situation may be due to several reasons. Suppose that one firm commands the sole telecommunications (fixed and mobile) infrastructure available in a given territory. Customers have to deal with that firm if they want to use telecommunications services, whatever the conditions imposed by the firm. So it is for competitors who would attempt to enter the market and would be forced to use the infrastructure during a temporary period.

N22 In the present competitive world, undertaking are aware of the importance of retaining customers' confidence. Most of them understand that restrictive practices may undermine that confidence. Resorting to, for example, tying-in may cause customers to turn to other suppliers.

N23 Customers dissatisfied as a result of the restrictive practice turn to other suppliers.

Figure 1

Changes Related to the Information Age

The Regulatory Mechanism

Let us identify the mechanism underlining the approach in both sets of regulation. Three ingredients may be isolated (Figure 1).

- *Information* In both sets, the purpose is to avoid a behaviour deemed incompatible with social values. To ensure that goal is achieved, information is provided to the potential trespasser: 'Tying-in is not deemed desirable and the company resorting to that practice will face a sanction'[N24]. In the absence of such information, the behaviour is likely to be adopted as business people may have the impression that their benefits will increase if the customers are forced to buy goods or services they would otherwise have left aside.
- *Learning capacity* Information is not sufficient per se. A change will only occur if the undertaking is able to assess the situation, to understand the necessity of a change and to modify effectively behaviour on the basis of that assessment.
- *Incentive* Another ingredient must be added to information and learning skills, to ensure a change will occur in behaviour. No change will take place if the agent is not encouraged, or pushed, to do so. The incentive may take several forms. In this case, we have examined the incentive which may be derived from the application of a sanction should the criticised behaviour be adopted[N25].

Impact of the information age on regulatory mechanism

Here comes the final word of the analysis. The ingredients examined above may be associated to a substantial extent with the information society.

- *Information* With the information society, we witness the development of electronic means of storage, analysis and transmission which all affect the volume and quality of available information. These media offer unprecedented manners to influence business behaviour—which is the purpose of regulation. A good example may be taken from the discussion concerning privacy protection for data transmission through the Internet. The proposal put forward by the US administration is to ensure respect of privacy via information provided by rating agencies, that would assess policies enacted by the undertakings which collect information about customers. That proposal appears to provide new regulatory tools which are in complete line with the possibilities opened by the information society.
- *Learning skills* A similar explosion may be witnessed in the capacity of actors—businesses or customers—to use learning capabilities. That development weighs on the design of regulatory tools for the future—especially for those which are related to the information society. The proposal put forward by the US administration, which has been briefly alluded to above, appears to be based on such process. In that proposal, customers may obtain information as to how data concerning their behaviour might be used by companies they are in contact with. On that basis, they may decide to alter their behaviour; for example, break their relation with a company that would not respect privacy.
- *Incentive* An incentive may be found in competition, which dominates the information society. Competition is greatly enhanced by new possibilities for customers to compare offers (monetary union, bargain hunter software, etc.). It is also reinforced by the pace of technology innovation, through which undertakings are forced to offer wider choice to the customers.

Yet, competition acts as a powerful incentive in the business community: it brings increasing wealth to those who can find their way in the business labyrinths (positive incentive), and sanctions to others (negative incentive).

Conclusion

With the information age, we are entering a new stage in society. That development has an impact on regulatory tools available to ensure compatibility of economic activities with social values. Regulation implies the use of information and incentive to obtain changes in behaviour. It also requires learning skills, to ensure effective change in undesired business behaviour. These ingredients are now undergoing major changes: information can be produced, collected and transmitted more easily; with competition, businesses must provide better answers to customers; learning skills ensure that economic objectives and targets are effectively attained. As a result of these changes, we are in a better position to overcome the opposition—traditional in the industrial age—between customers and businesses. That development may offer a model for the solution of other conflicts of interest as well.

Biography

Paul Nihoul
Université Catholique de Louvain and Rijskuniversiteit Groningen

Paul Nihoul is Director of The Telecom Unit. He holds a Licence in Philosophy/Letters (UCL), a Licence in Law (id.), a Master of Laws (Harvard) and a Doctorate in Law (UCL). He teaches in Belgium (with K. Van Miert), France and The Netherlands. He sits on the Board of Info—the *Journal of the International Telecommunications Society*. Formerly, he was an Attorney (New York, USA), a Counsellor of the Minister of Finance (Belgium) and a Referendaire at the European Court of Justice (Luxembourg). His last book was: *Droit européen des télécommunications – L'organisation des marchés*, Bruxelles, Larcier, 1999, 450 p.

[N24] As we have seen, the sanction may take the form of a penalty imposed by an authority. In that case, information is provided via the publication of civil or criminal laws. The sanction may also be administered by the customer, who may choose another supplier. The information is then communicated through another channel; for example, business books explaining to would-be managers the importance of customer satisfaction.

[N25] In that regard, two possibilities have been examined: the imposition of a sanction by an authority or by a customer. In both cases, the incentive take the form of an economic loss should the behaviour be adopted.

Dave Simpson

Internet Service Providers—Anarchy in the UK?

Few markets have shown more dynamism than the market for residential Internet access provision, where Internet service providers (ISPs) have to fundamentally reassess their business models on a regular basis. This paper looks at the ways in which ISPs have changed over the last few years, and presents some wild (informed) guesses as to the way that the market may develop in the future. In particular we will consider the role of telecommunications companies in delivering Internet access, and how ISPs must adapt to remain competitive in the long run.

Introduction

There are currently between 10 and 500 Internet service providers (ISPs) in the UK. The actual figure depends on your definition of ISP. This paper considers what an ISP once was and how this definition is evolving. There are lessons to be learnt by both ISPs and telcos, as well as those who have Internet-based services or content they wish to promote. We shall see that there is method to the madness rather than simply 'anarchy in the UK'.

History In The Making

As someone once said 'Events in the past may be roughly divided into those which probably never happened and those which do not matter.' One version of history goes like this:

In the mid-1960s, the latest technological advances involved time-sharing, which allowed terminal users to interact with computers and each other. Distant users would connect to the central computer using the telephone network. This was soon found to be an inefficient use of bandwidth, since the telephone line needed to be kept open even though the terminal was not transmitting data all the time. This would not have caused too much worry, except that the academics could not afford their long-distance, long duration telephone bills. So, in the spirit of innovation that is still present in the Internet today, they invented the concept of 'packet switching' which allowed multiple data streams to be broken down in to small units and interleaved over the same circuit. This removed the need for multiple telephone lines and hence brought down the cost significantly, especially on very expensive international lines.

This love/hate relationship between the Internet and telephone networks continues to this day, whether it be leased lines or public switched telephony network (PSTN) access. In the UK, PSTN charging models have significantly affected the way ISPs build their networks in order to maximise their revenue streams. This is further explored in this section.

POP along to your local ISP

The first commercial ISPs (Demon and PIPEX) started selling Internet access in the UK in 1992 by rolling out a national network of interconnected points of presence (POPs) (Figure 1). These ISPs offered customers Internet access, e-mail and newsgroups, and were accessible by either a local number (if a POP was within your local, or adjacent call area), or a national telephone number, charged at standard rates. ISPs charged a monthly subscription to customers, which was used to offset the costs of telephone lines into their modems, leased lines to make up their backbone connectivity and general operational and equipment costs. Other ISPs soon began to appear using the same charging model, often simply concentrating on specific geographical areas.

Around 1995, Compuserve and AOL entered the UK Internet market

The Obligatory Disclaimer

This paper describes the evolution of the ISP market in the UK, and in so doing mentions company names and brands. The inclusion of such names in this paper does not constitute an endorsement of their product, nor should a company's exclusion constitute the contrary. The very nature of Internet access provision means that different ISPs may be able to meet different requirements in different ways, at different times.

As can be seen by the paper, the ISP market is a dynamic market, and by the time this paper is presented and certainly before it is published, the market may well have changed dramatically once again. As such, this paper only represents a snapshot of the market at a particular moment in time (end of May 1999), pulled together from publicly available information such as press releases, newspaper reports and public presentations.

The observations and opinions stated are those of the author and may not represent those of OFTEL. Nothing within this document should be considered formal regulatory opinion or advice.

Dave Simpson:
Senior Technical Analyst
OFTEL
50 Ludgate Hill
LONDON
EC4M 7JJ
Tel.: +44 171 634 8865
Fax.: +44 171 634 8945
E-mail: dsimpson@oftel.gov.uk

with an on-line usage-based charge in addition to their monthly subscriptions and telephone access charges. This was justified on the basis that they provide content and value-add in the form of a proprietary user-friendly interface.

Virtual POP, real benefits

In 1995, Energis, a UK telco, introduced the concept of virtual POPs (vPOPs), where calls made by the subscriber to a geographic local number could be routed across the PSTN to a central POP remote from the customer (Figure 2). Mercury, another UK telco, suggested that this type of arrangement would be more appropriate behind a local-call rate non-geographic numbering range. It was from this point that ISPs started to use 0345, 0645 and 0845 number ranges.

This arrangement benefits ISPs immensely as it allows them to offer national coverage from a single POP, and dimension their networks more efficiently. A number of ISPs chose to base their businesses in the same building as LINX, the UK's main peering facility, thus reducing their costs even further. ISPs generally kept the same charging model (subscription + local rate calls), though some used it as an opportunity to introduce cheaper subscriptions.

This move led to an increase in the number of ISPs offering national local rate coverage, which benefited customers who before only had access to a small number of locally based providers.

Share and share alike

The need to have InterConnect and Accounting Separation (ICAS) compliant interconnect rates for 0800, 0345, 0990 and Premium Rate services led to OFTEL being asked to produce the interconnect formula for number translation services (NTS). This formula is used to calculate the revenue sharing between originating, transit and terminating operators, and compared to the formula for geographic calls, gives a higher share to the terminating operator. Figure 3 shows the revenue sharing arrangements.

Terminating operators soon began to realise that Internet traffic is a very lucrative source of revenue thanks to the NTS payments for long-duration calls. As such terminating operators began to offer free connection and line rental to ISPs, to encourage them to bring traffic to their network. After a while some operators were even willing to share some of their terminating revenues in order to attract ISPs to their network.

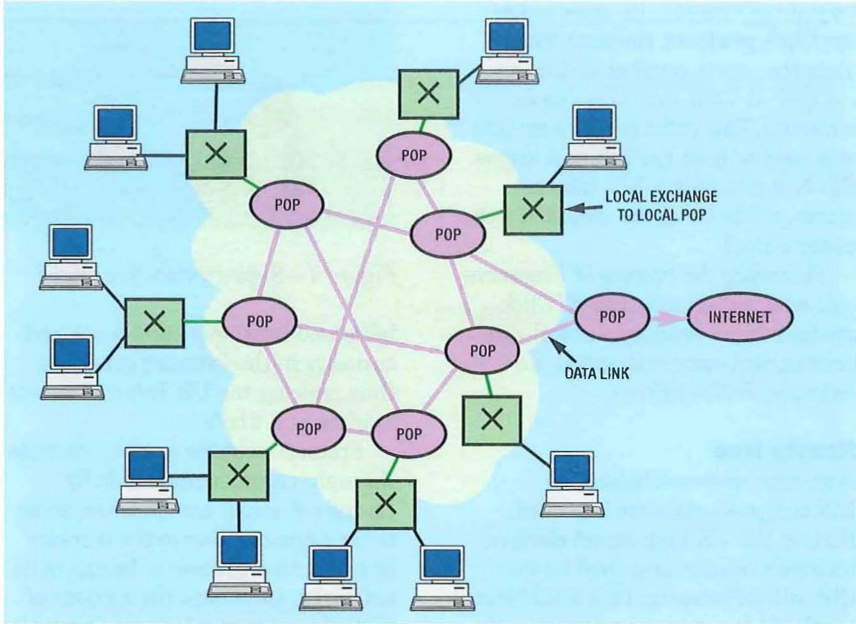


Figure 1 – Geographic POPs

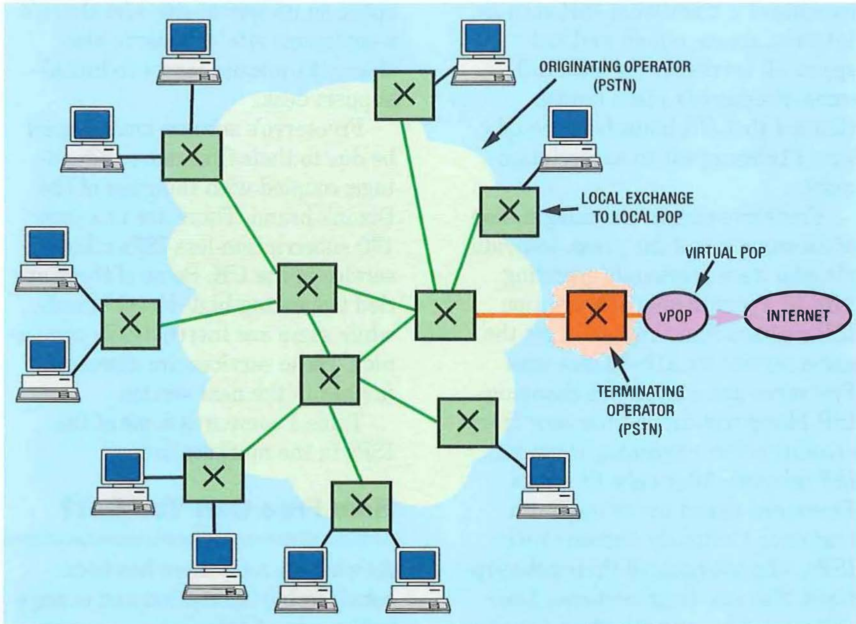
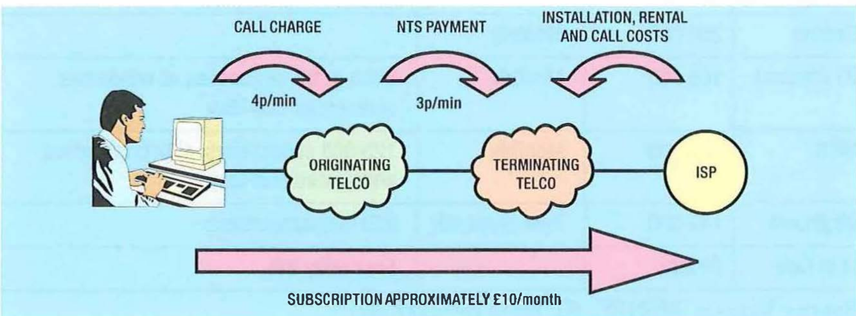


Figure 2 – Virtual POPs

Figure 3 – Number translation services



Pay as you go

On the 1 October 1998, BT launched its BTClick and BTClick+ products. This introduced a new concept of 'pay as you go' Internet access requiring no pre-registration. In order to use the Click products, the user simply dials the access number and is charged at 'local-rate plus one penny' a minute. This extra penny a minute was used to fund the Internet access. BTClick provided solely Internet access, while BTClick+ included web based e-mail.

Following the success of Freeserve (see next paragraph), the BTClick products were repriced at local-rate pricing, and some rebranded (for example, BTClickFree).

Nearly free

Just over one week before the BTClick products were launched, Dixons, the UK high-street electronics goods retailer launched its own ISP called *Freeserve*. This would have sparked little interest were it not for the fact that Freeserve was subscription free, yet still offered all the benefits of a traditional ISP, such as Internet access, e-mail and web space, all accessible for local call rates. Freeserve's press launch claimed that UK households could save £165m a year in subscription costs.

Freeserve caught the imagination of consumers and the press. Individuals who were previously spending £10–15 a month on a subscription ISP realised that they could get the same service for £10–15 less with Freeserve and considered changing ISP. Many test-drove their new 'free' account before cancelling their old ISP account. After only 18 weeks Freeserve signed up its millionth customer. Curiously certain other ISPs, who maintained their subscriptions, also saw their customer base enlarge (at least in the short term) as the hype surrounding Freeserve

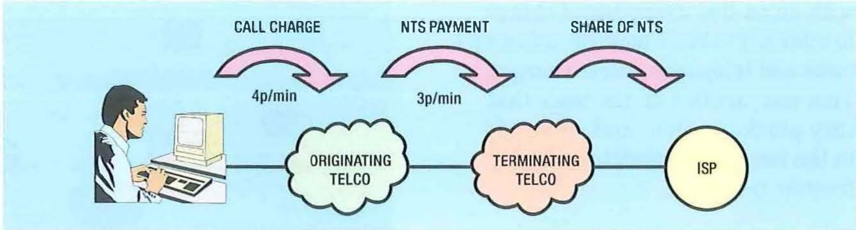


Figure 4—Subscription-free model

increased consumer confidence and curiosity in the Internet generally, thus growing the UK Internet access market as a whole.

Freeserve claims to cover its costs through out-payments made by Energis. Energis can afford to make these payments due to the increase in traffic that Freeserve brings to its networks, and hence the amount of NTS money generated (see Figure 4). Freeserve has additional revenue streams through advertisement space on its 'portal site' and Dixon's e-commerce site. Freeserve also charge £1/minute for its technical support desk.

Freeserve's success must in part be due to their first-mover advantage, coupled with their use of the Dixon's brand. There are now over 130 subscription-less ISPs offering service in the UK. Some of these are tied to existing high-street brands, while some are Internet-only companies. These services are discussed further in the next section.

Table 1 compares some of the ISPs in the marketplace.

How Free Can You Get?

As we have seen there has been considerable innovation and change in the area of Internet access, ever moving towards the panacea of completely free access. Already ISPs are fighting to be 'free-er' than anyone else. Nothing is ever 'free' of course, so the money must come from somewhere, either directly or indirectly. The rest of this section looks at how ISPs might achieve this in practice. Some of these models are already being tested by some of the big name players.

With all these services, once one ISP has offered a completely free product, commercial pressures mean that other ISPs have to match the offer, or provide value-add elsewhere. These pressures will be discussed in the final section of this paper.

One (monthly) price for as much as you can eat

So what sources of revenue might an ISP use to cover the cost of Internet access provision? ISPs could of course charge subscriptions to their customers in exchange for unlimited access, perhaps behind an 0800 number range. In this model the calls made by the customer are paid for by the ISP, offset against the monthly subscription. The obvious issue here is that the ISP must set the subscription rate high enough to cover the costs of the 0800 access (and hopefully make some profit), but not so high as to be unattractive to consumers.

An ISP may decide to set the subscription rate equal to subscribers' expected monthly usage. In this case it is especially important to set the subscription rate at a price point where low-users are attracted to the tariff, otherwise only high-users will join, skewing the average and forcing the ISP to make a loss.

ISPs have to balance the time-based outgoing charges for the 0800 number against their expected flat-rate subscription payments. In this model, the ISP carries a high level of risk, especially where it is known that subscribers' usage increases considerably, often by 400%, when they are using an unmetered access provider.

That said, a number of ISP are experimenting or already offering

Table 1 State of the market 1 April 1999

ISP	Subscribers	Subcription?	Updates (end May)
Freeserve	1·3 million	Free	1·5 million subscribers quoted in some publications
AOL + Compuserve	900 000	Monthly	Sub rate reduced from £14·95 to £9·99 announced end May
Demon	250 000	Monthly	
BT Internet	165 000	Monthly	0800 (toll-free) access at weekends announced end May
MSN	160 000	Monthly	125 000 subscribers, subscriptionless announced end May
Virgin.net	145 000	Free (April 99)	250 000 subscribers
Line One	70 000		Free (May 99)

Source: Various (MSNBC, FT, press releases etc) limited 0800 Internet access in exchange for a monthly subscription. Most notable in the last few days BT Internet has announced free access at weekends, and AOL is known to be trialling an option, both of which require monthly subscriptions.

Adverts

One ISP offered subscriptionless Internet access even before Freeserve. In March 1998 an ISP called *X-Stream* offered a subscriptionless service, where the only cost to the consumer was the local-rate telephony access. In exchange for this privilege, X-Stream placed banner ads along the top of the user's screen. In March 1999, when main stream ISPs were offering subscriptionless services, X-stream announced it would offer completely free access to its service at particular times of the day via an 0800 number, in exchange for customers displaying its banner ads. Following this X-stream announced it had now reached 305 000 subscribers.

Call revenues

Another model being explored is that of giving away completely free access to the Internet in exchange for customers changing telephone supplier. The concept here is that the profits of non-Internet calls can be used to offset the cost of Internet calls. When one supplier, LocalTel, in conjunction with Tempo (another high street electrical goods retailer) launched *Screaming.net* there were reports of brawls in stores as people fought to get hold of the sign-up CDs! Such was the interest that 15 000 CDs were distributed in the first week.

If you can't beat 'em, buy 'em

An ISP could reduce costs in a number of ways. One such way is to reduce the costs of the underlying network. This is one of the key drivers as to why telcos and ISPs are so closely affiliated in the UK, since by being an infrastructure provider, real cost savings can be made. When we look at the list of the top ISPs we find that nearly all of them, if not all of them, are owned or closely affiliated to telcos. This has not always been the case, as the telcos were relatively late arrivers to the Internet party. They made up for their late adoption by buying ISPs, together with their customer base, and in so doing injected capital funds in to needy ISPs, many of whom were still to make a profit.

Really virtual ISPs

In addition to running their own ISP services, a number of ISPs and telcos are selling solutions which enable anyone to set up an ISP without ever having to own any infrastructure. BT, Telinco, Easynet and others all have products which have enabled over 130 companies to enter the free ISP market—toy shops, football clubs, supermarkets, banks and newspapers all act as ISPs. Anyone with a brand or who represents a community has got involved. Estimates are that one new ISP is created every week.

The Identity Parade

The consumer now has a vast number of ISPs to choose from, all offering very similar services. With little to differentiate services other than a brand name (which in some cases is all that is needed!), ISPs are looking for new areas to set themselves apart. Even the cost of technical support is used as a differentiator. Some ISPs are even branching in to other telecommunications services, such as fax/e-mail gateways, voicemail and personal numbering.

Others, such as Freeserve see Internet access as a way of getting consumers to their on-line shops, and see e-commerce as both value-added content and an additional revenue stream.

Free and easy?

The subscriptionless model also creates its own problems for ISPs. Instead of having one ISP, many people have registered with multiple free ISPs, and may use only one or two regularly. This causes problems for ISPs who may not be able to accurately predict demand for their services. It also causes problems for statisticians trying to estimate the number of 'actual' subscribers, rather than simply those who have registered.

The subscriptionless model also gives consumers the impression that 'calls cost money, but the Internet bit is free'. Consumers often consider that free equals worthless which leads to high churn rates between providers, as there is little commitment to particular ISPs. As a result there are huge first mover advantages to be made in new products through a very portable consumer base!

Innovative sticky products

Whether the ISP is free or not, it is clear that the real money in the future will come from value-added products and content. One feature that needs to be considered is how 'sticky' that product or content is. For example, consumers who are greeted by the same pictures on an ISP's portal site each time they log on, will soon lose interest and make some other web site their start page, thus reducing the ISP to simply a conduit to the wider Internet.

This is the heart of the argument—what is left for an ISP? If their business is solely Internet access then how can they compete with the telcos and their ready built national data networks?

Quicker, unmetered and always on

This trend will continue as new technologies to access the Internet are rolled out. Technologies such as ADSL, cable modems, radio and satellite offer a higher-bandwidth permanent connection to the Internet, often for a usage-independent fixed monthly fee. Digital TV set-top boxes are also likely to have integrated Internet access. Many of these solutions require a national network of local POPs to collect traffic, meaning that once again the telcos are likely to be in the best position to exploit their roll-out. Regulation will ensure that any such service that BT provides will allow access to other operators.

ISPs concentrating on PSTN access may soon find themselves bypassed by these new technologies, and will therefore need to fundamentally rethink their access strategies in the longer term. Their immediate priorities should be getting experience of using these new technologies, and working out how they can incorporate them in to their portfolio of products.

These technologies could lead to a greater distinction between basic conveyance services, and value-added higher layer services such as content and services. Already we can see dedicated on-line e-mail services and sites offering free web space, where previously these services were only available as part of an ISP's offering. A look at the top 100 web sites reveals that only a small number (most notably AOL and MSN among others) are actual ISPs. All of these factors lead to less consumer commitment to their particular ISP.

Summary

ISPs have changed significantly since the early days of local POPs run by one man in his garage. Since then we have moved to the vPOP model, and have seen innovative use of the PSTN charging model to offer subscriptionless and even free access. In the future we are likely to move back to a national network of local POPs, where broadband access is the norm.

ISPs have constantly kept up with increasing customer numbers, each customer requiring more bandwidth thanks to improvements in PSTN modem technology and integrated services digital network (ISDN), and, in the not too distant future, new broadband access technologies.

Changes in the market have shown the effect (real or psychological) of the change to the subscription free model on subscriber numbers. Lots of genies have been let out of lots of bottles with regards to 'free' access, and we will inevitably see 'free', or at least unmetered, access to the Internet as standard soon. How this is financed, and whether it is sustainable in the long run remains to be seen, especially allowing for the increased customer usage which inevitably results from 'removing the clock'.

ISPs have driven Internet take-up in the UK and should be praised for their innovation in many areas. Once again they need to consider their position to avoid being swallowed up by the new world where access and conveyance are increasingly becoming independent of content and services. ISPs need to decide which side of the fence to be on, as only the largest will be able to do everything.

Glossary

ICAS Interconnect and accounting separation
ISDN Integrated services digital network
ISP Internet service provider
NTS Number translation services
OFTEL Office of Telecommunications, UK telecommunications regulator
POP Points of presence
PSTN Public switched telephony network
vPOP Virtual point of presence
Telco Telecommunications operator

Bibliography

ABBATE, JANET. A Tale of Two Networks: Early data communications experiments in England and America. IEEE History Centre http://www.rci.rutgers.edu/%7Ejea/papers/2Nets.html

OFTEL documents of interest available at www.oftel.gov.uk include:

- Interconnection charges for pay-as-you-go Internet services (for example, BT Click, BT Click+).
- Access to bandwidth: Bringing higher bandwidth services to the consumer (local loop unbundling).
- OFTEL Consultation Paper on the Relationship between Retail Proces and Interconnection Charges for Number Translation Services (NTS formula).
- Memorandum by OFTEL for Trade and Industry Select Committee: Inquiry into Electronic Commerce) includes section on Internet access).

Biography

David Simpson
OFTEL

David Simpson is currently working as a Senior Multimedia/Telecommunications Specialist in the Regulatory Policy section of OFTEL, the UK telecommunications regulator, where he is project manager for OFTEL's Internet Project. His responsibilities include the regulatory aspects of emerging technologies including the Internet. Before joining OFTEL he worked for Cable and Wireless Communications in its Network Development Division, where his responsibilities included implementing an international Internet telephony trial, the development of a fixed radio access product and general network access issues. Dave graduated from Sheffield University in 1993 with a degree in electronics (Information and Systems) Engineering.

Felicia Holness and John Griffiths

Multiprotocol Label Switching within the Core Network

The phenomenal growth of Internet traffic is attributed to the increasing acceptance of the Internet protocol (IP) as the predominant protocol in enterprise networks. The traditional design of a router upon which the Internet is based is becoming unable to support the current demand from users. This service has worked exceptionally well for non real-time traffic such as electronic mail but for real-time traffic such as voice and video, IP has only performed well across a lightly loaded network. To provide services for real-time traffic, new classes of services are being introduced into the Internet.

For networks to be able to support this increased user demand for more bandwidth and for multimedia and real-time traffic, IP routers need to be faster and to provide quality-of-service (QoS) guarantees. This demand for more bandwidth and QoS support has fuelled interest in the use of asynchronous transfer mode (ATM) as an underlying link-layer technology in the Internet. The aim is to exploit the potential benefits of ATM while maintaining the inherent strengths of IP.

Hop-by-hop routing normally works well for non-delay sensitive data traffic but, in cases where prioritisation is required, it gives unacceptable performance since the current Internet is unable to differentiate between different sorts of packets, and thus stipulate that the network provide designated levels of servicing. Since voice and video are delay sensitive they will add additional pressure to the routers. Additionally the increase in traffic is creating the need for higher bandwidth which the packet-forwarding model of router processing cannot provide effectively. The goal would be to expedite movement of information without forcing users to abandon router-independent protocols. Ideally, one would want to move the forwarding function of routing all the way out to the periphery of the network, leaving a protocol-independent network core that would be focused on providing reliable high-performance standards-based connections.

The emergence of the multiprotocol label switching (MPLS) architecture[1] aims to improve the scalability and performance of the prevalent hop-by-hop routing and forwarding across packet networks. Its primary goal is to standardise a technology that integrates the label swapping forwarding paradigm with network layer routing. This label swapping is expected to improve the flexibility and robustness in delivering new routing services without alteration to the forwarding paradigm. The word multiprotocol *suggests that this architecture will be applicable to any network-layer protocol in addition to IP; label switching indicates the underlying forwarding mechanism. This paper will start with a basic tutorial on IP switching techniques and then explain how it can be utilised to accelerate IP packets through the core of the network. The paper concludes with the description of a proposed architecture, quality-of-service label distribution protocol (QoSLDP) whose objective is to provide bandwidth guarantees.*

Introduction

An IP switch is a device that can forward IP packets at layer 3 as well as switching packets at layer 2. The mechanisms used by an IP switch enable it to be able to classify packets which will be forwarded at layer 3 and which will be switched at layer 2 and then to redirect those packets over a layer 2 switched path[1,2].

The device consists of an IP controller running network layer routing protocols (for example, BGP, OSPF, RIP) on top of a switching component. Most devices use an ATM switching fabric where the ATM switch maintains a table of input ports, input labels, output ports and labels like a normal ATM switch. There are two distinct functional components to an IP switching system: a control component and a forwarding component. The control component uses standard routing protocols to build a forwarding table. On receiving packets, the forwarding component searches the forwarding table to decide on the next hop for the packet, and switches the packet from the input interface to the output interface.

The ingress and egress components, otherwise known as a *label edge router* (LER), are positioned at the edge of an IP switching system; they can consist of software residing in a router. The concept of an IP switching ingress and egress is as follows: assume that the IP switching system decides to forward some packets over an established end-to-end layer 2 switched path. Packets arriving at the ingress are classified for either default layer 3 IP forwarding or layer 2 switching based on IP layer criteria such as IP source/destination addresses. Packets that fall into the layer 2 classification are forwarded over the established layer 2 path, while all other packets are

Felicia Holness:
E-mail: f.holness@elec.qmw.ac.uk
John Griffiths:
E-mail: j.m.griffiths@elec.qmw.ac.uk,
Department of Electronic Engineering
Queen Mary and Westfield College
University of London
London, E1 4NS.
United Kingdom
Tel: +44 171 415 3755
Fax: +44 181 981 0259

forwarded at layer 3 using normal IP forwarding procedures. Upon arrival at the egress, either at layer 2 or layer 3, they will be passed to layer 3 for normal IP processing and forwarded to the destination.

In conventional IP routing, packets travel from a source to the destination by hop-by-hop forwarding through routers. At each router hop the IP header is examined and the next hop decision is made on a number of functions such as the destination address, TTL counter, and so on. This introduces delay for packets traversing the path.

The layer 2 switched path bypasses intermediate layer 3 hops from source to destination; this path is called a *cut through* or *short cut* path[3]. The basic functions of a cut through path are:

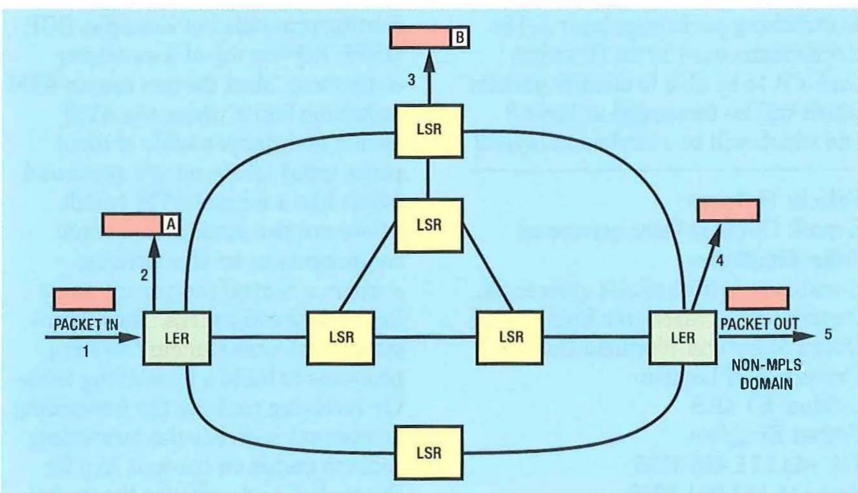
- it bypasses intermediate layer 3 routing;
- if the cut through is removed, packets can still be forwarded by conventional hop-by-hop routing;
- it can be point-to-point, point-to-multipoint or multipoint-to-point; and
- a cut through path can be built based upon data or control traffic.

Label Distribution Protocol

The main component of the multiprotocol label switching (MPLS) architecture is the label distribution protocol (LDP)[4, 5]; it is defined as a set of protocols by which a label switch router (LSR) communicates with an adjacent peer by means of exchanging labels. The LDP protocol is a set of procedures and messages exchanged and maintained between adjacent LSRs in the process of establishing label switch paths (LSPs) through a network by mapping network layer information to a label representing a data-link layer path. Each LSP has a forwarding equivalence class (FEC) associated with it; at the ingress of an MPLS domain all incoming packets will be assigned to one of these FECs. The classification of packets into specific FECs identifies the set of packets, which will be mapped to a path through the network. Examples of FEC elements are a complete host address or an address prefix.

The LDP protocol has four types of messages, discovery, session, advertisement and notification. The cumulative operation of these messages allows LSRs to indicate their presence in the network to establish and maintain sessions between peers and to perform operations on the labels they exchange between themselves. A label is a short fixed-length identifier used to identify a packet assigned to an FEC. Adjacent LSRs agree on the label that will be used to represent an FEC flowing in a particular direction; packets belonging to that FEC will have this label appended to it. The label can either reside within the data link header or exist between the layer 3 and layer 2 headers. The downstream LSR makes the decision to bind a label L to a particular FEC F and sends this assignment to the upstream LSR. This label is then used by the upstream node to forward all packets belonging to that FEC to the downstream node. It is these mechanisms of label assigning and distributing that are employed in an MPLS network. Figure 1 shows a simple example of an MPLS domain. (1) An unlabelled packet reaches the ingress of an MPLS domain, where the ingress node (2) examines the network layer header and decides which FEC to assign it to. Once this decision is made the packet is assigned a label representing the path it should take. The labelled packet is then forwarded to the next LSR within the MPLS domain. When it is received at (3) the LSR examines its labelling information base and uses the label to determine the next hop for the packet, and the new label that should replace the incoming label. The packet is then forwarded to the egress node using the newly assigned label. At the egress node (4) the LSR removes the label, examines the network layer header and forwards the packet from the MPLS domain into the non-MPLS network.

Figure 1—MPLS domain



Partnership of MPLS and ATM

This section outlines a number of scenarios in which MPLS and ATM can coexist; it explains the problems with the various implementations and direction in which this research will take.

It is possible for MPLS to operate over any data link layer, but ATM is by far the most attractive technology because it is the only link layer protocol that can transport video, voice and data with the offer of a defined QoS.

There are a number of different ways in which ATM[6, 7] can support MPLS; the first is as a peer model. An ATM switch is converted into an LSR called an *ATM-LSR* that runs an instance of the LDP protocol and network layer routing protocols, and is capable of forwarding packets at layer 3 and label swapping at layer 2. ATM-LSRs are directly connected to each other and they exchange routing protocols and LDP messages along serial links. In this instance, all ATM Forum protocols have been removed and the switches use LDP to establish LSPs through the network. The peer MPLS/ATM[3] implementation places less routing burdens on the system than an IP/ATM overlay model because each ATM-LSR only forms routing adjacencies with its neighbour, and not each router at the other end of a VC. However, in this model the loop-free QoS-based path selection and the native QoS support is lost. If one is not exploiting ATM QoS this model might seem to defeat the initial objectives.

The second example is the integrated approach in which ATM-LSRs communicate with each other through native ATM switches. ATM-LSRs may establish PVCs or SVCs using native ATM signalling to send IP traffic, control information or LDP messages over. When a stream is mapped to a VC, an identifier, VCID, is associated with it if it passes through intermediate switches. The integrated approach allows MPLS to be integrated into an existing ATM network[8].

In the third model the physical ATM model is divided into two topologies, one maintained by PNNI and the other by IP and MPLS. This requires the partitioning of the VPI/VCI label space into MPLS and ATM portions. Additionally, some ports may be configured for MPLS and others for ATM. Switch resources are then under the control of coresident, but mutually exclusive, MPLS and ATM switch control points. The switch component would run a VC routing protocol like PNNI. Thus a network employing the integrated approach would support concurrent MPLS and ATM topologies, routing and VC management. Rather than implement the overlay model that could delete VC resources, an installation of this model would be an option. This model also allows the provider to continue to provide ATM services while introducing an additional control plane that is optimised for scalable IP routing.

A number of scenarios in which MPLS can be used in ATM have been described. We have seen how the integration of IP and ATM has allowed IP to route its packets quickly through the core of the network, but the fundamental problem with the current version of MPLS is that it offers no QoS. With MPLS in its current form you would need to use the complex and cumbersome ATM Forum signalling. The authors propose a novel QoS scheme that will enable an IP switching network to reserve resources using LDP as its base.

QoS Label Distribution Protocol

In order to devise a QoS scheme it is essential to detail the component(s) that is deemed necessary. The authors believe that in order to offer more than one service level efficiently, there is the need for the 'reservation of router resources'[9].

QoSLDP expands on the fundamental component of the MPLS architecture—the LDP protocol. An ingress node will receive the application's QoS requirements in terms of ATM's QoS and traffic parameters; that is, cell loss priority, cell delay variation and cell transfer delay. The ingress node will formulate a *reservation request* message to its downstream neighbour with the requested QoS parameters. The node will carry out a connection admission control algorithm to determine whether or not it can accept this request. If it is unable to grant the request it will immediately reply with a *reservation reject* message. Upon receipt of a *reservation reject* message, the recipient will try alternative paths until it is accepted or until all possible paths are blocked.

If an LSR determines that it can grant a reservation request, it will select a virtual connection that can meet the QoS requirements and identify it by assigning a label to it, which it will record in its information base. It will then send the reservation request to its downstream neighbour. The downstream neighbour will determine if it can accept the request; if it can it will assign a label identifying the virtual connection. *Reservation request* messages will continue to propagate through the network until an egress node for the flow has been reached.

When a node realises that it is the egress node for a particular path, it will reserve a VC and send a *reservation accept* message to its upstream neighbour setting a field in the label indicating that this is the egress node for that particular flow.

Upon receiving a *reservation accept* message, intermediate nodes will record the label received from their downstream neighbour alongside the reserved virtual connection.

When the ingress node receives a *reservation accept* message it checks the label for an indication that its request has been received by the egress node before assigning all packets belonging to that flow with the label. An end-to-end virtual connection has been established for that flow to transverse with the requested QoS parameters.

A node that is unable to process a *reservation request* message will not propagate the request message to its downstream neighbour; it will instead send a *reservation reject* message to its upstream neighbour.

A node only propagates a *reservation accept* message to its upstream neighbour when it has received a label indicating it has originated from the egress node for the flow, thereby indicating an end-to-end virtual connection has been established for the flow. In the case where the connection is no longer required a *reservation withdraw* message will be sent to the appropriate routers.

In the QoSLDP architecture, in the event that a node is unable to allocate resources a message is immediately propagated to the initiator. If the router alongside the ingress router is unable to reserve resources it will reply with a *reservation reject* message instantly. In QoSLDP the network layer header is examined only once at the edge of the network and within the network, the routers need only examine a short label to determine where the next hop is.

QoSLDP naturally allows the encapsulation of QoS messages within LDP messaging format.

In the design of QoSLDP we have made the assumption that IP applications at some time in the future will have the ability to signal their QoS needs to the ATM layer, in terms of ATM QoS parameters.

At the moment QoSLDP is still in the experimental phase and as a result of this the authors accept that performance issues will arise. A few of the problems that can be determined qualitatively are scaling issues being introduced when a single VC is used for each flow as the number of VCs in any ATM network is limited. As a result of this the number of QoS flows that can be accommodated by any one device is limited strictly to the number of VCs available to a device. Another primary concern is how to integrate the many-to-many connectionless features of IP multicast into the one-to-many point-to-multipoint connection-oriented realm of ATM.

References

1 Rosen, E.; Viswanathan, A; and Callon, R. Multiprotocol Label Switching Architecture. draft-ietf-mpls-arch-02.txt, July 1998.
2 Newman, P.; Edwards, W.; Hinden, R.; Hoffman, E.; Ching Liaw, F.; Lyon, T.; and Minshall, G.; Ipsilon's General Switch Management Protocol Specification, Version 1.1, RFC 1987, Aug. 1996.
3 Metz, C. IP Switching Protocols and Architecture. ISBN 9 780070 419537.
4 Andersson, L.; Doolan, P; Feldman, N.; Fredette, A.; and Thomas, B.

LDP Specification. draft-ietf-mpls-ldp-02.txt, Nov. 1998.
5 Davie, B.; Lawrence, J.; McCloghrie, K.; Rekhter, Y.; Rosen, E.; Swallow, G. Use of Label Switching with ATM. draft-davie-mpls-atm-01.txt, July 1998.
6 Bruce, D.; Lawrence, J.; McCloghrie, K.; Rosen, E.; and Swallow, G. MPLS using ATM Switching. draft0ietf-mpls-atm-01.txt. Nov. 1998.
7 Clark, M. ATM Networks Principle and Use. Santa Clara, California, Frankfurt, Germany: Wiley Teubner 1996.
8 Awduche, D., Malcolm, J.; Agogbua, J.; O'Dell, M.; and McManus, J. Requirements for Traffic Engineering over MPLS. draft-ietf-mpls-traffic-eng-00.txt, Oct. 1998.
9 Keshav, Sharma. Issues and Trends in Router Design. *IEEE Communications Magazine*, 1998.

Bibliography

Dorling, B. *et al.* Internetworking over ATM. An Introduction. Prentice Hall, New Jersey, 1996.

Katsube, K.; Shobatake, Y.; Mogi, A.; Matsuzawa, S.; Jinmei, T.; and Esaki, H. Toshiba's Flow Attribute Notification Protocol (FANP) Specification. RFC 2129.

Marshall, G. Classical IP over ATM: a status report. Data Communiations, 1995, **24**(7), p. 5.

Nagami, K.; Demizu, N.; Esaki, H.; and Doolan, P. VCID Notification over ATM Link. draft-ietf-mpls-vcid-atm-01.txt, Aug. 1998.

Newman, P.; Edwards, W.; Hinden, R.; Hoffman, E.; Ching Liaw, F.; Lyon, T.; and Minshall, G. Transmission of Flow Labelled IPv4 on ATM Data Links. Ipsilon Version 1.0, RFC 1954, May 1996.

Newman, P.; Edwards, W.; Hoffman, E.; Ching Liaw, F.; Lyon, T.; and Minshall, G. Ipsilon Flow Management Protocol Specification for IP4, Version 1.0, RFC 1953, May 1996.

Rekhter, B.; Davie, B.; Katz, D.; Rosen, E.; and Swallow, G. Tag Switching Architecture. http://www.cisco.com/warp/public/732/tag/tagsw_ov.htm

Biographies

Felicia Holness
Queen Mary and Westfield College, University of London

Felicia Holness is currently studying for a doctorate degree at Queen Mary and Westfield College, University of London, which is being sponsored by British Telecommunication.
Prior to joining the University of London, Felicia obtained a BEng Honours Degree in Electronic and Electrical Engineering and a Masters of Science in Satellite Communications both at the University of Surrey.

John Griffiths
Queen Mary and Westfield College, University of London

John Griffiths graduated in Electrical Engineering at Manchester University. He joined the Post Office (later to become BT) Research Laboratories working on early PCM transmission systems, CATV and ISDN. Until 1994 he was manager of the Core and Global Networks Division at BT Laboratories at Martlesham. He is now Professorial Fellow at Queen Mary and Westfield College of the University of London, working on various aspects of modern broadband networks.

Stefano Giacometti and Roberto Mameli

Tunnelling Effectiveness in the Access Environment

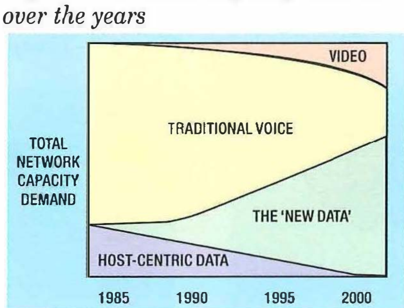
The current trend in the Internet access market is moving towards the SOHO (Small Office Home Office) target. Many analysts agree that the access market will continue to grow at a rate of approximately 30% to 40% annually through the year 2000. On the other hand, the availability of different access network infrastructures (often specialised for different kinds of media) leads to a full IP access solution able to overcome such a heterogeneity. This aspect is also justified by the ongoing convergence between telecom and datacom worlds, in which video, audio and data communications are delivered to the user in an integrated solution. In such a scenario the need for advanced access techniques arises.

This document proposes the use of tunnelling techniques in the context of a full IP access network. Tunnel mechanisms, which simulate point-to-point connections over connectionless networks, have been typically developed for applications based on virtual private networks (VPNs). IP tunnels solve traditional visibility problems that arise when we attempt to address a host within a private IP domain from a public one. Moreover the flexibility and the advanced security mechanisms offered to protect user data justify the introduction of the tunnel technology in the access domain.

Introduction

The telecommunication and networking worlds are going to be characterised by the 'convergence' process taking place during the last few years between telecommunications and datacommunications. Several technological and commercial factors justify this convergence. Historically telecommunications have been exploiting analogue connectivity, while data communications and networking have always been based on digital techniques. The development of powerful and cost-effective digital signal processors (DSPs) has decreased the distance between these worlds: many analysts consider them as one of the main reasons that will lead from circuit switching to packet switching. This 'paradigm shift' will probably take place in the medium term and it will allow bandwidth efficiency to be increased up to a magnitude order. Internet protocol (IP) telephony is certainly one of the areas in which this phenomenon is rapidly growing. By combining the best features of real-time voice communications and data processing, new kinds of applications can be foreseen. Figure 1 demonstrates this phenomenon: it depicts the redistribution of the total network capacity demand among different services.

Figure 1 – Network capacity demand over the years



In such a scenario the access environment plays a relevant role. Currently, the Internet access method for residential and SOHO (small office home office) users is based on the switched telephone network and exploits the dial-up mechanism. Users connect to an Internet service provider (ISP) by means of an analogue modem or a digital interface for integrated services digital network (ISDN) lines; when a connection is established an IP address is negotiated and, only after this phase, IP packets can be exchanged by the point-to-point protocol (PPP). Note that before address negotiation the user PC does not have an IP address at all; that is, it does not belong to an IP domain, neither public nor private. However, following the previously outlined convergence process we can easily imagine a medium-term scenario in which the IP connectivity will start directly from the users' premises, to provide integrated access to a set of services, both involving real-time and non-real-time applications.

In the previously described full-IP-access scenario some problems arise, such as those typically related to visibility, security, transparency to applications and so on. The use of proper access methods can solve all of them. A possible solution is given by tunnelling: basically these techniques are used to transfer data from one network to another one by means of an internetwork infrastructure, eventually using different protocols. The data to be transferred are typically contained in a payload packet (or frame) that is encapsulated within a transport packet and sent through the carrier network. The additional header introduced by the encapsulation mechanism carries routing information to let the

Stefano Giacometti:
Ericsson Telecomunicazioni S.pA.
Tel. +39-06-20410028
Fax +39-06-20410037
stefano.giacometti@ericsson.com
Roberto Mameli:
CoRiTeL
Tel. +39-06-20410038
Fax +39-06-20410037
mameli@coritel.it

payload packet traverse the transit internetwork. Once the payload packet has been encapsulated it can be routed between tunnel endpoints over the internetwork. The tunnel itself consists of the logical path through which the encapsulated packets travel. When the encapsulated packet reaches its destination on the internetwork, the frame is unencapsulated and forwarded to its final target. Note however that the intermediate internetwork can be any internetwork: in the context of virtual private networks, the latter usually coincides with the Internet, but this is not mandatory, since a different use of tunnelling could also imply a private transit internetwork.

This paper is organised as follows. The next section provides a technical overview on tunnelling techniques. Then the use of tunnels in the access network is explained, together with the advantages and disadvantages of this approach. Finally, the main points of interest are summarised and conclusions drawn.

Technical Overview of Tunnelling

The idea of tunnelling is not recent, since some examples have already been used in the past (for example, SNA (System Network Architecture) tunnels or IPX (Internetwork Packet eXchange) tunnels for Novell NetWare, both over IP internetworks). But recently, especially due to the interest in emerging scenarios (for example, virtual private networks (VPNs) or mobile IP environments) some new tunnelling technologies have been introduced. These newer technologies include:

- point-to-point tunnelling protocol (PPTP),
- layer 2 tunnelling protocol (L2TP), and
- IP security (IPSec) tunnel mode.

VPNs and IP mobility are two emerging applications of tunnelling. The basic idea behind VPNs is to use tunnelling in order to cross transparently a transport infrastructure (usually the Internet), making it possible for a remote user to reach its private network (for example, a corporate Intranet). In this way the remote user appears to belong to the private network domain and all the features concerning visibility, authentication and security, typically needed in a VPN scenario, are supported by the tunnelling technique. The mobile IP constitutes another application of tunnelling; in this case tunnelling is used between the foreign agent and the home agent. The former is the router in the visited network that manages visitor hosts, while the latter is the router in the home network that is aware of the current position of the mobile host. When the home agent receives data directed to the mobile host it forwards them through a tunnel to the foreign agent, which in turn delivers them to the target host. In this way the mobile host can keep its IP address (public and static at the same time) while travelling around the world.

Given its importance, it is worth providing a more detailed explanation of the tunnel mechanism from a technical point of view. Tunnels are based on the classical client-server paradigm: both the client and the server communicate by means of a tunnelling protocol, which can be either a layer 2 or a layer 3 protocol. Layer 2 protocols, such as PPTP and L2TP, encapsulate a layer 2 data unit (for example, a PPP frame) inside the transport packet of the carrier internetwork. In contrast, layer 3 protocols, like IPSec, encapsulate layer 3 data units (for example, IP packets) in an additional IP header before sending them across the IP internetwork.

A further distinction can be made between voluntary and compulsory tunnels. In the former case one of the tunnel endpoints coincides with the end user, that is, the tunnel client is located in the user's equipment, and the virtual connection is set up on demand (see Figure 2). In the latter case, the tunnel endpoint is physically distinct from the remote user equipment; thus at first a connection to the tunnel client must be properly set up and only then data coming from the user can be tunnelled through the internetwork towards the tunnel server. In this case, the tunnel client typically allows different users to share the same tunnel; that is, there are no separate tunnels for each user as in the case of voluntary tunnels (see Figure 3).

In spite of the previously outlined differences, the basic mechanism is always the same. For example, when the tunnel client has to send a payload to the tunnel server it appends a tunnel data transfer protocol header to the payload. Then it sends the resulting encapsulated payload across the carrier network, which routes it to the tunnel server. The tunnel server accepts the packets, removes the tunnel data transfer protocol header, and forwards the payload to the target network. In this way the tunnel client becomes visible in the server domain; that is, it behaves as a node physically located within that environment. Note that this process is completely transparent to applications, which are not aware of the underlying tunnelling mechanism. Tunnelling also involves some other aspects: among them a relevant role is certainly played by security issues. These include not only aspects concerning user authentication (to avoid unauthorised access to a private Intranet), but also problems related to data encryption (in order to guarantee a high level of confidentiality to transactions).

One of the main disadvantages of tunnelling is certainly represented by the introduction of some amount of header overhead. This is mainly due to the encapsulation mechanism that can be better understood by looking

Figure 2 – Voluntary tunnel

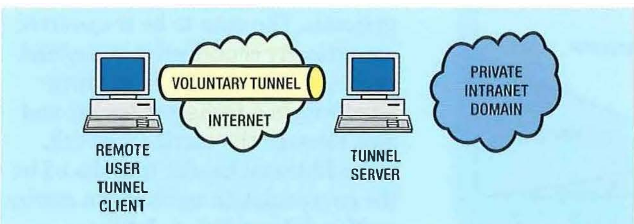
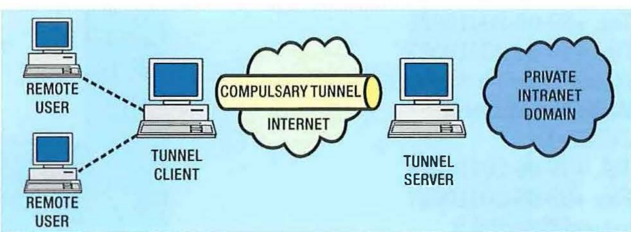


Figure 3 – Compulsory tunnel



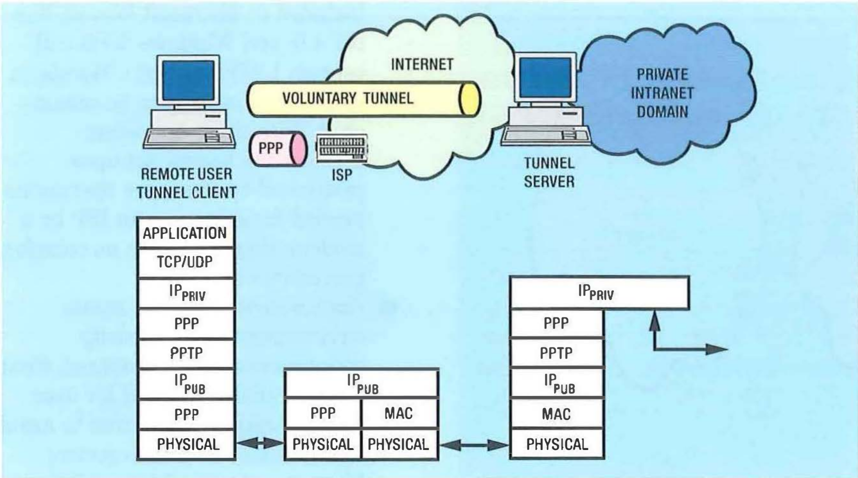


Figure 4—Access of a remote user to a private Intranet domain

at the protocol stack increase. In Figure 4, a typical scenario is represented, in which a remote user connects to its intranet by using a dial-up connection. At first the user sets up a PPP connection, by which it acquires a public IP address (that is, an IP address with a worldwide scope). After this it opens a voluntary tunnel to the tunnel server, by which it acquires a new IP address, whose scope is now limited to the private domain. From now on it is virtually located in the private domain and its packets are encapsulated (at the client side) and unencapsulated (at the server side) in order to be routed through the Internet without problems.

The tunnelling mechanism obviously involves several topics. The main ones are listed below along with a brief explanation. It is worth noticing that layer 2 protocols based on PPP usually inherit some of its mechanisms; for example, for authentication or address assignment.

- *Dynamic address assignment* This is the first and foremost problem to be solved. In order to make the tunnel client visible in the server's domain, the tunnel server must assign an address of its domain to the client. Usually layer 2 tunnelling protocols support dynamic assignment of client addresses based on the network control protocol (NCP) negotiation mechanism of PPP. In contrast, layer 3 tunnelling schemes assume that an address has already been assigned prior to initiation of the tunnel. Schemes for assignment of addresses in the IPSec tunnel mode are currently under development and are not yet available.

- *User authentication* As previously stated this is one of the main security concerns. The approach followed by layer 2 tunnelling protocols usually relies on the user authentication schemes of PPP (such as password authentication protocol and challenge handshake authentication protocol), while layer 3 tunnelling schemes generally assume that the endpoints authenticate themselves reciprocally before tunnel establishment. IPSec represents an exception, since it provides authentication during tunnel set up (by means of ISAKMP—Internet security association and key management protocol)

- *Data encryption* This problem constitutes another relevant security topic. As before, layer 2 tunnelling protocols exploit PPP-based data encryption mechanisms. For example, Microsoft PPTP implementation uses Microsoft Point-to-Point Encryption (MPPE), based on the RSA/RC4 algorithm. In contrast, layer 3 tunnelling protocols provide themselves proper encryption mechanisms. For example, IPSec defines several optional data encryption methods, which are negotiated during the preliminary ISAKMP exchange. L2TP can address this problem in two different ways: the former relies on PPP mechanisms, while the latter uses IPSec encryption to protect the data stream from the client to the tunnel server.

- *Key management* This is strongly related to the previous point. Layer 2 protocols generally use an initial key generated during user authentication and then refresh it periodically (an example of such a mechanism is given by the previously mentioned MPPE). Similarly IPSec explicitly negotiates a common key during the ISAKMP exchange and refreshes it periodically.

- *Data compression* Many tunnelling techniques also support data compression mechanisms. As before layer 2 tunnelling protocols are PPP-based; for example, the Microsoft implementations of both PPTP and L2TP use Microsoft Point-to-Point Compression (MPPC). The Internet Engineering Task Force (IETF) is investigating similar mechanisms (such as IP compression) for the Layer 3 tunnelling protocols.

- *Multi-protocol support* Another relevant topic is constituted by the possibility of supporting multiple payload protocols. This is a peculiar characteristic of layer 2 protocols, that are able to support various encapsulated protocols, such as IP, IPX, NetBEUI and so on. In contrast, layer 3 tunnelling protocols, such as IPSec, typically support only target networks that use IP.

Tunnelling in the Access Network

As previously outlined, the tunnel technique is normally used in order to support VPNs and IP mobility. However, the new emerging needs motivated by the convergence between telecommunications and datacommunications worlds justify its adoption in the access section of the network. The scenario we refer to is the one depicted in the Introduction. In such a situation there is an IP access provider (IAP) that allows full IP connectivity to residential and SOHO users. Note however that the access provider can be administratively distinct from the Internet service provider (ISP) and, moreover, it usually supports different services. In fact the former provides exclusively physical and IP connectivity, while the latter grants primarily worldwide visibility (that is, the possibility for a host located in a private domain to interact with other parties located anywhere in the world). Other types of service can also be provided by the ISP, such as

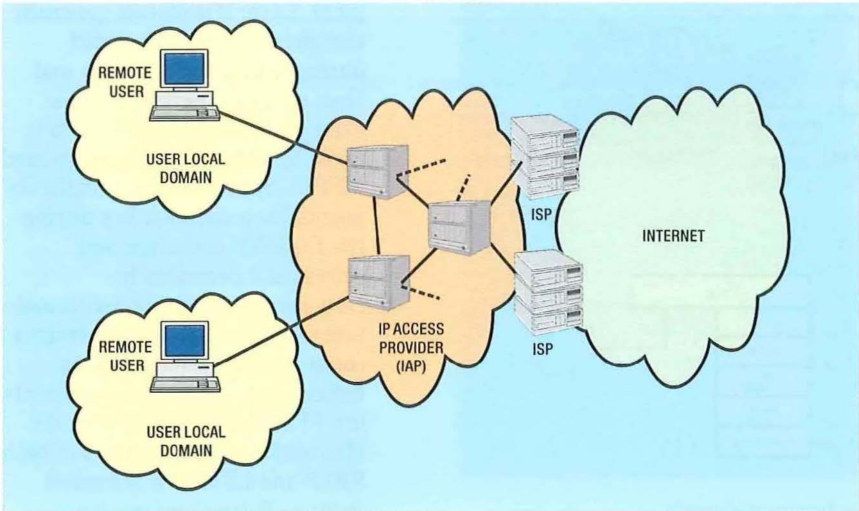


Figure 5 – Full IP access scenario

mail, proxies or newsgroup services; the choice among different ISPs may also be supported by the IAP, making it possible for the user to select the one more suitable for his purposes (see Figure 5).

Consequently, the access environment can be administered by a private entity with only a small set of available public IP addresses. To allow the growth of the access network the administrator may take advantage of the statistical multiplexing of IP addresses allowed by the dynamic address negotiation of the tunnelling mechanism. This is one of the several advantages in the use of tunnelling in a full IP access scenario; the main ones are explained in more detail below.

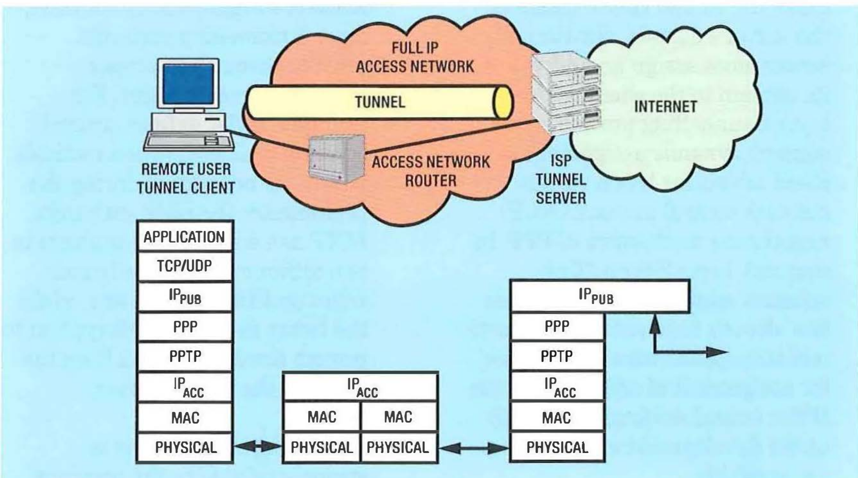
- *Statistical multiplexing and flexibility in the address re-use* The user (either a residential user or a small local area network (LAN), as in the case of SOHO environment) would be able to set up a tunnel on demand to access the Internet whenever necessary; for example, to make an IP telephone call. This is called *virtual dial up*, in contrast with the traditional dial up currently used. The main difference with the latter is, when not needed, the user is confined in its private domain; that is, with the user's private IP and with a limited scope. In this case, for example, the user would be able to make an IP telephone call with a scope limited to the access domain, thus without the need for a public IP address.
- *Transparency to applications* The tunnel is completely transparent to applications, which do not need to be modified to adapt to the tunnelling mechanism underlying. This happens because no address translation is performed and the visibility is obtained by packet encapsulation: some applications, for example, FTP data transfer, insert the IP address inside the packet payload, and a suitable address translation mechanism should be aware of it. In such cases we should have an address translation mechanism specific for each application, with all related problems of complexity and modularity. In contrast, the tunnelling mechanism encapsulates packets on the client side and extracts them at the server side in a completely transparent way.
- *Availability and user friendliness* Another advantage concerns the user, who would not be required to perform an expensive upgrade of its equipment, since newer operating systems include tunnelling support (for example, PPTP is included in Microsoft Win 98/Win NT 4.0, and Windows 2000 will contain L2TP support). Moreover, user friendliness is to be considered. With these operating systems the tunnel set up is performed by the same operations needed for accessing an ISP by a modem, therefore with no complex procedures to learn.
- *High security* In the access environment some security concerns must be considered. First of all a reliable method for user authentication is required to avoid unauthorised access to service. Moreover, the need for confidential transactions may require data encryption. Both these aspects are covered by tunnel implementations; this is especially true if referred to IPSec, which is a protocol designed mainly for security purposes.

In the near term the previously mentioned advantages, along with the reliability and the widespread use of the tunnelling mechanism, justify its adoption in the context of a full IP access network. It represents a simple and effective solution in view of the rapid deployment of the scenario depicted in Figure 5.

There are obviously some open issues that should be addressed in order to achieve better performance. The main one concerns the overhead introduced by tunnelling; Figure 6 shows the situation obtained using PPTP as the tunnelling protocol in a full IP access scenario.

The situation is slightly different from the one depicted in Figure 4. Now the user does not have to set up a PPP connection to an ISP in order to acquire a public IP address. In fact, thanks to the full IP access

Figure 6 – Overhead introduced by a PPTP tunnel in the access



network, the user already belongs to an IP domain and therefore has an IP address (even if a private one); this suffices in the case of transactions within the access domain (for example, an IP call or a mail message to another user in the same domain). In contrast a tunnel must be properly set up to reach the public Internet (for example, for a data transfer from a content provider located outside the domain). Note that in the latter case the tunnel server is located in the ISP, which is therefore responsible for assigning a public IP address to the user, while routers in the access network are transparently crossed by the tunnel. The overhead obtained is due to the protocol stack increase caused by the encapsulation mechanism. It can be reduced by proper header compression techniques, such as those proposed for L2TP, or otherwise by using a layer 3 tunnelling protocol (for example, IPSec), which introduces fewer overheads. Along with this problem there are some other minor aspects to consider, that are outlined in the following paragraph along with a brief explanation.

Conclusions

In this paper a novel use of tunnelling techniques is proposed, quite different from the classical applications in which tunnels are used to cross a public network infrastructure, either to reach a mobile host (for example, in the case of IP mobility) or a private domain (for example, in the case of VPNs). In more detail the idea is to use the tunnelling technique to transparently cross the private access environment (that, as previously mentioned, should be supposed full IP) in order to reach the public IP network (for example, Internet). The tunnelling mechanism well adapts to this purpose, since it allows advantages in terms of:

- statistical multiplexing and flexibility in the address re-use;
- transparency to applications;
- availability in standard operating systems and user friendliness; and
- high security.

In spite of all these features, there are some disadvantages to be kept in mind. Tunnelling introduces some amount of overhead, which limits the bandwidth for user data. In an access environment, where typically resources are limited (especially a wireless one), this can represent a problem. Moreover additional time for encapsulation, unencapsulation and processing of tunnelled packets is required, possibly introducing unpredictable delays that could affect end-to-end delay (this is especially troublesome for real-time traffic). Finally, the implementation of QoS mechanisms (for example, following the differentiated or the integrated services paradigm) on tunnelled flows may not be simple. This can be easily understood by observing that both these QoS mechanisms exploit the knowledge of information contained in the internal IP header, which is hidden in the encapsulating packet.

None of these problems represents an insuperable obstacle; for example, the overhead due to the encapsulation mechanism can be reduced by proper header compression techniques or by using a layer 3 tunnelling protocol. Moreover delays introduced by the encapsulation-unencapsulation process are usually negligible if compared to the end-to end delay experienced by the tunnelled flow. Finally the problem of supporting quality of service for tunnelled flows is currently under study by IETF: some solutions have already been proposed for the use of resource reservation protocol (RSVP) with IPSec and for a differentiated service extension for L2TP. All the reasons explained above justify the adoption of the tunnel mechanism in the access environment; tunnels provide a fine and easy way to deploy a solution to some of the intrinsic problems of a full IP access structure, such as those related to addressing and worldwide visibility.

References

1 Casey, L. An extended IP VPN Architecture. draft-casey-vpn-extns-00.txt, Nov. 1998. http://www.ietf.org/
2 Hamzeh, K.; Pall, G. S.; Verthein, W.; Taarud, J.; Little, W. A.; and Zorn, G. Point-to-Point Tunnelling Protocol (PPTP). draft-ietf-pppext-pptp-10.txt, April 1999. http://www.ietf.org/
3 Townsley, W. M.; Valencia, A.; Rubens, A.; Pall, G. S.; Zorn, G.; and Palter, B. Layer Two Tunnelling Protocol (L2TP). draft-ietf-pppext-l2tp-15.txt, May 1999. http://www.ietf.org/
4 Gupta, V. Secure, Remote Access over the Internet using IPSec. draft-gupta-ipsec-remote-access-01.txt. Nov. 1998. http://www.ietf.org/
5 Virtual Private Networking: an Overview. http://msdn.microsoft.com/workshop/server/feature/vpnovw.asp?
6 Remote Access: Understanding and Implementing Virtual Private Networking (VPN) Services. White paper, Bay Networks, http://www.baynetworks.com/products/Papers/2746.html

Biographies

Stefano Giacometti
Ericsson Telecomunicazioni S.p.A.

Stefano Giacometti received his degree in Electronic Engineering from University of Rome 'Tor Vergata' in 1997. He was then with CoRiTeL as a scholarship holder. His research interests focused on architectures and signalling protocols for broadband networks and integration of IP and ATM. He then joined CoRiTeL, being involved in ABR congestion control for ATM flow via satellite. He now he works for Ericsson Telecommunications, Reasearch and Development department, being involved on high-bandwidth access and on real-time traffic on IP networks.

Roberto Mameli
CoRiTel

Roberto Mameli received his degree in Telecommunication Engineering from the University of Rome 'La Sapienza' in 1997. Since September 1998 he has worked for CoRiTeL, where he takes part to the research activity of the SOFIA group (SOlution for Full IP Access). Within this project he is mainly interested in the investigation of problems related to quality of service and aspects concerning wireless access techniques and mobile IP issues.

Luciano Gavi

Service Level Management for IP Networks

This paper gives a general overview of the service level management (SLM) approach, which can offer significant advantage to service providers competing in the current European market. Starting from a theoretical approach, concepts are then applied to telecommunications scenarios focusing on the provision of data services and the Internet protocol (IP). Major concepts related to service level agreements (SLAs) and SLA management systems are analysed in greater detail defining how such concepts can be applied to IP networks. Finally some details are given concerning possible parameters to be applied to the provision of IP services, including both permanent and switched services.

Introduction

In the current European telecommunications environment the advent of a liberalised market gives rise to complex scenarios where a number of players are involved in providing services to end-customers. Players include network carriers ensuring the availability of infrastructure, network operators offering basic transport services, service providers providing basic and/or advanced services, service brokers etc.

The study of the complicated interactions that may occur among different players in providing telecommunications services to end-customers is outside the scope of this paper.

Luciano Gavi:
Systems Division,
Sirti S.p.A.
Cassina De' Pecchi
Milano, Italy
Tel: +39 02 6677 3699
Fax: +39 02 6677 3215
E-mail: L.Gavi@sirti.it

What is worthwhile mentioning here is that organisations tend to categorise their telecommunications assets and activities in terms of services that are supported and relevant service levels that can be monitored. Analysis of service levels means managing, in a business oriented perspective, the network technologies and applications adopted as well as the administrative and organisational structure that enable service life cycles.

This paper starts by describing some fundamental concepts of service level management. Then such concepts are applied to the telecommunications environment, focusing on data networks. The analysis is not intended to be exhaustive but the paper tries to give some flavour of the importance that SLM is getting for data networks. Some details are given about service level agreements and related management systems that will soon represent key issues in the support of IP services.

Basic Concepts of Service Level Management

The aim of this section is to describe the SLM approach and related concepts. Performing SLM means describing the activities carried out in terms of services made available to some specific customers. For each service some parameters are identified that allow the definition of metrics and service levels. Thus the monitoring of the parameters associated to a service gives a measure of the performances related to the service being provided.

This approach focused on SLM gives rise to substantial benefits in terms of operational efficiency and effectiveness, allowing a continuous comparison of service levels achieved against levels targeted or requested by customers.

SLM can be applied to different environments. For instance the provision of IT services within an organisation can be managed using SLM. SLM can be applied also to other fields like transportation and energy. In the following more details are given about SLM in general. Then SLM concepts are applied to the telecommunications field.

Defining service level management

To implement SLM the following aspects should be taken into account:

- definition of service level components or *objects* that can represent the service with a number of parameters showing how well the service is provided;
- definition of the SLM policy to be adopted determining rules, activities and constraints;
- identification of a proper organisation ensuring that the right people are involved at the right time to carry out SLM;
- definition of formal service level agreements; and
- identification of proper operations support systems, if any.

The definition of SLM objects can vary depending on the environment involved but the related parameters usually describe such aspects as availability and quality of service. Aspects related to the object's criticality and time-of-day relevance are also taken into account. Objects and related parameters have to be instantiated to represent real service instances.

As an example in the IT environment an SLM object can be a workstation and related parameters can be system availability and file system status.

The definition of policy comprises the specification of the processes and interactions that should run to ensure SLM operations. Processes include:

- monitoring, which describes how the SLM objects' parameters are measured and collected;
- data processing, which describes how data made available by the previous process are aggregated and elaborated;
- services maintenance, which describes which actions should be carried out (acting either proactively or in reaction to problems) to guarantee adequate and constant levels of services; and
- reporting, which describes the scope and format of data to be recorded and made available.

To enable SLM an appropriate staff of people has to be assembled, defining who is responsible for doing what and identifying a work flow that minimises possible arbitrary choices. Organising a proper staff is a hard job and can impact significantly on traditional structures, introducing new interactions and responsibilities.

A service level agreement (SLA) is a sort of contract that formally defines what the service provider commits to ensure to the customer in terms of SLM objects and related parameters. An SLA usually comprises some major SLM objects and related parameters that represent the service being provided. Considering the same example mentioned above, if a workstation is an identified SLM object, an SLA could be defined comprising that object and, as related parameters, system availability and file system status: such parameters can be measured and the following thresholds can be agreed upon:

- system availability > 99·8%
- file system occupancy < 90%

SLAs are usually associated with penalties that apply when agreed thresholds on service levels are exceeded.

Identifying clear SLA thresholds on well-defined objects and significant parameters enables a better relationship between a service provider and a customer:

- the service provider has to check continuously that thresholds are not exceeded (aiming if possible to keep some conservative distance from the thresholds); and
- the customer is aware of which level of services can be expected and is allowed to claim only if thresholds are exceeded.

To establish effective SLM, the adoption of an operations support system (OSS) can be very helpful. This becomes fundamental if SLAs are defined. Only the adoption of a proper OSS permits the service provider to monitor the service levels and adopt preventive and quick corrective actions to avoid penalties. Furthermore, use of SLA management OSSs enables the performance of the resources that are involved in the service provision to be studied, resources to be tuned to get maximum performances and differentiated levels of services for different customers to be dynamically defined.

Details about OSSs performing SLA management on telecommunications networks are given in a later section.

Service Level Management Scenarios in a Telecommunications Environment

All the SLM concepts illustrated above, including the definition of policies and the identification of a proper personnel structure, can be applied to a telecommunications environment. This paper, however, aims at focusing on the definition of SLAs and the architectural description of SLA management systems. Such aspects are becoming more and more important in the field of data networks because of the specific nature of those networks.

Basically, two main scenarios can be mentioned here that occur in the field of telecommunications networks:

(*a*) interactions occurring between an operator and the end customer; and
(*b*) interactions occurring between two operators (either network operators or service operators).

In case (*a*) the operator and the customer agree to define a formal contract referred to an SLA which clearly states all objects involved and expected levels of their parameters.

In case (*b*) the two operators interact in such a way that, at least for any specific exchange of data, one is acting as the service provider and the other as the customer. In this case the agreements that can be defined are referred to as *operational level agreeements* (OLAs). From a conceptual viewpoint an OLA is not so different from an SLA, but from a practical viewpoint some relevant differences may apply. Firstly an OLA is often tied to some extent to existing regulations and laws, while in the definition of an SLA a major role is played by commercial and marketing aspects. Furthermore, in several cases (for instance when an incumbent and a newcomer interact) service level objectives for the SLAs can be deeply affected by the service level objectives defined in OLAs.

For the sake of simplicity only case (*a*) will be considered in the following, focusing on an operator which plays the role of service provider (SP) providing services to end customers.

The adoption of SLAs has become consolidated when transmission services (for example, leased lines) are involved. In this case a few metrics can be taken into account—provisioning time, recovery time, availability time—and the models to define an SLA can be simple enough. More complex scenarios will occur when availability parameters calculated according to sophisticated algorithms (based on ITU M.2100 for transmission lines) are adopted. Similarly, when basic voice services are provided (for example, interconnection of new operators to the access network of an incumbent) some standard metrics can be defined that can easily be monitored by using capabilities available in voice switches of central offices.

Scenarios including the provision of data services are usually much more complicated. In this case the definition of SLM objects and the handling of related parameters can become difficult tasks.

A key issue is the collection and accounting of measures related to the objects' parameters. This can be accomplished in several ways:

- recording of rough data in major network components and batch transfer;
- collection of more sophisticated and aggregated data through intelligent devices; and
- use of distributed probes disseminated in strategic sites of the network

Such data can be collected taking into account that real time is not required but a reasonable time interval should be adopted if proper reactions to failures and degradation are addressed. Collection and accounting of measures data is a particularly complicated issue when IP services are involved. A specific analysis can be developed describing topics such as the identification of IP flows across a network and the implementation of application aware devices.

The need for deploying OSSs arises for data networks, as the complexity of interactions involved cannot be handled without such systems.

Modelling Service Levels Agreements

Defining an SLA involves a mixture of administrative and technical issues. Administrative issues refer to the definition of all terms and conditions that make the SLA contract applicable and enforceable including identification of validity times, responsibilities, exceptions, auditing policies, penalties, etc.

As far as technical aspects are concerned, a major issue is to define the right objects to represent a service for the purposes of the SLA. Such objects are also called *service access points* (SAPs). For instance, in the provision of frame relay or ATM services, permanent virtual circuits (PVCs) can be identified as the SAPs. Logical and/or physical local area network (LAN) interfaces can be considered when IP services are involved.

Two approaches are possible for defining and handling SAPs:

- identifying SAPs as logical entities only, related to the provision of a service; and
- identifying the mapping between a SAP as a logical entity and the underlying physical components that are involved in the provision of a service.

The approach adopted has relevant consequences on techniques that have to be adopted to collect and manage SLA data and on the general SLM policy that can be implemented.

Considering all the above aspects described above a typical SLA can be summarised as including:

- customer identification and related data,
- contract details,
- services profile and related data,
- SAPs involved and related parameters (for example, SAP weight),
- services instances involved,
- time intervals applicable and exceptions handled,
- levels of service and penalties defined depending on SAPs and service profiles adopted, and
- reports to be made available to the customer.

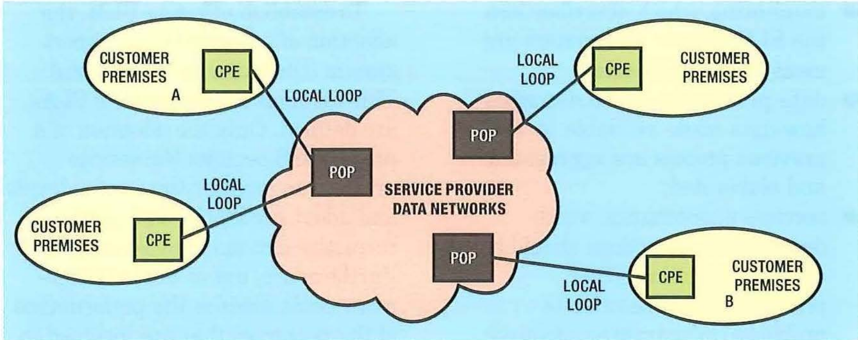


Figure 1—Architecture for data service provisioning

Focusing on data services, the following architecture can be defined to represent the provision of a service to a customer located in sites A and B (see Figure 1).

A point of presence (POP) is defined as the interface between the local loop and service provider's network. In the case of IP services a POP can be an edge router for permanent services and a network access server for switched services.

Two different situations can be envisaged to define SLA in the above scenario:

- the service provider (SP) is in charge of service support POP to POP; or
- the service provider is in charge of service support end-to-end.

The second case applies when the local loop and customer premises equipment (CPE) are under the responsibility of the SP. This usually happens when the CPE is provided and managed by the SP itself. SLA thresholds can vary remarkably from the first case.

In order to manage complex contracts, SLA management systems can be implemented, as described below.

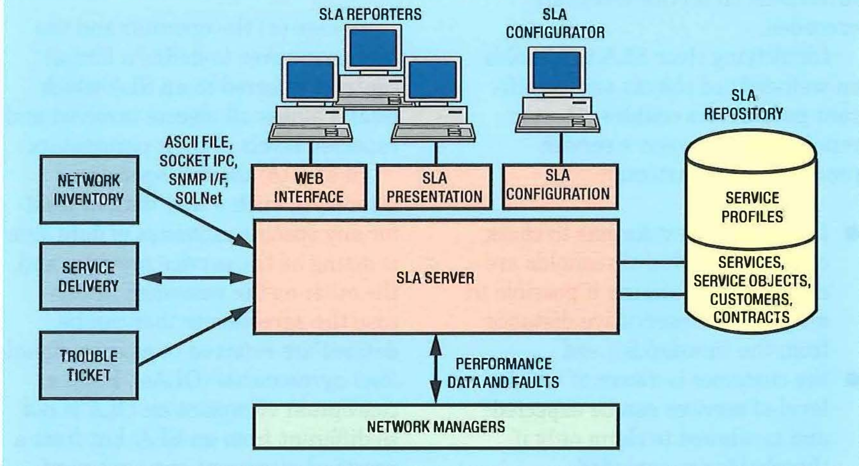
SLA management systems

Systems to carry out SLA management are now becoming more common. Such systems implement the modelling concepts described above, adopting different solutions. A general architecture of an SLA management system is given in Figure 2.

The following considerations can be applied when choosing a SLA management product:

- models implemented in the system (availability of a well-defined object model, mapping between logical entities and physical entities);
- scope of the systems in terms of networks technologies that are supported (synchronous digital hierarchy (SDH), frame relay, ATM, IP...);
- availability of a database (external versus internal, standard versus proprietary, etc.);

Figure 2—Architecture of an SLA management system



- configuration capabilities available (dynamic definition of SLAs' parameters, etc..);
- reporting capabilities available (easy definition of forms and reports, Web interface, etc.); and
- interfaces available to other OSSs (trouble ticket management, network inventory, service delivery, network performance management).

Effective interfaces with other OSSs like network inventory and service delivery are particularly important for IP services as in this case detailed information is to be exchanged concerning service/ customer features and SLA activation procedures.

If the system adopted is well focused on the services provided, significant advantages may occur including:

- automation of SLA administration activities;
- centralised control of the status of services being provided;
- full support for preventive and corrective maintenance with semi-real time monitoring;
- the possibility of defining SLAs differentiated for customers' location, time of day, etc.;
- recording of data on a dedicated repository and availability of sophisticated reporting; and
- the possibility of carrying out detailed analysis to tune network performances (capacity planning) and marketing strategies.

As far as IP networks, the adoption of a proper SLA management system, using the right models and sophisticated functions must be combined with appropriate data collection/accounting mechanisms. Implementing SLA management of IP VPNs and IP-based applications means managing high volumes of complex data.

SLA of IP Networks

The development of datacommunications networks is enabling the definition of sophisticated services for customers. Service have evolved from time-division multiplex (TDM) services through frame relay to ATM and IP services. Such technologies are much more complicated and require huge investments to ensure that networks are robust, flexible and manageable. Huge investments are however needed to compete in a liberalised market where customers can be captured only by a sound mixture of technological choices and marketing strategies.

As a result of the considerations developed in this paper investments must be applied not only to the network infrastructure but also to the definition of a proper service level management framework with a related SLA management system. This is the only approach that can guarantee a strategic advantage and a satisfactory control of cost to benefit trade-offs.

To give a flavour of the complex scenarios that have to be managed in providing IP services, a list of basic SLA parameters specific to IP services is given below (considering both permanent and switched services). Such parameters should be used in conjunction with the general parameters given in the section on 'Modelling Service Level Agreements'. Thus for instance proper measurement intervals should be identified as well as proper weights should be given to SAPs.

Parameters for permanent services:
- global availability of SAPs,
- SAP unitary availability,
- access guaranteed bandwidth,
- end-to-end guaranteed bandwidth,
- transfer delay, and
- IP packets loss/duplication rate.

Parameters for switched services:
- global availability of SAPs,
- SAP unitary availability,
- successful call rate,
- transfer delay, and
- IP packets loss/duplication rate.

In both cases, parameters related to maximum response time and maximum recovery time are usually considered as well.

Conclusions

Service level management is becoming more and more common as service providers are competing in the growing European data services market. A good approach to SLM implies the definition of an appropriate SLA and the implementation of a SLA management system. This is particularly important in the field of IP networks. However, the high level of complexity of such networks and the low operational expertise available to most service providers represent significant hurdles. Large effort is now devoted to clarifying the enabling technologies and the service scenarios that will rule the real advent of IP.

Having achieved a better understanding of the 'IP world', service providers will very soon need to differentiate from competitors and to pursue best control of their resources. At that stage definition of sound SLAs will become a must to attract customers. Implementation of SLA management systems will be necessary to manage a comprehensive portfolio of IP services.

References

All the concepts presented in this paper are the result of the work carried out by the author and his colleagues in the field of SLM over the last three years. Considerable experience has been achieved by performing requirements analysis for SLM OSSs to be either developed internally in Sirti or integrated taking third-party products. Technical documentation of all such systems (including applications for both IT and telecommunications environments) can be considered as a background of this paper. No reference to Sirti's and third party's products is given here because this is considered outside the scope of the paper.

Biography

Luciano Gavi
Sirti S.p.A.

Luciano Gavi joined Sirti in 1989 and worked about five years in the research and development department, taking part in ESPRIT research projects devoted to the development of prototypes for new telecommunications networks (MANs). From 1992–1994, he participated in two European RACE projects, respectively devoted to telecommunications service engineering and security of telecommunications networks. In 1995, he joined Sirti Systems Division, participating, as a senior system engineer, in several bids in Italy and abroad in the field of network management. From 1996–1998, he managed projects devoted to the development of network management systems and ATM networking solutions. Since mid-1998, he has been a proposal manager, responsible for coordinating all projects related to network and service management for a major Sirti customer.

Johan Kardol and Paul Kallenberg

Benefits of Network Planning and Design and Advantages of Voice Telephony over ATM in a Public Network

This paper describes the need for a business-driven network planning process and the network vision of Lucent Technologies in a fast-changing, almost exploding, world of telecommunications. At the moment operators cannot predict what the cost and profitability of new service investments will be. An integrated method has to be used to link business objectives and network investments in a continuing way. An example of the network planning process in an application of voice telephony over ATM (VToA) in the infrastructure of an incumbent operator is given.

Business Driven Network Planning Process

Investment decision-making in the dynamic revolutionised telecommunication industry environment must take into consideration the strategic value and implications of the three fundamental external forces that drive telecommunications change, namely *technology, user demands,* and *industry structure.* The selection of basic technology options is expanding and for each option there is a growing array of products from an increasing number of suppliers. Technology advances are producing continuing price/performance improvements in microelectronics, increase performance in computing and software technology and have lead to a dramatic emergence of photonics and lightwave communications.

The considerable decrease in technology cost and the associated increase in bandwidth/capacity present opportunities for the creation and delivery of high-value-added services and applications, along with opportunities for substantial increase in business volume. The growth in the on-line and Internet-based services is one (important) example of this effect. The result is a major shift of the telecommunications industry focus, from provisioning of the basic resource (that is, bandwidth) to competing in delivering high value-added services to users.

Network planning and design are key activities that have to be done before a network can be implemented. Results of the network planning and design process determine the technology and type of equipment to be used. This in turn determines the cost of the network itself and the cost of its implementation and operation.

The benefits of a well-planned network are paramount in the daily operations of a telecommunications service provider. An effective design maximises the revenue potential from the telecommunications company's network, and allows it to satisfy the demand of all its customers while minimising the required capital and operational investment.

The network planning process enables the following:

- identification of the network implications of business directions, and other driving forces such as competitive pressures and environmental effects;
- identification, analysis and justification of investment decisions in the emerging dynamic telecommunications environment, using a variety of engineering economic methods and tools to evaluate alternatives;
- identification and analysis of the implications of technology advances and timelines in terms of time of introduction, maturity, acceptability, and standards, on business processes, functions, applications, and services; and
- development of network architectures that take advantage of technology capabilities to efficiently support business application requirements.

Winning in the competitive environment of today requires an integrated network planning business model that focuses (in a continuing way) on giving feedback on the performance of the network compared with the requirements (see Figure 1). At the moment operators quite often do not know what the cost of their operations are and cannot predict what the cost and profitability of new service investments will be. Therefore an integrated method

Johan Kardol:
Lucent Technologies
Tel: +31 35 687 15 18
Fax: +31 35 687 58 39
E-mail: kardol@lucent.com
Paul Kallenberg:
Lucent Technologies
Tel: +31 35 687 1276
E-mail: kallenberg@lucent.com

has to be used to link business objectives and network investments in a continuing way.

Some aspects of this model are:

- scenario planning,
- economic valuation techniques, and
- risk analysis.

Scenario planning

The above mentioned modelling methodology is based on a frequently repeated study of scenarios narrowing down from a wide view in the early phases of a project (or network operation for say new entrants) towards minor differences in later phases of a project (or in a mature network of an incumbent operator).

It is essential to consider a number of possibilities and study all of them in some detail, and compare the scenarios with respect to strengths, weaknesses, opportunities and threats. A situation analysis has to be done for: business mission, markets attractiveness, existing competencies and challenges, service opportunities, competition situation and regulation.

To give an example for a new entrant the following considerations can be the starting point for scenario planning.

In such a case there could be great uncertainty (for example, about customers or services); solutions are chosen which are easy to change (like the use of leased lines from the incumbent)—these solutions tend to be costly for starters, but flexible.

One important aspect is coverage and hit rate, which could force starters to invest heavily without much return in the first phases; in these situations (fixed) wireless solutions can be an excellent and flexible vehicle.

When, after a while, the situation stabilises, then more cost-effective solutions can also be chosen, like the use of own fibre or exchanges.

Of importance for the scenarios is that data transmission in the early phases is the most costly part (estimated 50–70%) of the network, much more than switching. This is different from the situation in incumbent networks. Less costly (estimated 20–30%) are the interconnection bills from the incumbent for the termination of calls. Even less costly are the equipment costs (estimated 10–15%).

A main theme in all scenarios should be to find the right *balance* between (un)certainty and costs.

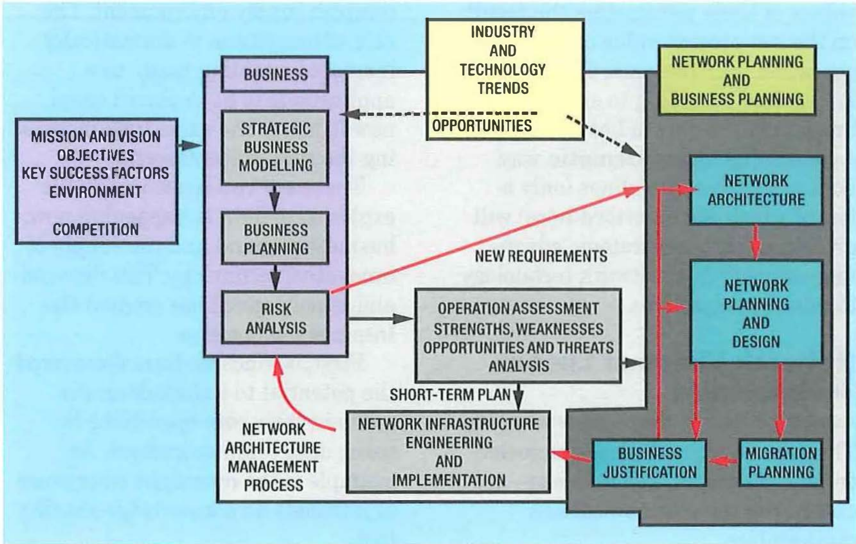


Figure 1—Integrated approach to network and business planning

Economic valuation

An economic valuation technique always used in this respect is the net present value technique and cumulative discounted cash flows. As a simple example, in Figure 2 investments of $100 M are made in year one and returns are valued as high in year 2–5. The free cash flows can be higher, but the net present value is less then $100 M taking into account depreciation, capital expenditures etc. In the example, the investment becomes profitable after about 1·5 years.

Risk analysis

The Tornado diagram (Figure 3) is a means to see very quickly what are the most important factors influencing revenues in a particular scenario. Dependant on the likeliness of change choices can be made for one scenario above the other. Tornado diagrams take into account several parameters and calculate for extreme

Figure 2—Example of cumulative discounted cash flows

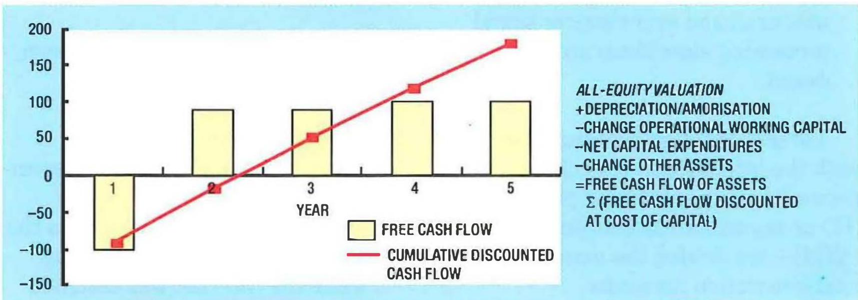
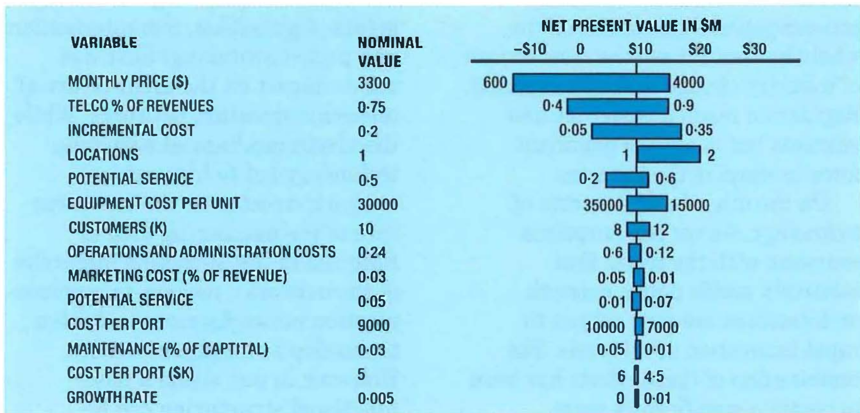


Figure 3—Typical Tornado diagram



values of these parameters the result on the net present value of the operation. Furthermore, the results are ranked according to size. Most important is to do studies like this in a continuing and **systematic** way. Following these guidelines (only a few of which are described here) will give an important strategic advantage and will link network technology to business objectives.

Network Vision of Lucent Technologies

Three basic technologies—microelectronics, photonics, and wireless—are disrupting the communications marketplace.

- *Microelectronics*—or chips—are continuing to follow the famous Moore's curve and are doubling in capacity roughly every 18 months.
- *Optical transmission capacity* is doubling every 12 months—by increasing the capacity of a single wavelength of light and by putting multiple wavelengths of light on a single fibre. Presently a 1 Terabit/s system running 400 km—the typical long distance span—is possible. 1 Terabit/s is enough capacity for all the voice calls that take place around the world at any point in time.
- *Wireless capacity*—the number of calls in a given volume of air—is doubling every 9 months as smart antennas and ever cleverer signal processing algorithms are introduced.

These technology changes, along with the introduction of packet networks—either Internet protocol (IP) or asynchronous transfer mode (ATM)—are driving the march to next-generation networks.

Two areas, technology and regulation are shaping the equipment supply environment. *Regulation* has maintained a modestly pro-competitive stance but on the whole has not caused the level or rate of industry change that was expected. Regulation made it easier for new entrants but is not the dominant force in shaping the industry.

On the other hand, the rate of *technology change* has surprised everyone with the result that relatively stable public network architectures are now subject to rapid innovation at all levels. The combination of these effects has been to create a significantly more complex supply environment. The rate of innovation is dramatically increased enabling many new applications to be deployed (good news), but at the same time increasing the rate of obsolescence.

There are two reasons why the explosion of data is happening now; business demand and the supply of innovative technology. This demand-and-supply spiral has created the Internet phenomena.

First, businesses have discovered the potential to unlock dramatic value in their core operations by using data communications. An example is the overnight emergence of intranets as a knowledge-sharing tool.

Second, on the supply side, it is now possible to build applications (intranets) using modular standardised technology popularised by the Internet.

An important implication for service providers is that the business of serving this demand will operate under new competitive rules. It will lead to a new industry structure offering aggressive prices, new network architectures, and finally an unprecedented drive to offer services across national and service provider boundaries.

Three major technological developments can be recognised, looking at the history of telecommunications since the introduction of automatic telephony:

- *digitisation,* first of transmission, later of switching systems;
- *computerisation* of switching control functions; and
- *packet technology*—in the telecommunication world, first in the signalling systems and now as the alternative to circuit switching, while the Internet was designed for packet switching from the very beginning.

The three technological developments (digitisation, computerisation, and packet switching) have had major impact on the architecture of telecommunication networks. While the electromechanical switching technology led to hierarchical network structures (the tree structure of the numbering plan is reflected in the switching hierarchy of the network), modern telecommunication networks seem to 'flatten', to develop as an organic entity. However, in our vision a more functional structuring can be recognised in modern network architectures.

First we will take a closer look at how the three technological developments have changed, and are changing, the structure of networks.

- Digital transmission technology brought high-bandwidth transmission. The transmission price per circuit is falling and the bottom price will not be reached within the coming years owing to the progress that is being made with optical transmission.
- The introduction of software program control (SPC) created the possibility for address translation. While the classical telephone network encompasses a number of functions inside the network in a distributed way (routing, charging, user identification, etc.), with SPC call control functions were split off and separated. The new network consists of a number of separated functions realised by separate units in a flexible way (call control; transport, access, signalling). A strong example of this phenomenon is the functioning of the Internet. A user can be reached on a domain name address. The domain name system (DNS) translates the given domain name into an IP number, that is used by the network for routing purposes. Thus, the computer (the DNS server) facilitates unbundled network functions by translating the user addresses (domain name) into a technical number (the IP number). In fact, this possibility can be recognised too in the concept of intelligent networks.
- Packet technology was initially applied to achieve reliable non-real-time transport over a network, where data integrity was far more important than the speed of conveyance. Efficiency, since transmission capacity is only used when needed, played an important role as well. There was a clear trade-off between reliability and speed of network transport: packet switching stands for reliable but slow (store and forward); circuit switching is fast but prone to bit faults. All this has changed dramatically. With the developments in optical technology and with ATM and the new developments in IP, the technology of real-time transport with packet (cell) switching is mastered, thus

combining the advantages of both techniques: real-time transport, efficiency in voice and data networking and flexibility for wider bandwidth applications. So the principle of packet switching may become superior to circuit switching when it comes to combining speed and reliability.

It is without any doubt that the technological innovations described above have greatly influenced the architecture, planning and design of modern telecommunication networks.

Below, the vision of Lucent Technologies on the structure of the near future structure of networks will be explained.

Rather than a hierarchical switching structure, a functional architecture will be seen. In the Lucent Technologies view on networks, a grouping in three functional levels will occur (see Figure 4):

- the *user services network and service management* level,
- the *core network* level, and
- the *access network(s)* level.

The influence of the technology drivers can be clearly pointed out in this vision.

Servers provide for address translation functions and network management (upper network level). On this level services applications are handled: voice traffic, data traffic, intelligent network (IN), IP routers etc. These servers use the underlying transport layer and the access layer for connection with the end users.

The availability of very high bandwidths based on synchronous digital hierarchy (SDH) and optical transmission technology (dense wave-division multiplexing (DWDM)), combined with the notion of packet switching being superior to circuit switching, points to the occurrence of wideband, data switched transport networks: the core network. This network has become a unified transport for different services: voice, data, video and leased lines. The question whether the concept of a unified core network will ever enclose broadcast services (for example, television), remains unanswered for the moment.

Another question is what is the nature of the future access technology. Ongoing is the process of 'fibre closer to the user': with optical access rings, fibre-to-the-business and

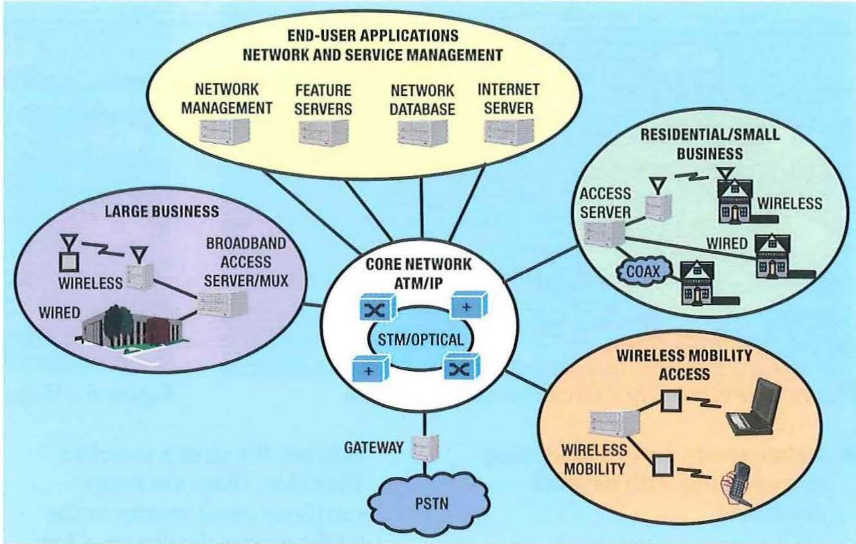


Figure 4 – Network vision

passive optical networks (PONs). Without doubt user mobility will be of paramount importance. If, as a result, wireline access will become less important over time is to be seen. With technologies like asynchronous digital subscriber loop (ADSL), commercial exploitation of copper wireline may be promising for the future.

Example of the Network Planning Process in an Application of VtoA

Changing traffic mix is driving new networks. Today, data traffic is generally carried over circuit networks that were designed for voice traffic. In the future, as the data traffic exceeds the voice traffic, voice traffic will be converted to packets and carried over packet or cell networks. This will take place because packet networks will be more cost effective and because the packet networks can more effectively carry—and combine—a variety of different multimedia services.

Data and voice traffic have different characteristics. Data traffic has different holding times, bit rate per call and revenue to the service provider per bit transmitted, which is another force driving new networks designed for the traffic specifics.

Different types of operators will have different network strategies as a function of whether they are starting from a 'voice' network or a 'data' network and as a function of the market segments they wish to serve.

At some point in the future, however, these differing networks will converge around the packet-optical core. The real trick will be transitioning today's networks around that core in a way that rich services—both existing and new—can be provided to customers.

As a consequence of the above, many incumbent operators face the question of how to extend the capacity of the circuit switched network. Increase in traffic demand caused by Internet traffic, IN services and interfacing with other licenced operator (OLO) networks leads to capacity constraints in the current circuit switched network.

One step towards a fully packetised network can be to consider planning for a *packet core backbone* network. Some of the advantages of such a plan will be highlighted in the following. Voice-over-packet technology offers new network architecture capabilities. This holds both for IP- and ATM-based networks. Currently, ATM offers better quality of service guarantees than IP. Therefore, in this context of the future evolution of voice networks, we focus the discussion on VToA.

Current circuit switched toll networks usually consist of a full mesh of toll switches (see Figure 5), typically in the order of 40–70. This leads to a number of logical routes of the order of 1000. In the voice-over-packet architecture, devices consisting of three components will provide the toll switch functionality:

- *voice gateway*, for circuit/packet termination, packetising digit collection;
- *signalling gateway*, for signalling interworking; and

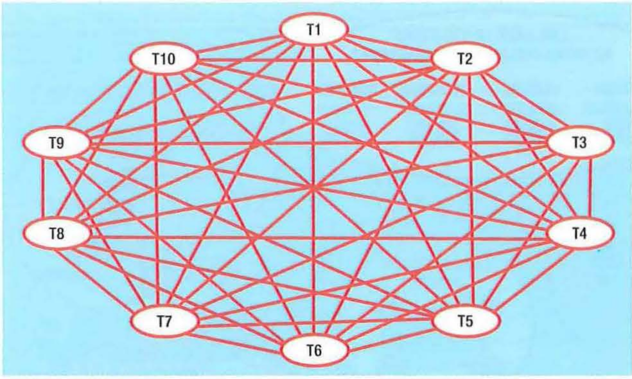


Figure 5—Fully meshed circuit switched network

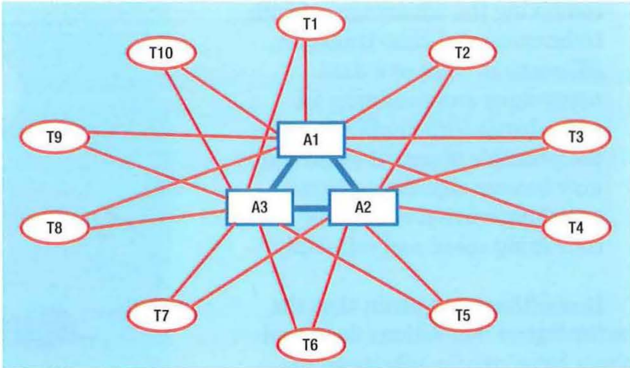


Figure 6—Voice-over-packet architecture

- *feature server*, for call processing in cooperation with network databases.

The switching function between the voice gateways is done by a limited number of ATM backbone switches (see Figure 6). The number of ATM backbone switches is considerably less than the number of gateways. The reason is that the core switches are designed for large traffic streams, optimised for STM-4 interfaces. Having one ATM switch per voice gateway location, connected in a full mesh, would result in a poor link efficiency, and thus in high cost, both of the ATM switches and of the transport network.

Therefore, the number of ATM switches should be minimal, taking into account their maximal capacity, and sufficient network and routing redundancy in case of a failure of one of the ATM switches. Typically, there will be 5–10 ATM switches.

The architectural differences lead to a difference in cost of the time-division multiplex (TDM) and voice-over-packet network, both from a one-time investment and from a yearly operational point of view. The main characteristics are as follows:

- *Cost per port of the equipment* In the TDM network, a call going through a switch requires two trunk terminations. In the voice-over-packet case, the voice gateway interfaces to the circuit-based side of the network via a traditional trunk termination. At the other side, the voice gateway has an ATM or IP interface. This interface has lower cost per circuit-equivalent than a TDM interface. This even compensates for the extra, but relatively very small, costs of the core ATM switches.
- *Trunk efficiency* The TDM network consists of a full mesh between the circuit switches. Therefore, there are many, relatively small entries in the traffic matrix, leading to a low trunk efficiency on these direct logical trunk groups. In the ATM case, there are a limited number of trunk groups between the voice gateways and the ATM backbone switches. Each has a quite high traffic load, thus leading to a high efficiency. The trunks mentioned so far, are E0s. Rounding these up to E1s, again gives difference in efficiency between the two architectures, for similar reasons. The difference may even grow further if higher granularity (E3 or STM-1) exists in the transmission network. The difference in capacity is reflected in the cost of transmission. Its relative impact on the total price depends on the cost structure, in particular whether or not leased lines are used.
- *Silence suppression and voice compression* ATM and IP allow application of silence suppression and voice compression techniques. These will result in further reduction of the required bandwidth, and thus in savings in the transmission network. However, this has to be applied with caution. In particular, these techniques are not suited to fax and modem traffic, and may even lead to extra capacity requirements because of the additional overhead.
- *Operational savings* As mentioned before, the TDM architecture has a much larger number of trunk groups than the voice-over-packet architecture. From an operational perspective, there is a close relation between the provisioning costs and the number of trunk groups. First of all, the size of the trunk groups has to be determined, based on traffic figures obtained from measurements and forecasts. Secondly, the actual provisioning requires operations on each trunk group. This means that the costs of the provisioning will be substantially lower in a voice-over-packet network than in a TDM network.

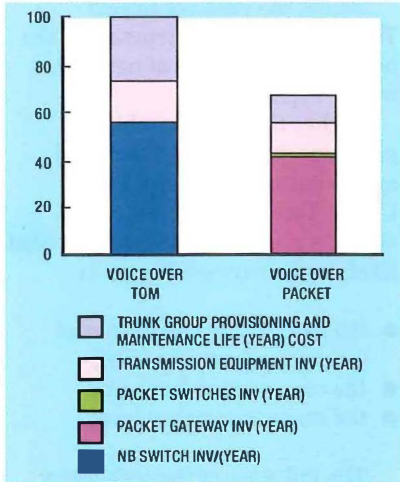


Figure 7—Life cycle cost comparison of voice over TDM and voice over packet

Cost results

As an illustration of the effects mentioned above, an example life cycle cost (10 year) comparison is shown in Figure 7. It is based on a network with 50 nodes. The traffic offered to this network is 375 000 Erlang. No silence suppression or voice compression is applied.

Acknowledgements

E.Drakopoulos, Lucent Technologies; Regional Director Network Planning EMEA; R.Cornejo, Lucent Technologies; Manager Network Planning; and H.Fischer; Lucent Technologies; Manager Network Planning EMEA.

Biographies

Johan Kardol
Lucent Technologies

Johan Kardol is regional Director Network Planning and Design with Lucent Technologies and is responsible for the design of public voice and data network infrastructures for customers of Lucent Technologies in the European region. Customers are incumbents and new entrants. Since 1970 he has worked at Philips Telecommunication Industry, AT&T and Lucent Technologies. He has experience in telecommunications in a number of assignments: responsibility for development of telecommunications equipment, project management of telecommunication projects, operational responsibility for business networks and consultant. Johan received a master's degree in electrical engineering at the TU Delft.

Paul Kallenberg
Lucent Technologies

Paul Kallenberg received his master's degree in mathematics in 1975 at the University of Leiden, the Netherlands. In 1979 he received his Ph.D. in mathematics, also at the University of Leaden. From 1979 to 1980, he was member of the department of Statistics of the Centre of Mathematics and Computer Science in Amsterdam. Since 1980, he has worked at Philips Telecommunication Industry, AT&T and Lucent Technologies. He was involved there in teletraffic engineering, in particular on overload control, network dimensioning, congestion control and queuing problems. Since 1990, he has worked on network planning. He has experience in modernisation plans for switching and transmission networks in many countries in Europe and the Middle East, cost comparison studies for TDM and voice-over-packet (ATM/IP) networks, in access network planning projects and planning of SDH and ATM networks.

Andy Pelling, Steve McNickle and John Hooper

Network Management in the 21st Century

The story so far....

Chapter 1—The Quest.
On a small blue planet in the outer reaches of the Milky Way a rapidly developing civilisation is becoming increasingly hungry for information.

The dawn of the age has arrived where it's people have recognised that 'Knowledge is Power' and global telecommunications is the key to unlocking that power.

Back on Earth, most of the leading telcos have now embraced the concept that superior operational support systems (OSS) help to keep them ahead of the pack by permitting optimal network utilisation, reduced cost, improved lead times, greater efficiency and flexibility. Many operators now have state of the art 20th century systems in place such as network infrastructure management and process automation tools, others are in the process of developing or deploying them now.

Challenges addressed to date have included:

- reducing the need for detailed equipment configuration knowledge in the planning community, and
- automating the planning processes.

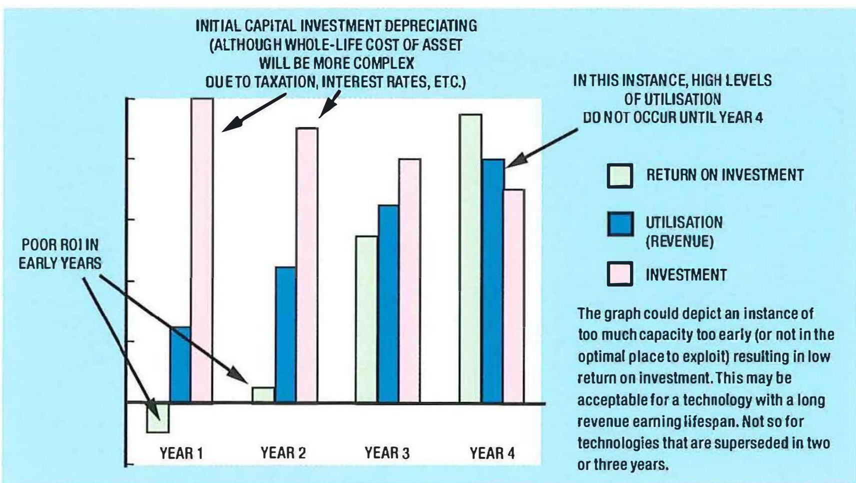
The combination of these can dramatically increase productivity and asset utilisation, facilitate multi-skilling and planning transportability as well as enforcing best practice and ensuring the adherence to minimum standards.

This paper now poses the intriguing questions of what are the next giant steps for mankind in the telecommunications industry; where do we boldly go beyond the OSS of this century? As we all know, standing still means the competition will draw ahead and the race will quickly be lost. It is therefore not an option. Axarté is working with customers to try and anticipate the operational support systems requirements for the next millennium. In such a rapidly developing industry it is not practical to plan too far ahead as technology advances and changing customer expectations are likely to revolutionise many sectors of the market beyond recognition in just a few years. For example it was not so long ago that many access providers believed that they would have to bear the incredible cost of replacing their copper networks with fibre. This strategy has now been radically changed by the development of technologies such as asynchronous digital subscriber line (ADSL), which will give many access networks a new lease of life, enabling them to support the products of the next few years.

One area worthy of significant development is capacity management. This in essence means a move from reactive to **proactive** planning. Some operators have networks with inherent values in billions of pounds and annual investments of hundreds of millions. If these are not optimally managed then the increased cost can be huge. A saving of 1% may seem small but in this context can be worth several millions and is likely to pay for the OSS many times over. Good capacity management, in simple terms, means having the right amount of capacity in the right place and at the right time. Put too much in too early and return on investment (ROI) is low. Provide too little, too late and potentially huge costs are incurred in reactive provision (fire fighting) of short-term, poorly planned network growth. In the worst cases of course, customer orders are lost and thus revenue reduces.

Figure 1 shows the first of these shortcomings (that is, over capacity, too early).

Figure 1—Over capacity, too early



Andy Pelling, Steve McNickle and John Hooper:
Axarté (formerly Accunet),
Votec House,
Hambridge Lane,
Newbury, Berkshire,
RG14 5TN
United Kingdom
Tel: +44 (0)1635 528001
Fax: +44 (0)1635 528085
E-mail: smcnickle@axarte.com

How Do We Get There?

To be able to manage a network efficiently, the starting point is accurate records. Many have trodden the expensive path of data cleanse and the list is long and distinguished of those that have spent vast sums cleansing data but have failed to put the systems and processes in place to maintain that accuracy.

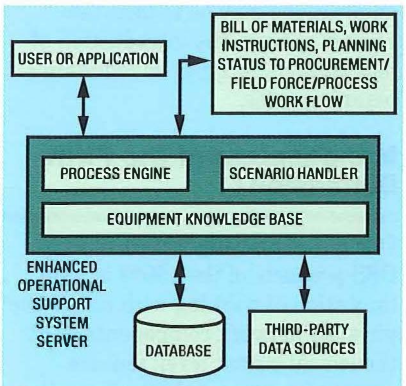
Emerging operators are faced with the pressures of quickly establishing new networks and being able to meet large volumes of customer orders. Starting with a 'clean sheet' of records, they often believe that data accuracy is only a problem for the incumbents, it will however, soon become their problem too and of course the sooner it is tackled the less the cost of data cleanse. Their options are:

1 Incorporate systems/processes to maintain data accuracy from day 1.
2 Wait until it becomes a significant problem then undertake data cleanse and option 1.
3 Undertake data cleanse every couple of years or so.
4 Ignore the problem and reap the consequences later.

To maintain accurate records, network operators and equipment suppliers require a process automation tool which can add to and make changes upon the record data automatically, removing the possibility of human error and not allowing the process to be completed without full databuild to predetermined standards; and which updates, and checks the integrity between, dependent systems via synchronisation. Figure 2 shows the key elements and operations.

Assuming that an operator has accurate records of all network assets and the varying status of each

Figure 2—Process automation tool



component, there are three main considerations for using that data.

- Firstly, the investment needs to be exploited to the maximum; as much available capacity as possible must be utilised for revenue earning products, or leased to other operators or moved to a location where it can be better utilised.
- Secondly, the operator must be able to maintain the network at minimum cost, by being able to anticipate potential areas of deterioration in terms of loss of service or lower operational parameters. When components do fail they need to be located rapidly and alternative network solutions need presenting on-line.
- Thirdly, consideration needs to be given as to how this network will evolve. In many cases this will be growth although it must be recognised that some platforms are supporting dying or migrating product streams and planned downsizing can give gains in reduced operating cost. Alternatively networks may need to be enhanced to support future services. Individual platforms can no longer be looked at in isolation as most are linked by the service streams they support. A new generation switch network for instance will have significant impact upon the transmission network.

Within the third point above the current focus has tended to be on monitoring for exhaustion. Simply identifying redundant equipment and turning off the power (which in turn is loading vent plant) can reduce operating and maintenance costs.

If we look at network growth, which will account for a significant proportion of a telco's capital investment, the following factors need to be considered.

Initially there are some key business decisions to make, such as:

- What will be the ROI for this expenditure over its useful lifetime?
- Should I buy or should I lease?
- What are the trade-offs between incremental build (just-in-time) and the economies of scale from bulk build?

A good OSS will give vital information, on demand, to support these types of decisions. Various scenarios should be easily entered into the system, which will be able to return costs and resource requirements for each scenario along with the capacity profile that such a solution will deliver.

If the decision is made to add to a network for increased capacity, trending and projections based upon forecasts and weighting factors should be available to the OSS user. Armed with this information an optimum build can be planned in terms of topology and quantity. Where a scenario for incremental build or relatively small-scale bulk build is identified as a suitable option that will support projected demand, this allows deferment of the capital spend or redirection of it to other requirements. Alternatively, for certain types of networks where intrusion can be costly† and components are not; bulk capacity build may be desirable.

The enhanced operational support system (EOSS) developed by Axarté allows multiple scenarios to be built, tested against and costed. One scenario can have associated sub-scenarios. Once a scenario is promoted to the chosen option, a full history trail is maintained of what went before, and this history provides a very important source of actual time-based network changes.

Axarté is developing ways to manipulate the network data to give projections of capacity exhaustion, over-provision, under-utilisation, non-optimal topologies and opportunities for directing sales to areas of potential capacity exploitation. A user will be able to display the results of such network queries in a multiplicity of formats. Automated warnings can be applied or reports can be initiated and then represented as anything from a model of the network at some pre-determined future time to simple graphical displays, such as pie charts and line graphs or tables.

Thresholding, But Not As We Know It!

The ultimate step would be to trigger a build at defined thresholds using an automated plan-and-build process

† Intrusion is manual work within the network infrastructure with costs possibly due to faults caused by these interventions, or out-of-hours working to minimise disruptions, etc.

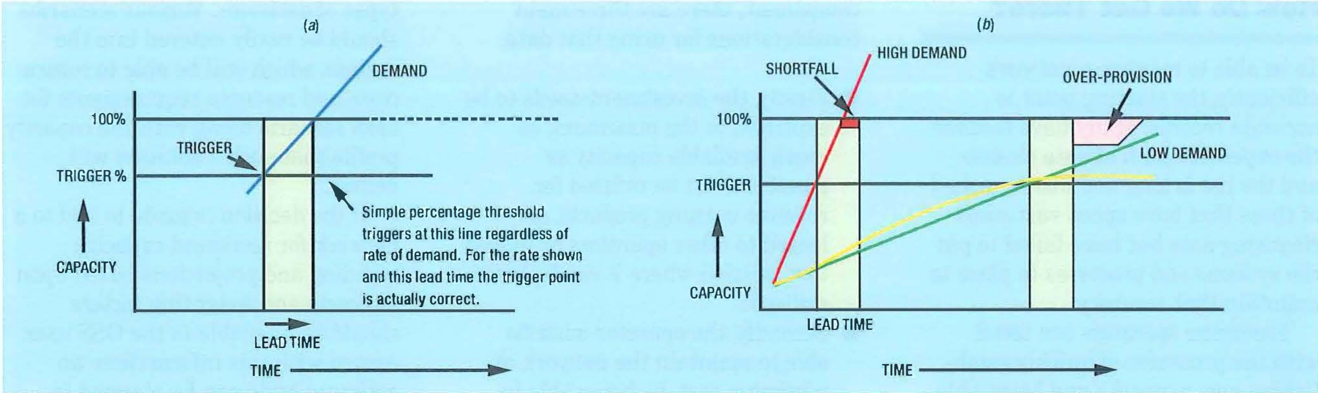


Figure 3—Thresholding the old way (before time travel)

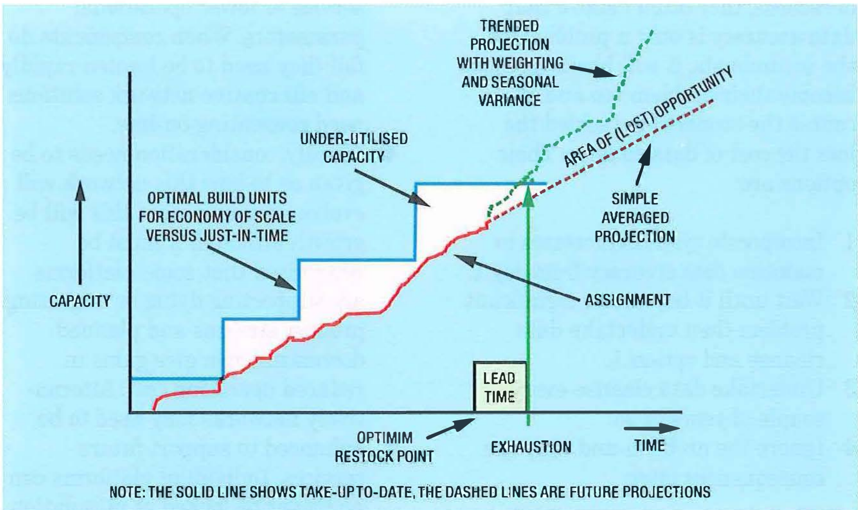


Figure 4—21st century thresholding

manager with no manual intervention, and for some straightforward types of network build this may be appropriate. A threshold would not be a simple percentage remaining as this is incredibly inefficient (although widely used at present). Experience shows that this type of triggering will almost always get it wrong and as such lead planners to build in excess safety margins which results in under utilised networks with poor ROI. The graphs in Figure 3 show that only in the case of the medium demand (blue line) in graph (*a*), which also happens to be linear, does the trigger point lead to the optimum network build time. Figure 3(*b*) shows that a higher demand will lead to exhaustion in advance of the delivered capacity; a lower or tailing off demand means that capital has been invested far too early and in some cases there may never be a return on part of that investment.

The two graphs show four possible rates of growing assignments to a network segment, which in simple thresholding would not be trended. New build would simply be triggered when the set percentage utilisation was reached. Taking the lead time between planning/ordering and installed capacity being available for use as the green rectangles it can be seen that the optimum point for re-ordering will vary depending upon the demand profile.

Capacity planning tools need to be much smarter than just employing simple thresholding. An application must consider historic trending at an appropriate level of granularity, and take account of non-linear demand (that is, seasonal variances, market driven peaks, etc.). They must also allow the input of forecast data and weighting factors. Lead times and safety margins can be built in to remove the over-ordering of cautious planners. Where rules are breached or overridden, flags or exception reports can be raised. Another requirement will be the ability to consider both the physical and logical/virtual components with regard to capacity management.

Such a tool is effectively giving the user a glimpse of the predicted future based on the best information to hand and calculations using pre-determined logic rather than 'intuition'.

Figure 4 depicts both a simple averaged projection (which is better than percentage thresholding), based on actual time-based information, and a more sophisticated version where other factors are taken into account.

We Have the Power

Another key requirement for the OSS of the future must be to empower the customers to perform an appropriate amount of system configuration. It is not reasonable to expect that the users, or even in some cases the support/administrators, will have software programming skills for the application. They could however, be guided through reconfigurations by intuitive 'wizards'.

This is true today for many off-the-shelf home and office type products. Not many users would attempt to re-write their Internet software but to reconfigure a new modem for instance is a relatively simple task using a step-by-step wizard. Future systems will have such tools available to the user although full support will still be available for those that prefer to devolve such work to the experts.

Typical user or client configurable applications may be to facilitate process change, the introduction of new equipment types or redefining planning rules.

Management of the Fifth Dimension?

One of the limiting factors of some OSS products of the 1990s is their limitation of working with either the physical network components or at the logical circuit level; few are designed to manage both. Capacity management tools and configuration systems usually operate at the logical level. Most planning systems can only consider equipment components and the connection between them, in other words, three-dimensional entities. We have already seen previously that EOSS will allow scenario planning and capacity projections which are effectively the fourth dimension (time). Beyond this there is the concept of virtual paths and virtual circuits which are common with today's technologies. These *logical entities* (the fifth dimension?) need to be managed in an entirely new way from their physically constrained bearers. If all of these 'dimensions' can be handled within one tool..........

Open Interfaces and Intelligent Message Handling

It is already the case that no one system can provide a complete solution for the many and complex requirements of the telecommunications industry. This requires a range of interface technologies being supported by the OSS applications and intelligence within the messaging sub-system.

In today's global market place these applications could be geographically remote and in countries of differing native languages. As telcos expand they may take on new partners around the globe, who must also be incorporated into the overall OSS environment.

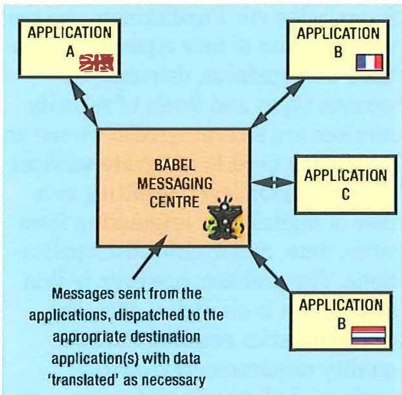
Ideally a platform independent messaging centre will handle this—the Babel messaging centre (Figure 5).

This message centre will automatically interrogate and direct messages between applications and guarantee delivery. But, what if this system actually translated in real time the appropriate data within a message to the native language of the receiving system (or indeed native languages of multiple receiving systems). How much more useful any instructions and information would become to the receiving party.

As an interim, applications comprising an OSS need to be capable of handling language issues, such as menus, tool tips, icons, but also including value translation.

Axarté is addressing these international issues today.

Figure 5—Babel messaging centre



Conclusion

Limited automatic planning is here today. The next millenium will bring automated proactive planning to achieve maximum ROI and efficiency, and integrated globalisation of telco OSSs.

For the 21st century, it looks as though there will be life out there, beyond the millennium, at least for those telcos that have been planning ahead and boldly going where no system has gone before.

Now as for the 22nd century, we can percieve networks on chips that intuitively plan and automatically reconfigure themselves in anticipation of the next second's capacity requirements in direct communication with satellites. These will provide access to a virtual datastore the size of a peanut containing all known data in the universe. The location of the end user chip—a cranial implant! Management of that we will discuss next millennium!

Biographies

Andy Pelling
Axarté

Andy Pelling is a telecommunications consultant for Axarté. He has 24 years experience working with one of the world's leading telcos, specialising in networks, systems and process improvements.

Steve McNickle
Axarté

Steve McNickle is the Technical Director for Axarté. He has some 20 years experience in the IT industry and is focusing now on flexible solutions to telco network, systems and process management challenges.

John Hooper
Axarté

John Hooper is the BT account manager for Axarté. He worked for 28 years within BT in various operational roles. Latterly he was seconded to the BT 'Breakout' initiative investigating process re-engineering within key operational areas.

Jos Gerrese

Managing Quality of Service in an IP Infrastructure Environment

This article explains that quality of service (QoS) depends on the type of services, the construction of networks and the requirements of different customer groups. A brief historical overview is given on the developments and transitions happening in transporting electronic information over cable systems, with major changes in the areas of transmission protocols and analogue/digital conversion. An explanation is given on the new converged networks concept, highlighting some consequences, an important one being the fact that converging means compromising and possible sub-optimisation.

The actors, the network, the services and the customers are each further analysed. The consequences of increased QoS requirements are investigated. The conclusions indicate that costs will increase compared to today's IP pricing level and that a lot more attention needs to be paid to QoS supporting technical and administrative facilities.

Jos Gerrese:
AUCS Communications Services
E-mail: jos.gerrese@aucs-europe.com

Introduction

Moving from the well-organised world of the public switched telephone network (PSTN), integrated services digital network (ISDN), X.25 and private leased lines into the wilderness of Internet protocol (IP) networking creates a number of issues. This together with the fast-increasing demand for applications based upon Internet/browser technology makes quality management in IP networks a real challenge.

This article describes some of the background, some of the major issues and ways in which they can be brought to a solution.

Brief Historic Overview

Looking from the early days of telecommunications to current practice we see a development starting a long time ago with analogue voice networks; telex and telegraph entered the world as the first digital forms of data, running over analogue circuit-switched voice networks; and about 30 years ago we saw the first data modem traffic, running on the analogue voice network. Later on we saw data using its own infrastructure facilities like X.25 packet-switched networks, evolving into frame relay, some asynchronous transfer mode (ATM) and now transmission control protocol/Interet protocol (TCP/IP).

In the meantime, voice has transformed largely into a digital format. The growth in volume of data networks and the totally different cost structure of those networks make it feasible and attractive to run digital voice on the back of digital data networks. This is called *convergence* in modern jargon.

Over time we have seen a complete turn-around from all traffic over a circuit-switched network to all traffic over packet-switched networks. Will this be the end?

Present Situation

The developments as we see them over the last few years show an incredible growth of data traffic with recently a strong focus on IP, mainly supporting public Internet usage. The main suppliers state that network facilities are doubling every 3–6 months. This has resulted in a frenzy of contruction of new transatlantic cable facilities.

Another important effect of the Internet explosion is that huge numbers of people are familiar with browser-based applications, fast access to all sorts of information and global networking. It is no wonder that the business community wants to benefit from those developments and opportunities, to carry out their existing business better and create new ways to perform new business.

However, rightfully they require the same sorts of service and quality assurances as they are used to with the traditional communication networks, like voice, telex, X.25, frame relay etc. Furthermore, we see that, because of new types of applications and business, demands for various types and levels of security services are emerging. Also we see an increasing need to integrate services at the desktop level, resulting in a flow of digital data originating from voice, data, and multimedia applications. The problem, however, is that although it is all digital data, the characteristics and therefore the quality requirements from the various applications and sources are very different. Real-time and on-line services like voice and video have different critical quality parameters than X.25 or TCP/IP data.

The Challenge

The challenge is how to bring law and order into a network so that individuals and corporations can perform their businesses in a way that is reliable and sustainable, supported by service level agreements, while keeping the flexibility which makes the IP so attractive to an increasingly large part of the world population.

Directions of Solutions

A number of various service, network and customer characteristics need to be addressed to reach a situation as described above.

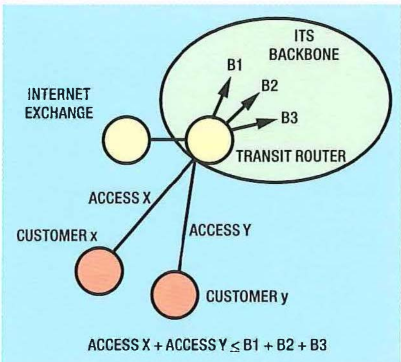
The Network

IP networks and the TCP/IP protocol are of the 'send and pray' type. Packets will find their way through the network up to their address destination. This may go fast but it could also take a very long time before a packet arrives at the destination and the response back is received.

Basic QoS has three key parameters:

- network availability—up-time of the facilities;
- round-trip delay—how long does it take for a package to arrive and to receive a response; and
- packet loss—if no transmission path is available the service will drop packets, which is registered as such, but they do not reach their destination.

Figure 1—Basic quality through non-overbooking



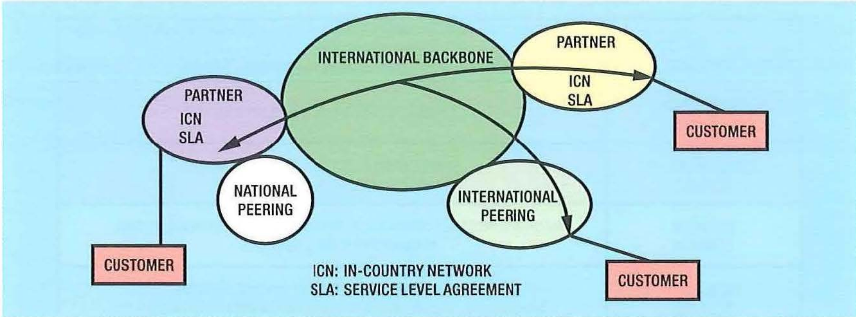


Figure 2—Quality reach extension through peering and interconnects

As long as traffic flows in a network which, up to its boundaries, is controlled by one network provider or IP service provider it is possible to have strong controls in place through network design, network management and traffic observations. As soon as packets start hitting other network domains which are beyond controlled reach, QoS can no longer be managed. One fully depends on management levels applied by the other network supplier(s).

Even when the other networks are managed with the same QoS across all connected networks, then still the final quality is determined at the end address by the design and performance of information servers, home pages etc. There is very little opportunity to enforce quality at the very end of the service.

The optimum quality arrangements can be obtained by building QoS agreements across networks. This can be achieved:

- first and foremost by expanding the reach of the direct controlled network as far as possible through IP backbone structures; and
- secondly, through peering and interconnect negotiations it is possible to expand the QoS reach in an indirect manner by agreeing QoS parameters and values to be respected by the interconnect partners (Figure 2).

Interconnection of IP networks takes place on three layers:

1. *Peering† in public Internet exchanges* Peering in public Internet exchanges like in Amsterdam, Stockholm, Paris, Geneva and London is the most simple way to connect IP domains. In generally there are no QoS objectives or guarantees.
2. *Peering in commercial Internet exchanges* Commercial interconnect exchanges like Telehouse in London can apply at least a certain level of quality requirements for those participating.
3. *Bilateral interconnect arrangements* These are by far the best instrument to get QoS across IP networks.

† Peering is a form of interconnect based on the peer principle. There is no accounting and no exchange of money ('peer' = equal).

The Services

A most common split in IP services is based on the level of accessibility:

- *Internet World Wide Web (WWW) traffic*, which is fully open and public. No special quality measures are in place and quality is largely determined by the various Internet service providers (ISPs).
- *Intranet traffic*, which is fully IP/browser based but access is limited to a defined closed user group through firewalls, filters and passwords. This is mostly used by corporations to run their internal workflow and reporting processes.
- *Extranets*, which are intranets with wider access facilities, creating a virtual business community to exchange information, to run critical processes and especially to support logistic processes: the order to delivery chain.

From a different angle we see for all those types the rise of electronic commerce in various stages of sophistication. Security is the key issue for this type of applications.

Services can further be divided into: network services, communication services, information services, transaction services (Figure 3). They

TRANSACTION SERVICES	ELECTRONIC DATA INTERCHANGE (EDI)/WEB EDI, ELECTRONIC PAYMENT SERVICES, TRUSTED THIRD PARTY/CERTIFIED AUTHORITY, CALL CENTRE PLATFORM
INFORMATION SERVICES	WEB HOSTING, DIRECTORY SERVICES
COMMUNICATION SERVICES	400NET, IP MAIL, IP FAX, MANAGED VOICE, UNIFIED MESSAGING, GROUPWARE
NETWORK SERVICES	MANAGED INTERNET SERVICES, IP-VIRTUAL PRIVATE NETWORK, IP-DIAL, MANAGED FIREWALL SERVICES
TRANSPORT SERVICES	WHOLESALE ACCESS/INTERNET TRANSIT SERVICES TRADITIONAL: FRAME RELAY, MANAGED BANDWIDTH SERVICES, X.25

Figure 3 – The IP services value stack

each have their different quality requirements.

The Customers

The users of IP facilities can be split roughly into three categories:

1 private and residential users;
2 business-oriented users, small businesses and corporations; and
3 professional Internet companies like ISPs, application builders, and information providers.

The main difference between categories 1–2 and 3 is that 1 and 2 are users of the IP facilities but it is not their core business. They use it as a support facility to run their business. Businesses in category 3 depend on IP for their core business.

It is clear that these categories have different requirements with respect to IP quality. Cost is the most important issue for some customers, with not too much emphasis on quality, while for corporations that run mission-critical applications reliability and security are the most vital points.

Consequences of Increased and Differentiated Quality

It is not possible to discuss QoS and SLAs without the proper measuring, reporting and management facilities in place. These are in fact new to the IP environment and will therefore have related effects on costs, operational staff, customer care, etc.

It can safely be stated that the cheap IP Internet has to become more expensive when all sorts of quality measures and management are added. At this point in time it is still hard to quantify those increased cost effects but they will be there for sure.

As an example we can look at the requirements from a voice-over-IP perspective, if we require a quality level that matches with today's voice virtual private network (VPN) facilities:

- We need to implement guaranteed and delay-less throughput. In IP this can be done fundamentally in two ways:

  – make sure there is always and everywhere an abundance of bandwidth capacity so that congestion never occurs; and
  – create special pre-reserved (like circuit switching) facilities in the network like RSVP, MPLS etc.

- We need to implement the necessary operations and management tools.
- We need to implement the required administrative facilities that can measure and invoice the usage of special quality services within the network and across its boundaries.
- It may be necessary to build more resilience in the network itself

It may be clear that each of those four mentioned points will increase costs.

Conclusions

Because of the increasing need for functionality and QoS it is most likely that IP will not be the final network stage and some new converging structure with a mix of packet- and circuit-switched elements will be developed.

As soon as QoS needs to be delivered the costs of service provision will rise sharply above the current prices as now applied for the public IP Internet environment.

Through the fundamentally different cost structure and more efficient usage of the IP type networks the total costs of converged services networks may still be lower than in the traditional network environment.

QoS depends on the requirements of the various user groups and at the same time on the functional requirements of the various service types that run over the network. It is therefore a complex mix of influences which all need to be valued and priced to reach and maintain a profitable business.

Biography

Jos Gerrese
AUCS Communications Services

Jos Gerrese graduated with a Masters Degree in Telecommunications from Technical University of Delft, Telecommunications in 1973. From 1973–1984 he undertook various technical, commercial and management functions in PTT Telecom, Netherlands. From 1984–1989 he was Director of the consultancy organisation NEPOSTEL for South East Asia, stationed in Jakarta. From 1989–1993 he was General Manager of International Services in PTT Telecom, Netherlands. From 1993–1996: he was Director of International Services Portfolio for AUCS and in 1996 became Director of AUCS Internet Business Services, responsible for wholesale access and transit IP services and IP offers. He has been president and chairmain of various consortia, responsible for building multinational broadband networks in the European territory (EBIT, GLOBAND). He is a frequent speaker at conferences and seminars around the world.

Alcibiade Zaganiaris and Achilleas Kemos

Economic Viability of Advanced Telecommunications Services

This paper presents a methodology for the assessment of the access network for the residential and small business user. This enables a global system assessment and the identification of minimum-risk introduction strategies. The various network alternatives are further assessed using risk methodology. The impact of critical variables such as key component costs, civil works, operation, administration and maintenance (OAM) costs, competition and demand forecasts is modelled through probability distributions.

Introduction

The world economy has experienced a significant proportional shift of the major sectors over the past 30 years—agriculture, industry, services—in the gross domestic product. The main thrust of this movement has been a steady decline in the shares of manufactured and agricultural products on the one hand and an expansion of both market and non-market services on the other: finance, transport, tourism, distribution consultancy, education, etc.

Alcibiade Zaganiaris:
France Telecom-CNET, 22307 Lannion, France.
tel: +33 2 96 05 26 09
fax: +33 2 96 05 33 91
E-mail:
alcibiade.zaganiaris@cnet.francetelecom.fr
Achilleas Kemos:
ENST Bretagne, BP 78, 35512 Cesson Sévigné-France.
tel: +33 2 99 12 70 39
fax: +33 2 99 12 70 13
E-mail:
Achilleas.Kemos@enst-bretagne.fr

While employment in industry has declined over the past decade, services have been the major job creator, compensating for losses from industry, though unable in Europe to create enough jobs to prevent an overall decrease in employment. Advanced communications is a key factor in the service sector's role as a job creator. The share of information-based jobs is rising fast. Two basic elements are needed for such services: unambiguous standards and critical mass. The attraction of a telecommunications service depends directly on the number of compatible users. Thus a new service cannot really take off until a certain number of users is reached. Once this critical mass has been achieved, growth rates can increase dramatically, as in the case of Internet.

A virtuous circle of supply and demand can be created only if a significant number of market testing applications, based on information networks and services, is launched Europe-wide, to create a critical mass. The issue of critical mass is even more crucial for peripheral regions, where economic, social, cultural and technical barriers are a hindrance to the diffusion of such services. Also, the business opportunities for small and medium enterprises (SMEs) are limited by a variety of factors such as human resources and finance and the difficulties they encounter accessing appropriate information and being integrated in the global industrial fabric. The rapid growth of the Internet illustrates the enormous potential of new global networking possibilities. However, specific services adapted to SMEs' needs are poorly defined or missing.

Innovative use develops best when driven by user communities together with manufacturers and network providers. Such uses assume the availability of service-oriented networks and open access for service providers. Moreover, in a competitive environment, new network infrastructures have to be viable in the usual economic sense; that is, by minimising the present worth of the annual charges. Furthermore, the investment must generate positive returns, consistent with investors' typically short-term expectations. The decision to invest in new technology is subject to the risks and uncertainties inherent in the competitive marketplace. Technology-driven innovation is the most costly route to a new product. Diminishing the profitability of a product limits a key source to new investment. In the former monopoly environment, network deployment was implemented by a single regulated firm. Product life times as well as user numbers over time were then predictable and the risks associated with the investment were readily known and manageable.

A number of studies have been carried out in order to provide a sound support to the strategic decisions of the main actors in the field of access network techno-economics. As a main result, guidelines are provided to the network operators for both the selection of the candidate set of services and the deployment of the appropriate network architecture in different operating conditions. Additionally, considerable information has been made available to the manufacturers to make them aware of the risks and opportunities of developing broadband equipment. Finally, cost/benefit analyses highlight the user's point of view regarding the opportunities of the various telecommunications services.

Overview of the Techno-Economic Methodology

The methodology is being used for the evaluation of advanced multimedia services and networks in a competitive, deregulated environment. Fast improving technology gives birth to new alternatives, yet society's ability to fully exploit them still lags by years. The scenarios studied include gradual provisioning of services like plain old telephony service (POTS), narrowband integrated services digital network (N-ISDN), asymmetric switched broadband, symmetric switched broadband and broadcast broadband. The quality of the existing infrastructure, the average subscriber capacity demand and demography are among the key elements likely to affect the economic viability of access network upgrade strategies.

The importance of key engineering parameters such as subscriber density and civil works are highlighted through various sensitivity analyses. The various network alternatives are further assessed using risk methodology. The impact of critical variables like key component costs, civil works, operations, administration and maintenance (OAM) costs, competition and demand forecasts are modelled through probability distributions. The effect of a set of values of the critical variables is calculated by one program run. The risk is then assessed by generating many values from the probability distributions and subsequently performing a large number of program runs. The parameters relevant to the engineering of the network, such as subscriber density and type of civil works are then defined. These are the inputs to a geometric model which calculates the overall cable length of the various sections of the network. The costs of the network components are calculated using an integrated cost database, containing data gathered from many European sources. The resulting discounted system cost can then be derived. Finally the overall financial budget is calculated for the various architectures by integrating demand and tariff data with the life-cycle cost. The database includes costs at a given reference year for components, installation, civil works, operations and maintenance. Learning curve data together with demand estimates classified in relevant groups, are included in the database. The demand for the services, which can be provided over different network architectures, is estimated through a Delphi survey carried out in the countries participating in the project. The potential market and demands are derived for each service. The revenues, determined from tariff and demand estimates, together with the life-cycle cost, lead to the overall financial budget of the system considered. The impact of key engineering parameters is highlighted through various sensitivity analyses.

Different architectures which can provide a selected group of services are considered. Some key model outputs are time series over the study period, others are constant values:

Investments are usually spread over the study period. To get a single figure of merit for the total investment, the future investments are discounted to the start of the study period using the conventional discounting formula. The total discounted investment cost is usually called *First Installed Cost.*

Most often, the calculation includes the *revenue* from the services. This is derived from a service connection tariff and by estimating an annual tariff for each service per connected customer.

Profits are calculated from the revenues, investments, depreciations and taxes.

Cash Flows are derived from the other relevant economic parameters in a standard manner.

Life Cycle Cost is defined as the sum of global discounted investments and global discounted running (OAM inclusive) costs. This value represents the total cost for constructing and running the network over the study period.

Cash-flow is the difference between the cash to be paid (negative cash-flow) and the one to be received (positive cash-flow) during a certain period.

Component Price Versus Time

The cost trends are described using the learning curve relation according to Wright's law:

$$P_{2n} = K \times P_n$$

where P_{2n} and P_n are the prices for the completion of the $2n^{th}$ and n^{th} units respectively. Network components have been clustered according to the K factor, subsequently used in the tool.

By combining a standard demand logistic curve for the growth over time of the accumulated component volume with a learning curve, every component can be classified by four parameters. These are the component price today $P(0)$, the learning curve coefficient K, the time it takes for the accumulated volume curve to go from 10% to 90% of the saturation value ΔT and the value of the accumulated volume today $n(0)$, normalised to the saturation value. The last parameter is the inverse of the assumed market potential. The ΔT and $n(0)$ are the only two parameters that need to be estimated by forecast methods. The forecast function for the evolution of the relative accumulated volume $n(t)$ is given by

$$n(t) = [1 + \exp\{\ln(n(0)^{-1}-1)-(2\ln 9/\Delta T)\times t\}]^{-1}$$

where the only inputs are $n(0)$ and ΔT. The previous relation is now inserted into a learning curve:

$$P(t) = P(0)\times[\, n(t)/n(0)]^{\log_2 K}$$

This yields the final expression of the price of the component versus time:

$$P(t) = P(0)\times[n(0)^{-1}\times\{1 + \exp[\ln (n(0)^{-1}-1)-(2\ln 9/\Delta T)\times t]\}^{-1}]^{\log_2 K}$$

which contains all the four parameters involved in the component price versus time evolution. The asymptotic price level for long times only depends on $n(0)$ and is given by:

$$P(\infty)= P(0)\times [n(0)^{-1}]^{\log_2 K}$$

and the slope of the price curve for small t is proportional to ΔT^{-1}.

Cash Balance is the cumulative cash flow. A typical cash balance curve for a network scenario goes first deeply down to the negative side because of the high initial investments. If the scenario is profitable, the cash flow turns positive fairly soon and the curve starts to rise. The point in time when cash balance turns positive gives the *Payback Period* for the project.

Net Present Value (NPV) gives a single figure of merit for a project. Its definition is the sum of *Discounted Cash Flows* plus *Discounted Rest Value* of the project. NPV is a good indicator for the profitability of the scenario especially in these cases where the Payback Period cannot be used because major investments are spread out in time. The weakest point in this figure of merit is the definition or calculation of the rest value of the network.

Internal Rate of Return (IRR) is the discount rate at which the NPV is zero. If the IRR is higher than the opportunity cost of money (that is, interest of an average long-term investment), the project is viable.

Cost/Benefit Analysis

Cost/benefit analysis is a prerequisite to the 'before' phase of an investment project. This is a process of making shared assessments of investment projects which identifies, measures and compares actual and potential costs and benefits of an application over the entire application life cycle. It incorporates all costs and benefits incurred during the entire life cycle and includes operating costs as well. Cost/benefit analysis in TERA† focuses on the end-user point-of-view. The goal of cost/benefit analysis is to assess the costs and benefits of new telecommunications services to a specific group of users compared to the costs and benefits of traditional solutions for this user group. This is a critical step for decision makers, since it helps to identify and manage costs, run the project in a more business-like manner, optimise return on investment and compare the financial and non-financial impact of a project.

In TERA a number of appropriate business cases have been chosen in which the cost/benefit analysis is applied.

† Techno-Economic Results for Acts, project No. 364 of ACTS programme of the European Commission.

The framework for the cost benefit analysis consists of six steps:

- identification of key user groups, life span of the application and discount rate;
- identification of benefits (by business owner) and costs (by cost/benefit project manager), both direct and indirect over the life span of the application;
- calculation of net present value;
- definition of key parameters and performance indicators;
- sensitivity and impact analyses; and
- assessment of the worthiness of the project, evaluation of options and go/no go decision.

The above framework has first been applied to two selected cases, for which reliable cost inputs were available: tele-medicine and tele-learning. The first results quantify the relative savings due to new multimedia services.

A cost/benefit analysis must be completed before the final decision. As the project goes along, the business owner and project manager track the costs and benefits to see if the promises of the cost/benefit analysis are fulfilled. If the scope of the project is changed, the cost/benefit analysis needs to be revisited and readjusted to the new scope.

Cost/benefit studies are closely linked to externalities. In models for market demand in the case of network externalities it is important to estimate the willingness of customers and potential users to pay for the services. Cost/benefit analyses can help in quantifying the willingness to pay for different services and also in different markets. Therefore the two analyses should be combined when applying them in business cases. Cost/benefit studies require considerable resources, especially for gathering the necessary data. Therefore, a simplified cost/benefit model will be derived and inputs from other projects will be possibly used to determine the costs and benefits for the user. The objectives of the cost/benefit studies are the:

- identification of the costs and benefits of a telecommunications project;
- use of the discounted cash flow technique;
- calculation of the net present value for after-tax cash flows and comparison with the conventional alternative;
- calculation of the user's willingness to pay; and
- cost/benefit analysis as a comparison of alternatives over a study period.

Cost/benefit studies are likely to estimate the impact of costs and benefits of an advanced multimedia service upon the user's activity as compared to the old conventional alternative. In today's competitive environment, the end user increasingly needs tools for measuring the relative impact of various alternatives in order to determine his/her strategy. The final choice is not always straightforward and a combination of alternatives might be an acceptable approach; for example, a short-term option chosen as a stop-gap measure and a long-term option also chosen to maximise benefits over time.

Commonly, the time period considered for economic evaluation ranges from 5–10 years. For tele-medicine and tele-learning, a time period of five years was considered appropriate, since new emerging technologies may thoroughly alter a scenario over a longer period.

Conclusions

Demand and tariffs have a major impact on the viability of a particular service, whereas the cost of technology turns out to be of secondary importance. The number of technological options is significant smaller, if an adequate infrastructure exists and the main geographic characteristics are known. The uncertainties associated with revenues are much higher than the uncertainty on the cost of technology.

The broadband upgrade of the access network is of paramount importance for the incumbent operator. The costs associated with such an upgrade are equivalent to the overall costs of building a new access network. Hence the existing infrastructure has to be utilised in an effective way.

The results achieved reveal strong indications for high-tech teleservices such as telelearning, for which the most important costs are for labour. Therefore, telelearning allows substantial savings by sparing labour cost during transportation time. However, the labour costs of the remote learning centre (help desk essentially) may be key for the economic success of the operation. Thus, depending on the quality of service required, the degree of interactivity wanted by trainees can strongly increase the cost of telelearning operations, if the time of individual support greatly increases. The transportation and accommodation expenses appear to be the second important item, especially for long distances. In comparison, the telecommunications costs turn out to be fairly low, even in the case of a high-price connection time policy. The same observation can be made for the telelearning software tools.

Teleworking, information services, and teleentertainment stand out as the most promising services for broadband applications in the future. There will be a substantial demand for broadband services in the residential and the small office/home office (SOHO) market during the next 10 years. However, households are not willing to pay too much more for additional broadband applications and additional capacity.

There is a wide range of telemedicine applications already available but before any real market implementation, deontological and regulatory problems must be addressed. There are two main attributes of telemedicine: bringing the patient closer to the health centre and retrieving the appropriate pieces of medical expertise. Their objective is to favour the right diagnosis and therefore to help take the correct decision. Telemedicine implies an improvement of communication within the health world and a wider dissemination of medical expertise. Both telemedicine and telelearning are of great social importance. However, technologically advanced network infrastructure is a prerequisite for an efficient take-up.

In conclusion, from a strict and limited financial point of view, telelearning is always profitable when there are several hours of transportation between the trainee and the learning centre. However, on-line telelearning also brings additional benefits, in particular a better reactivity for high-tech companies needing to be on top of available technology. Combined with information available from the Internet and specialised media, it provides a tool of efficiency enabling companies to be more competitive.

References

1 Zaganiaris, B. T.; Olsen, K.; Stordahl, L.; Ims, M.; Tahkokorpi, I.; Welling, M.; Lähteenoja, P.; Nieminen, J.; Dostal, F.; and Fleuren, P. Multimedia tecno-economics. *ECMAST '98*, Berlin, Germany, 26–28 May, 1998.

2 Kemos, A. and Zaganiaris, A. Teleservice Business Cases, FITCE '98, 24–28 Aug. 1998, London, UK.

3 Ims, L.A. et al. Upgrading the PNO and cable operator access networks to advanced broadband services: technology alternatives and their economic implications in a competitive environment. Telecom'96, 10–15 June 1996, Rio de Janeiro, Brasil.

4 Kemos A.; and Hamdi, M. Cyberspace and the opportunities of regional student co-operation in economic development. Telecommunications Summit: The international Student Seminar on Human Resources Development, Bandung, Indonesia, September 1995.

Acknowledgement

This paper refers to the studies performed within project ACTS364/TERA. The authors wish to express their gratitude to the European Commission for supporting this work.

Biographies

Alcibiade Zaganiaris
France Telecom-CNET

Alcibiade Zaganiaris graduated in mathematical physics from the University of Athens. He received a postgraduate certificate in telecommunications electronics and a doctor's degree from the University of Rennes, France, then a Ph.D. on telecommunications economics from the University of Lyon. He joined the research centre of France Telecom-CNET, where he carried out research on optical materials and fibre properties, local networks and techno-economic studies. He is author of 75 publications, conferences and patents. He is currently in charge of international cooperation projects involving high-tech SMEs.

Achilleas Kemos
École Nationale Supérieure des Télécommunications de Bretagne

Achilleas Kemos graduated from the Electronic and Electrical Engineering Department of University College London. He then moved to France where he obtained two post-graduate degrees: a Mastère on 'European Business Engineering' at the École Nationale Supérieure des Télécommunications de Bretagne and a Diplôme d'Études Supérieures Specialisées on 'Project Analysis for Industry and Agriculture' at the Université de Rennes I. He is a research engineer in the Economics department of the École Nationale Supérieure des Télécommunications de Bretagne and project manager of the Copernicus—INTACCOMP project, evaluating the impact of research and development in eastern European countries. He is also participating in the European Project ACTS 364 TERA as a techno-economics expert. He has also worked as a consultant for the European Union Directorate General XVI, evaluating the impact on cohesion of community policies for telecommunications.

Mark Hoogenboom, Patrick Steemers and Albert Kuiper

Quantifying the Market Demand for In-Vehicle Telematics Internet Services

Three consultants give their views on in-vehicle telematics Internet services. They explain the approach for creating a suitable business case, in particular the way to look at customer needs and the playing field of the service provider. Service pricing and market size are estimated and the business case issue is concluded with a description on how to structure in-vehicle telematics Internet service development projects, using the Web-enabled Enterprise Framework. There follows a description of the technology required and what is available today. Specifically global system for mobile communications (GSM) operators are mentioned as they clearly are moving towards services beyond speech. The article concludes with a business example of a company using in-vehicle telematics.

Business Case

When looking at the business case for in-vehicle telematics Internet services one must focus on the perspective of the occupant. This focus on an actual user of the in-vehicle Internet service is the same as the focus one needs to develop services on the 'fixed' Internet. An individual (rather than a company or government etc.) should be the focal point as an individual will generate the high-volume services demand that can generate revenue.

Albert Kuiper:
e-mail: albert.kuiper@capgemini.nl
Mark Hoogenboom:
e-mail:
mark.hoogenboom@capgemini.nl
Patrick Steemers:
e-mail: patrick.steemers@capgemini.nl
All the authors are from:
Cap Gemini Netherlands, Division Telecoms, Utilities and Services.
Tel: +31 30 252 7685
Fax: +31 30 252 7686

The Driver and The Passenger

For a driver who is a potential user of the in-vehicle Internet services, the journey is fundamental: it starts at some moment and place, it ends at some future moment at the destination. In between the driver and the vehicle are en route for some time.

The geographical elements are: the starting point, route and destination. (Note that in this setting the driver and optional passengers are taken as the point of view. Someone will be the owner of the vehicle, possibly the same person as the driver, but owning a vehicle is a completely different setting, only touched upon here.)

For any person in a vehicle, three aspects of the journey will be of interest:

- the duration of the drive,
- the route, and
- the current location and its vicinity (a sphere around the vehicle).

For each aspect of the drive, the driver or passenger will have a different set of desires for in-vehicle telematics Internet services. For example, if the duration of the drive is important because a meeting is scheduled at the destination, the driver will have a keen interest in route information to ensure the journey is completed within the limited time. Another example is when the driver or passenger(s) develops a desire for a snack. Then it is useful to know which shop will be able to provide the desired snack within a radius of, say, 2 km from the current location.

In short, human behaviour during a drive is important in assessing which in-vehicle telematics Internet service will be viable and profitable.

The Service Provider

From the point of view of a service provider contemplating a business case for any in-vehicle telematics Internet service, providing information in volume is important. Information has a broad definition here which will be addressed extensively in the following sections.

Providing information in volume implies that a service provider will make a profit from selling a small amount of information often. Hence, the service provider should focus on high-volume low-cost in-vehicle Internet services.

To ensure that the vehicle is capable of receiving the in-vehicle Internet services, some basic hardware must be available but manufacturers of in-vehicle telematics hardware must realise that real profit will be made in providing the services through the hardware, not by selling the hardware.

Now, the customer for any of these services is only available while present in the vehicle. This may sound limiting, but remember that while the driver and passengers are in the vehicle, they have virtually no opportunity to spend money on anything. However, they do have the time to select and buy, as they would when cruising around in a huge department store. Moreover, as the trip becomes boring, they will undoubtedly develop the desire to use their time in a more entertaining way. And that is a perfect moment to have them spend some money!

The price of a service—a packet of information—should be low. So low that one readily decides to purchase the service, just like making a local telephone call (who considers the cost of such a call in advance?). Hence the need for low-cost and high-volume in-vehicle Internet services, like the services on the fixed Internet.

There is one evident distinction between a fixed Internet user and a mobile one: being en route at a specific moment in time with specific unique requirements for that specific journey will enable the service provider to price a service based on these unique aspects. Take two drivers, Roger and Gerwin, who both start to drive from A to B—this does not imply they are two identical potential customers. If Roger is on his way to an important meeting, he will readily spend money on route information that will ensure that he arrives in time. If Gerwin is just driving home, like he does every day, route information is not a service he is likely to buy. For Gerwin it would be nice to know whether he will be home in time for dinner, but acquiring this knowledge has little business value. He would settle for a rough estimate like 'between 30 and 45 minutes', taking into account all route-related data that is available from traffic information from the past (historical extrapolation). Roger, on the other hand, will readily spend a Euro or two for exact, on-line, route information that will guide him around any obstacles to get to B in time. Of course, the next day Roger may be the ordinary commuter, driving from A to C, while Gerwin could then be the person pressed for time on the same route.

Thus the profile of the customer is not based solely on who the customer is, but more on the situation the customer is in. The customer's situation will have a huge influence on the pricing of the service, and therefore service diversification on the quality of the information is essential.

The Size of the Market

Simple calculations show the sheer size of the market. In 1995, people in The Netherlands drove 138·1 billion passenger kilometres. Assuming most of these kilometres were on highways, an average of 100 km/h seems a reasonable estimate (it may be less in practice). This implies about 1·381 billion passenger hours were spent driving or waiting as a passenger to get from one place to another, without spending much money. (The only money they could actually spend while driving was by using their mobile telephone.)

If each journey lasts two hours and two people were in each car, then 345 million journeys were made or about one million journeys per day. Every 700 km or so the car needs refuelling, the only contact the driver and passengers have with the sale of goods and services during a drive.

So, in a country with one million journeys every day, about two million potential customers are available for about two hours. Many are on their way to work or to relatives and a few are in a hurry and pressed for time.

Drive-In-The-Life-Of-A-Driver

To understand how someone works or acts, one should observe the person for a day or so. This shop-floor investigation approach, also known as *A-Day-In-The-Life-Of*, is very useful to understand what situations a person gets into, what moods the person is in, and what the person actually does in the time available. Analysing all the events will provide a profile of the person's day.

Let's take three examples of *A-Drive-In-The-Life-Of-A Driver*.

Drive one—short and familiar

The first drive is 10 minutes long, to get from home to the shopping mall. The driver is familiar with the route and there seems very little to consider in order to provide useful in-vehicle telematics Internet services to such a driver. Apart from providing a parking spot and having the essential groceries packed and ready.

Drive two—commuter

The second drive is a commuter trip. The route is not new, but there are a number of different routes if the standard route is really busy. The journey will take between 30–50 minutes. This driver will be interested in receiving:

- timely traffic information about options in the route, but only when traffic is busy; and
- entertainment through news, music or other content.

The traffic information service must be available before the route option has to be chosen. This is obvious, but as the exact moment of departure and the estimates of traffic congestion are dynamic parameters, it will be a challenge to provide the correct traffic information in time. The traffic jams at any given moment are not important, but the estimated traffic congestion when the driver will be at the congestion is. This kind of information requires a huge number of calculations, on top of the availability of historical data, to provide a decision about the optimum route. Moreover, if the internal decision process of a person is added to this, the mood the person is in that particular day will be an influence as well. The 'I wish to keep moving' mood results in a different decision from the 'I don't mind about anything today' mood.

Furthermore, as soon as a trajectory has been chosen from which no further deviance is possible, most of the value of the traffic information has gone. It can only provide improvements to the expected time of arrival.

In short, the traffic information must:

- provide route options decision support;
- help determine the estimated time of arrival; and
- be tuned to a user profile or actual user requirements.

Today, traffic information only tells you where traffic jams are at a specific moment. This is far from the service a driver would like to have.

The entertainment service should not be seen as a broadcast service as provided by radio stations. For instance, you may like to hear the news as you drive off—not exactly on the hour. For journeys less than 60 minutes, you may not be able to catch the news at all if you are dependent on radio broadcasts only. A *News-On-Demand* service within the first five minutes of any journey would be useful. Moreover, the news content can be based on a listener's profile.

For the more recreational entertainment, there are two important areas to consider. Obviously, *Music-On-Demand* would be an interesting feature, as it is practically impossible to take along all the music the driver would possibly like to hear in the vehicle. Music that the driver has a licence for, for instance because it was purchased on a compact disc, should be free of charge when requested, and new tracks could be bought on the spot.

The other important area is something that resembles a talking book. On request a text is read out loud. This may be an e-mail, a book or perhaps a lesson on some subject the driver is studying. The service, which can best be named *Content-On-Demand*, will keep the driver from getting bored.

Drive three—the Nanny-gator

Now the third drive is more complicated. The journey will take about two hours. The first hour and a half is to get from home to the driver's parents where the children are dropped off for the night. Then a 30 minutes drive is required to get to the theatre to see the Cirque du Soleil at 20:30 hours. It is Friday evening, the schedule it tight and traffic jams are likely.

In this case, the driver is not the only potential customer of a service—the spouse and children are potential customers as well. What would you personally pay for a service to entertain your children in the back seat for two hours? A limited survey came up with an estimate of about five Euro per child per drive†. For this sum you could have a Nanny-gator service, which is the electronic nanny to help a child navigate through the content it may select. This children's version of a personal digital assistant (PDA) is the personal digital Nanny-gator (PDN).

As the schedule is tight and rather complex because of stops and the delays these stops bring about, personalised traffic information is useful to guarantee timely arrival. The first leg in the route, from home to the parents' house, is familiar but as time is critical all route options should be considered to get to the destination in time. The second leg in the route is towards a parking spot as near to the theatre as possible. The route and destination are unfamiliar and the driver will require route information as detailed as to which turn to take. Also, as parking is required, a parking spot that is guaranteed to be available would be important. On top of this, the duration of the second leg of the route is certainly restricted, as the show will start on time.

All in all, for a drive like this, a total of about 15 Euro for in-vehicle telematics services can be considered reasonable, as it will ensure the journey and the event is a success.

Billing of In-Vehicle Telematics Internet Services

As most of the in-vehicle Internet services will be low-cost (because their benefit/value for the driver or passenger is relatively small) billing must be based on small amounts per service delivered. It must be possible to make a profit on billing a 10 cents service.

GSM operators, like fixed-line operators, are able to bill these kinds of services, as short telephone conversations will result in small amounts to be billed (such as 1 cent per second). On the Internet a huge effort is being made to be able to bill low-cost transactions and this is expected to be possible in the very near future.

Billing small amounts for services has a very interesting spin-off—the information acquired about a customer (based on the services requested) has premium rate value on which to actively base customer management. This alone may be a good reason to provide services with virtually no margin, as the analysis of the acquired services is a product in itself that can be sold for profit.

Future Trends For In-Vehicle Telematics Internet Services

If any provider can bring the Internet infrastructure into a vehicle, with a bandwidth comparable to ISDN (or better), a major outlet becomes available for a plethora of services.

The Content-On-Demand service will enable you to listen to your personalised *Half Hour News* where the news items are selected to your specifications on that specific day.

If you get personalised routing information, the system will know where you are. This enables those who you gave access to that information to track you and then your secretary will be able to tell anyone calling your latest estimate of your expected time of arrival.

If you provide the system with the route you are going to follow, in order to get route-specific traffic information, the system could easily provide you with car-pooling candidates waiting along the route. This would introduce a just-in-time car-pooling facility, in which the driver gets paid a fair amount for taking a passenger, while the passenger is able to relax or work.

Develop the hand-held Nanny-gator and you have the ultimate learning and relaxation device for any child.

How To Web-Enable A Car

To get a car on the Internet, to Web-enable a car, is much more than installing the hardware in the car. It takes a whole new look at system design and system development to ensure that a feasible system is developed. This look at in-vehicle telematics Internet services will be very close to the way we look today at designing and developing Web-based solutions (Internet solutions).

Some key features of such solutions are:

- *distribution is an issue*—the system is spread over a number of moving vehicles (clients) and a set of distributed servers (Internet).
- *reliability is not 100%*—the chaotic system architecture on which theInternet is based on will function most of the time, but not all the time.
- *fire-and-forget is a characteristic*—a service request result, for instance a music-on-demand selection, is sent to the customer, but only if the network is able to detect difficulty in delivery will the sender be notified.
- *a dynamic business case is required*—due to the unpredictability of services emerging, evolving and maturing, the business case must cater for quick changes: since competitors can easily jump in at any time, a fixed business case is out of the question.
- *long-term initiatives*—short-term benefits (quick wins) are necessary.

In short, it requires a adapted project approach to be successful. Cap Gemini developed the WEAPON framework for Web-enabled applications for just this situation‡.

---

† Based on interviews with: Steven van Battum, Petra van Krugten, Cobi Vroman, Lilian and Dagmar Hoogenboom.

‡ Coen, Bob; and Hoogenboom, Mark. Web-Enabled Applications Programmed On the Net (WEAPON) Framework, How to become a Web-Enabled Enterprise. McGraw-Hill, 1997, ISBN 0-07-011774-8.

Mobile Internet Will Fuel the Demand For In-Vehicle Internet Telematic Services

More and more customers will choose their service providers based on their offering of personalised, tailored value-added services. Mobile communication and technology will increasingly become commodities. In the battle for customers, vendors are wondering what services should be offered. One way leading to the answer is to look at what services are used by customers today in other, non-mobile situations.

The Internet is services galore

Today services offered on the Internet are being used by millions of users every day. If one is looking for services galore, surely the Internet is it.

First of all e-mail services over the Internet have become a ubiquitous, convenient way to communicate, both for business and private goals. Also 1000 new services are launched every day on the World Wide Web. These range from news, sports, financial and commerce to search-engines and 'portal' services, which help users find their way and drastically influence the way the Web is being used.

Internet technology is used for intranets

Internet technology is being deployed by companies to create their intranets. Corporate intranets offer essential applications and services (e-mail, calendar, business process control and information) to employees. Of course these services are used in company offices, but they can be used from a simple home telephone line.

Taking the Internet mobile

Once the Internet can be used fully mobile situations, potentially all Internet applications will stimulate the demand for in-vehicle telematics services. It can be expected that the person already using Internet services will take the applications and services he/she likes or needs into the mobile situation, for reasons of comfort, efficiency and mobility.

One could formulate the following rule: *People use services when mobile only when they already use them in a fixed situation.*

Mobile Internet enables easy service development and distribution

Upcoming technologies for mobile Internet will have a high impact on the way in which new in-vehicle telematics services (location specific, time-critical, sensor data) will be developed and re-developed in the next three years. This is because in-vehicle telematics services in the next three years can be developed, based purely on Internet technology. Internet technology offers an easy world-wide and open platform for offering any kind of application. The Internet is easy to access. Internet standards (for example for e-mail messaging, Web pages, file transfer and software development (JAVA)) enable the easy development and deployment of these applications.

Permanent mobile Internet connectivity

New mobile Internet technology will be based on a form of connectivity that is (virtually) permanent, comparable to the way people are connected to their computer networks (LANs) in the office. Ultimately, mobile voice services will also be offered over the mobile Internet as voice-over-IP.

Based on the rapid technology developments and the drastic uptake of Internet services, it can be expected that in-vehicle telematics services and other mobile services will be offered to individuals who are permanently connected to the public mobile Internet as well as to their office IT facilities (LAN, documents, services)—all in a mobile manner.

Third-Generation GSM Will Bring Mobile Internet into the Vehicle

GSM technology is rapidly evolving towards its third generation. This generation GSM is based on the universal mobile telecommunication system (UMTS).

Increasing GSM data speeds

UMTS will enable data speeds ranging from 384 kbit/s to 2 Mbit/s ideally, over 200 times the 9·6 kbit/s available today over GSM circuit switched data (CSD). The UMTS data speed will be higher than available over integrated service digital network (ISDN) lines.

Data speeds of up to 2 Mbit/s allow high-quality audio and video communications. Data access, file downloads, webbrowsing etc. all work very smoothly and rapidly at these speeds.

As a stepping stone towards UMTS, the general packet radio system (GPRS) has been defined. GPRS is available today to GSM operators who are implementing it in their networks. GPRS allows for data speeds in the range of 115–384 kbit/s, enabling data communication and audio streaming comparable to using ISDN lines at home.

The mobile Internet protocol (IP) is being defined

UMTS technology is expected to enable IP-based communication and applications. IP is the foundation layer of the Internet today and is constantly updated by the Internet Engineering Task Force (IETF). Under the latest Request for Change (RFC2002), IP is currently being optimised for the specific challenges of mobile communications. Here one can think of the challenges that come with hand-over, roaming etc.

When UMTS is massively adopted by the GSM network operators and when mobile IP standards have been defined, true mobile Internet access will be a reality. Looking at how fast the evolutions are taking place, this will be in three to five years time.

Third-generation handsets

A lot of development is taking place in GSM handsets. Leading manufacturers are currently introducing handsets that are supporting the wireless application protocol (WAP). These WAP-telephones feature a 'micro-browser' that enables the user to view wireless markup language (WML) pages. This simple variant of web browsing allows for such services as flight schedule information, stock information, traffic information, all in a much more attractive and intuitive presentation than seen on current GSM handsets.

One advantage of third-generation handsets is that service applications eventually can be downloaded to the handset dynamically. This allows for updating of services, adding new services or even completely renewing the user interface and menu structure of the handset.

Technologies such as WAP open up the way for the GSM operators to take a role as a service broker, selecting and offering third-party services.

Although currently WAP telephones use the circuit switched data and SMS facilities of GSM, later WAP telephones can be fitted with GPRS or even UMTS capabilities. In this way the high-data-speed advantages of third-generation GSM will be combined with the latest user interface developments. Third-generation handsets will evolve into powerful, flexible smart terminals.

Bringing Java to the handset

Evolution goes even further. Java is entering into the GSM industry. With its downsized Personal Java version, Java is positioned to challenge WAP in the future. A smart terminal that is fitted with the Java virtual machine will be capable of executing any kind of application.

One can imagine that a GSM handset running Personal Java is capable of micro-payments, web browsing, running audio and video applications, executing spreadsheet programs, e-mail applications, etc.

It is forecasted (by Yankee Group) that no less than 600 million smart terminals will be sold yearly wordwide by 2005. This is over three times as many PCs sold yearly today.

Clearly the future systems for using in-vehicle telematics services can be developed around third-generation GSM (smart) terminals that are fitted in the vehicle or that can be taken out from the vehicle for that matter.

Typical In-Vehicle Telematics Set-Up

Until now three major developments have influenced what in-vehicle telematics will look like soon:

- mobile Internet (mobile IP);
- third-generation GSM; and
- third-generation handsets.

A fourth source of influence can now be identified—the significant evolution of palmtop and handheld computers.

The market is divided between:

- palmtops/handhelds that run Microsoft's downscaled Windows, called Windows CE; and
- hardware based on using EPOC as the operating system, endorsed by the Symbian Consortium.

Palmtop and handheld computers are well positioned to take a role as the device to use for in-vehicle telematics services.

Microsoft has started the development of the *Auto-PC*. A device that can be fitted in any car like a car-radio, running Windows CE, offering the power of a PC to the driver.

With all these developments taking place, the in-vehicle telematics services set up will be very generic, for example consisting of:

- third-generation GSM (UMTS) connection;
- mobile IP protocols for communications;
- Auto-PC fitted in vehicles;
- additional sensor equipment in the vehicle, connected to the Auto-PC through the PC bus interface; and
- third-generation GSM smart terminals, running Personal Java, also for use when out of the vehicle.

Applications will characterise in-vehicle telematics set-up

The generic example hardware setup described in the last paragraph could run a set of applications for the user.

The set of applications that can be expected to be used in 80% of the vehicles in the future consists of:

- a voice-control application;
- a telephone application using voice-over-IP technology;
- a messaging application for e-mail, using text-to-speech technology for reading out messages and speech-to-text for dictation of e-mail messages;
- an Internet browser for accessing information and dynamically downloading new applications and services;
- a set of specific JAVA applications, developed for this particular user (home-grown);
- a set of industry defined and developed JAVA applications for general in-vehicle telematics services;
- personal radio, using IP audio streaming technology (for example, MP3); and
- video services, using IP video streaming technology (for example, RealNetworks' RealAudio)

Mobile E-Mail: An Example of Current Mobile Internet Services

The first step being taken by the GSM industry today is the offering of mobile e-mail. With this service the demand for e-mail services (regardless the user's location) is being fulfilled.

Mobile E-Mail challenges and solutions

With today's GSM standard there are basically two options for offering e-mail:

- CSD is used where the GSM user sets up a data call at 9·6 kbit/s; or
- e-mail messages are delivered as SMS messages to and from the handset.

The first option requires the user to use a PC or laptop to run the e-mail software. The GSM handset is connected to the computer with a GSM modem. Here, the GSM handset is only used for the data connection. And here the drawbacks of dropping connections, slow download speeds and hand-over trouble come in.

The SMS option forms a good alternative to connecting to a PC to the GSM handset. Firstly SMS allows the service provider to notify the GSM user that an e-mail has been received. The user does not need a computer linked to the GSM handset. Rather he/she can read the notification information on the handset's display.

The next step is to enable downloading of the actual e-mail message text to the GSM handset through SMS and then to allow the user to send commands to the e-mail service, by using SMS messages; for example, issue a command such as 'GET MAIL' to download e-mail.

A step further is to also allow sending of e-mails by using SMS messaging.

Example of a Packaged Service, Combining GSM, E-Mail and Internet: The Mobile Internet Service Provider (ISP)

Several GSM operators have already launched mobile Internet and e-mail services—including Europolitan, Bouygues Telecom and Itineris. One reason for providing this is that it makes it harder for users to churn when their e-mail address is tied to the GSM telephone service. With the upcoming of number portability, where users can switch operators and keep their GSM number, mobile e-mail gives operators a new weapon to fight churn.

GSM operators today invest in mobile e-mail and Internet platforms that have been developed by such companies as Comverse and SendIt AB. SendIt offers the comprehensive Internet cellular smart access (ICSA) platform that allow GSM operators to launch mobile Internet services in months.

In general, mobile Internet and e-mail service offerings consist of:

- an e-mail box on the GSM operator's mobile e-mail and Internet platform;
- SMS notification of incoming e-mails, with filtering options so that only notifications are sent for e-mails that meet certain conditions (for example, sender, subject, time of day, etc.);
- SMS e-mail receiving, sending, replying by SMS commands, all on the handset;
- a web interface through which the user can access his/her e-mail box, regardless of how they access the Internet; and
- a web site for self-care, where users can set their preferences.

Some operators package this service with dial-up Internet access. To enable good-quality Internet access when the user connects by using the GSM telephone (using CSD), the operators set up their own GSM access point, or GSM point of presence (GSM PoP).

This gives the operator full control over the quality of the Internet access, because:

- the operator controls the access point and the backbone network behind it; and
- the operator also provides the actual Internet backbone connection through which the users access the Internet.

Also, the operator is in control of the charges to the user for the GSM Internet connection. To stimulate traffic the operator can set this tariff to a very attractive level, comparable with a local call charge.

The GSM operators who offer these services today have created their stepping stone to offering full mobile Internet and mobile IP services in the near future.

Business Example: A Heavy Duty Construction Equipment Division

For a heavy-duty construction equipment division it is essential to identify important information and communication technology (ICT) trends in the short term, aligned to the business service strategy of the organisation's construction equipment operating companies.

Integration of information

Servicing for the heavy-duty construction equipment organisation is a complex process. It is tightly linked to service level agreements, to the history of the equipment and requests of the customer. The whole value chain, including the supply chain to the division's home locations and other suppliers, need to be integrated. The margin for the division is very dependent on the timeliness of actions—the move is to just-in-time service.

Current trends exist that have an impact:

- Intranets, with supporting applications that fit into the service areas will be a key element. The trend is to have dedicated IT platforms and IP-multimedia databases. An example of this trend is the new Oracle8i Internet server that provides low maintenance cost with high applicability.
- The Internet will become a de-facto standard for business uses. This will provide a common look and feel and promote user-friendliness.
- E-commerce is a trend that has an impact—for a heavy duty construction equipment division it is probably not deployed in the sense of web-based selling, but may be rather in the linking of procurement and authentication of parties in the transaction. In addition, customers can do self-administration directly through web access to define the specific planning of service needs.
- Remote integration of the workforce leads to the question of identification, authentication and authorisation of users. SIM cards will provide a level of security of the sensitive databases and transactions.
- The sharing of knowledge is key to servicing. Knowledge systems allow the field force to have access to relevant information and responses to their questions and queries.
- Another key element is the capability to detect trends and patterns from the information that is received from the field. The trend is to deploy datamining tools such as Omega that can be tailored to fit the purpose to detect patterns of breakdowns or correlation in service elements. These tools allow the sales-force to keep track of trends in sales and parts supplies and allow them to be proactive. It will provide relevant feedback to product development in a shorter cycle.
- Multimedia is a key area, with videoconferencing. GroupWare applications with shared user interfaces will allow group collaboration. Thin-clients will facilitate the task of distributing functionality to the field.

Mobility

The servicing on-location and in the field with tight integration in (near) real time with the back-office, warehouses and schedules implies that there are ways for mobile data communication. Repair process integration is the new way of working aimed to lower the time spent, the travel, and be ahead of events.

Current capabilities exist in GSM modems for file access and e-mail support. But the next few years show that improved datacommunications services will appear.

- The GPRS mobile data communications service will become available, first of all in the Nordic countries, in Q3 1999. The GPRS standard will allow multimedia support up to 115 kbit/s—allowing video-stills and videoconferencing to take place between locations—for review of instruction material to the field force and sharing (static) pictures of the serviced objects with back offices.
- The wireless access protocol (WAP) will provide the mobile user direct access to the Internet and become available before 2001.
- Portable PCs will become available with the direct-to-air interface on GSM standards, improving reliability—a key element for the rough environments that the organisation services.
- GSM air interfaces can be integrated to remote sensing appliances.
- Mobility and direct access to computer systems with vital information means that attention needs to be given to identification, authentication and authorisation. New-generation SIM cards will be able to be programmed such that the user can be securely accepted into corporate systems and initiate transactions. PCs can be linked to mobile handsets for these enhanced datacommunications services.
- The fast advances of datacommunication show a trend of a drop of the cost of transport.

- Where no GSM network is available, the datacommunication can be provided through satellite services such as the Iridium network (though with a cost disadvantage). Standards are being prepared for this access: the global mobile satellite services (GMSS).

Telemetrics

A key element in being pro-active is to have good information on the current status of the equipment. If the service level agreement identifies a maintenance based on hours of use, then the availability of this information is key to just-in-time servicing. Telematics applications can provide this information and provide this over the air to enable planning.

- In-vehicle telematics is a key research topic for Volvo. The information is provided over the air via GSM SMS messages. Also Siemens is developing a wide range of rugged GSM industrial grade transmitters that can be embedded in equipment and fitted with PC interfaces.
- Sensors are being developed for special purposes, such as analysis of wear and tear. Also for analysing acidity of oils, for gauging oil pressures and temperatures. Their information can be stored locally in the vehicles or equipment or transmitted to base stations for analysis.

The use of these new developments will give organisations an edge in attracting new and keeping customers, and the capabilities thus developed will enhance competitiveness in other lines of business.

Biographies

Patrick Steemers
Technology Consultant with Cap Gemini

Patrick Steemers is a technology consultant with Cap Gemini, the Netherlands. During his career in the telecommunications industry he has gained an extensive experience of the developing Internet business. He graduated from Eindhoven Technical University with a masters degree in Electrical Engineering in 1991. He has held several key-positions within Royal Dutch PTT (KPN). Since 1996, with the commercialisation of the Internet, Patrick has performed strategic consulting with international telecommunications companies for setting up Internet services. His projects have covered Europe (Belgium, Spain, Italy, Slovenia) as well as South-East Asia. Patrick is a conference speaker on issues related to the convergence of Internet and mobile communications technology.

Mark Hoogenboom
Senior Technology Consultant with Cap Gemini

Mark Hoogenboom is currently working as a Senior Technology Consultant with Cap Gemini. He was a speaker on several seminars related to web-enablement, Internet, intranet, etc. Since 1997 Mark has been working as a consultant, implementing the iterative application development method with several clients, in Japan, Korea and the Netherlands. He is a co-author of two titles in the area of iterative working: the Iterative Transformation Cycles Guide and the IAD Implementation Guide and he is a co-author of the book 'How To Become A Web-Enabled Enterprise', a book about the use of Internet-like architectures for business.

Albert Kuiper
Business Consultant with Cap Gemini

Albert Kuiper is a Business Consultant with Cap Gemini. In the 1990s he has worked extensively in developing the concepts of customer care and the relationship of telecommunications services and billing. In this role he has made numerous contributions to conferences and has conducted workshops internationally. He has developed organisational process models for operators and redefined the processes around billing for several German operators. He has managed the development of an ATM billing system in the CANCAN research and development project, that formed part of the EC ACTS framework. The work included development of service concepts and charging schemes for ATM that open up a path for commercialisation.

Alison Brady

Formulating Strategies for a Competitive Telecommunications Market

In constructing a market strategy, a telco needs to determine the sizing of the market it wishes to service. A balance must be achieved between the level of capacity and the reach of their various networks, matched to the predicted market demand for each of these networks. The decision for the telco to invest either through building its own networks or leasing from other operators, will depend on how the marketplace is changing. This in turn will be influenced by customer behaviour and how this will affect market trends. The investment decision will also be dependent upon whether the telco is an established operator or a new entrant.

This paper concentrates on the demand for residential services provided over fixed and mobile networks. In particular, the trend for some customers to disconnect from fixed networks and use mobile-only services for their telecommunications requirements will be explored. The implications for the balance that telcos need to achieve in the provision of their fixed and mobile networks are examined. An interactive model using system dynamics has been created to allow policy makers to explore different 'what-if' scenarios. Thus the implications of various proposed strategies can be investigated before a commitment to invest is made.

The model allows a holistic view of issues such as how the number of customers and usage will be affected by different tariffing structures. It also provides a method for determining the 'churn' of customers created by a competitor announcing a cut in call costs. A further benefit to a telco is that revenue per customer for the whole market can be calculated. Ultimately, a more in-depth understanding of the marketplace is developed, enabling the creation of network strategies that have a greater opportunity for successful revenue and profit generation, as well as attracting and retaining a large customer base.

Introduction

In today's fast-moving marketplaces, all companies need to achieve and maintain a strong competitive edge. This is irrespective of the types of products or services they provide. With an ever-increasing rate of change in today's telecommunications markets, this need is heightened. To achieve this, prices should be keenly priced with a low cost base and the highest possible return on investment.

The trend today is for companies to operate on an international scale. This means working in an arena where the working culture could range from being vastly or mildly different from that experienced in the potential incomer's 'home' market. Somehow an incoming operator, whether telecommunications or not, needs to be able to assess the market and decide where to position itself, its particular service offering with regard to demand, competition and availability. Prices also need to be set with regard to those of the present local competitors. Expensive mistakes can be easily made in this environment.

As it is unlikely that all consumers will prove to be equally profitable, it is important to identify which sectors are most likely to purchase the services and which sectors of the consumer market are likely to be most profitable. It is important to identify whether price reductions will boost the uptake by a particular customer sector. If so, how small can this change be and still make a difference?

In particular, the telecommunications operator needs to understand the delicate balance between the costs incurred by the customer, the demand for services and churn. Residential customers are particularly sensitive to price, so varying the tariff structure is one very strong way of influencing customer demand. It can be used to grow or restrict a particular market according to a company's ability to service that market. However, it can be difficult to appreciate holistically how the changes to one market sector would affect another.

A generic systems dynamics modelling technique, using the PowerSim tool has been developed by BT. It assesses the dynamic effects of market size, customer propensity to change, tariffing and revenue generation in a competitive marketplace.

Alison Brady:
BT Advanced Communications Engineering
Adastral Park, Martlesham Heath, Ipswich, IP5 3RE, England.
Tel: +44 1473 642758
Fax: +44 1473 620455
E-mail: alison.brady@bt.com

Understanding the Problem Space

The basis of this paper is to consider a fictional country of 10 million potential customers. The market has been split into different types of telephone user:

- *Fixed+mobile*: Users with fixed and mobile telephony are specified as definitely subscribing to fixed telephony. They may or may not also subscribe to mobile services and mobile telephony.
- *Mobile-only:* Users definitely have mobile telephony and do not have fixed telephony.

Initially, there is only an incumbent network provider who has an established fixed network and an established mobile network. The bulk of these customers use fixed lines for all communications from home, including all Internet requirements. This is partly due to cost and partly due to bad reception by the mobile receivers. So, it is altogether a more robust option to use the fixed line network for these calls. This is a growing market, with more and more of these customers going 'on-line' using the net for e-mail communications, shopping and surfing the Internet. The incumbent is satisfied that customers are happy with this, as it has deliberately made sure that its tariffing structure means that fixed calls are much cheaper than mobile ones. This reduces any demand to have its mobile network and/or handsets improved. The incumbent operator has also made a policy decision to make minimum investments in the mobile area. The fixed+mobile customers also have mobile telephone communication ability but this is very much for convenience when out and about, as the network and handset capability is more than ample for the demands of voice traffic.

Customers who have mobile-only communications tend to use the mobile telephone for convenience. The split of these customers between fixed+mobile and mobile is as shown in Figure 1.

This graph shows the tendency of the telephone population to switch away from mobile-only usage to using fixed+mobile for calls as their *average call minutes per month* rises. Consider as an example, if the population on average uses the telephone for just five minutes per month, then it would be forecast that 520 000 customers would use mobile-only and 480 000 customers would use fixed+mobile. Above 5 minutes per month average usage, more customers will use fixed+mobile compared to mobile-only. For example, at 100 minutes average usage per month it would be anticipated that 280 000 customers would have mobile-only compared to 720 000 customers with fixed+mobile. The rise in minutes usage may be due to Internet traffic or voice, but it still creates a tendency to switch away from mobile-only. This may be due to the higher stability of the fixed network for Internet calls, the level of the fixed network bills for voice or a mixture of both. When the call is of a shorter nature, then the convenience of mobile still ensures it has a reasonable take-up.

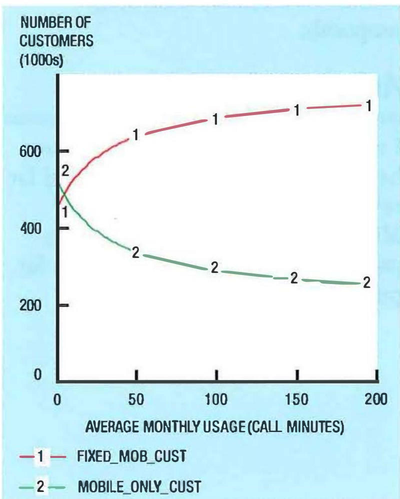
We will consider a new telecommunications operator who wishes to expand its operations internationally into this country, challenging the position of the incumbent. We will assume that this operator can legitimately operate in this country.

Scenarios

Initial market strategy

The new operator is intending to build and install only a mobile-only network rather than incur the larger costs (in both time and money) from building and installing a fixed network. Much of its own commercial success in its home country has been due to the time and money spent in developing its own mobile network and hand-unit products to improve reception. It has also noted that the incumbent is reluctant to invest in this new technology and is keen to reuse much of the development that it has already acquired. The new operator believes commercial success here depends on a combination of being able to acquire some of the incumbent's custom, both fixed and mobile, and being able to increase take up of mobile telephony overall.

Figure 1 – Incumbent operator's market base



Initially, the new operator decides to start up operations matching the incumbent's mobile tariff and relies on innovators moving across to try something new. It believes this strategy will allow a chance to gain a better understanding of the market with little risk.

Figure 2 shows the split of mobile customers between the new operator, (line 2), the incumbent, (line 3), and the sum of both of these customer bases, (line 1).

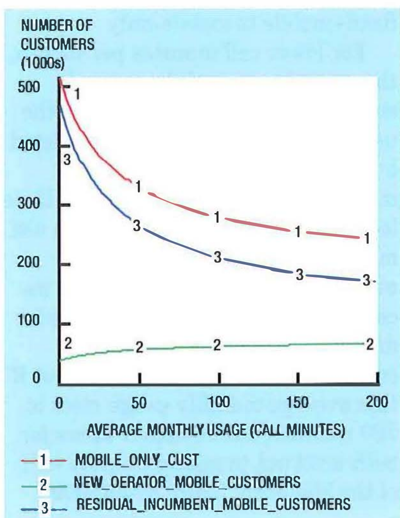
It can be seen that just by virtue of being a new operator in a previously stable market, a customer base of approximately one-fifth of the customers has been won over to the new mobile network. Assumptions made here are that there is a percentage of the customer base that is willing to trial a new product. It may be that, over time, these consumers will return to the original network operator's services, whether mobile, fixed or both.

The new network operator is therefore keen to try strategies to grow the mobile market and their share of it.

Active market strategy

The new operator is intending to compete directly on price. Mobile tariffs are going to be slashed across all market segments in an attempt to woo both mobile and fixed customers from the incumbent. It is confident that its strengths with regard to quality of service will attract and retain any Internet users who may be attracted on price rationale. This

Figure 2 – Initial market share for the incumbent and new operator



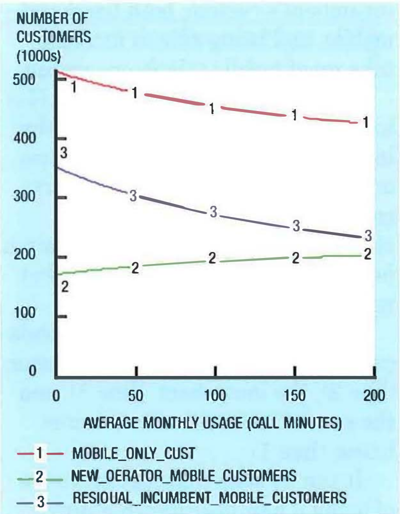


Figure 3—Market share for the incumbent and new operator during a price war

strategy is initially very successful as shown in Figure 3. This indicates the number of mobile-only customers that the incumbent and newcomer might expect to gain as a function of the average minutes usage for the population.

The previous overall mobile market levelled out at just below 300 000 customers (see Figures 1 and 2). These customers can be considered as mobile 'die-hards'. These are people who would use mobile no matter what the cost, whether for convenience, need, or image. Whereas the majority of the customer base switches to fixed once the particular average monthly usage rises above a certain point, these 'die-hards' do not.

Now, Figure 3 shows that in the price war, this market is levelling out around 450 000, an increase of 50%. These are people who have been influenced by the reduction in tariff to continue using mobile-only. This is a significant migration from fixed+mobile to mobile-only.

For lower call minutes per month, the incumbent's mobile network supports more customers but as the usage increases, people are attracted by the new operator's prices and move away from the incumbent. If we look in closer detail at users who use mobile telephone services for an average of 100 minutes a month we can see that the traditional operator manages to hold 100 000 more customers than the newcomer. But if this average monthly usage rises to 200 minutes, the customer bases for both level out to approximately half of the 'die-hard' figure of 450 000.

However, this is only an initial gain. The incumbent will try to retrieve these customers by a mixture of tariff reductions and fixed/mobile packages. The newcomer will need to exert some effort in retaining and growing their share of the market. One way of doing this would be to start identifying differing requirements for different sectors of the market and targeting each segment with particular packages: for example, cheap daytime Internet access for students and senior citizens.

Network Implications

Once we have gained an understanding of the average minutes usage of telephony by the customer base, the division of the market share between the incumbent's fixed and mobile networks and to the newcomer's mobile network can be calculated. Consequently, each operator will know how much demand will be placed on its networks. This enables calculation of network dimensioning, associated costs, anticipated revenues and ultimately anticipated profits.

Conclusions

We have demonstrated a technique that can be used to estimate the churn of customers created by the entry of new competitors, price cuts in the marketplace's tariff scheme. The effect of price on the take-up by particular market segments has also been examined. This technique can easily be extended to look at the particular nature of calls whether Internet or voice and any changes in quantities of calls. These and other 'what-if' scenarios can be created and the results interpreted as part of an overall assessment of strategy proposals.

Acknowledgements

I wish to thank Ann Matthews for her contribution to this paper and for reviewing it. I also wish to thank Mike Matthews for reviewing this paper and authorising its release for publication.

Biography

Alison Brady
British Telecommunications plc

Alison Brady received a B.Sc. (Hons) degree in Applied Physics and Solid State Electronics from Heriot Watt University in 1986. She has subsequently gained a Diploma from ACCA in Accounting and Finance and a Diploma in Management Studies from Henley Management College. Since joining BT in 1986, she has worked in a varicty of activities including design and testing high-speed Gallium Arsenide opto-electronic receivers, requirements capture predominantly for Multimedia Services and process modelling work. She presently works in the Business Modelling Group which develops simulation techniques to model stategic business problems and provide general business analysis and support to BT.

Eric M. Scharf

Meeting the Challenge of Charging for ATM

ATM provides a flexibility that is considerably greater than that to which users of traditional telephony services are accustomed. This paper will endeavour to establish a framework within which ATM charging schemes may be devised and offered to the customer. We shall look at some of the many candidates for charging parameters and will suggest how they may be grouped in order to assist the development of charging schemes and the specification of suitable facilities in ATM switching fabrics. A charging scheme structure comprising parameters, algorithms and price plans will be suggested. This paper will also look at how one may assess the merits of a particular charging scheme. Some proposed schemes will be reviewed in the context of the suggested framework and the assessment criteria. The conclusions to the paper will also provide a brief forward look to the directions that ATM charging may be taking.

Introduction

We are now entering an era in which many trial ATM networks are being established, and some ATM networks are starting to come into revenue earning service. The operational flexibility offered by ATM networks provides a challenge of new dimensions to the network operators and service providers looking for suitable charging schemes to reflect costs and to match the wide variety of services that ATM can support. A further challenge is to create charging schemes that can accommodate this new-found flexibility while at the same time being customer and user-friendly and easy to implement.

Eric M. Scharf: Department of Electronic Engineering, Queen Mary and Westfield College
Mile End Road, London E1 4NS, United Kingdom
e.m.scharf@elec.qmw.ac.uk

Charging is a customer-oriented function of communications networks. The charging process for a communications service may be influenced by at least three factors: marketing, political and technical. Marketing decisions will often take account of the perceived value of a service to the user or the need to attract users to new services. Political factors are usually associated with regulatory constraints. The cost to the service or network provider of an offered service involves traffic measurement and is thus intimately bound up with the particular communications technology being used; this aspect is the focus of this paper.

Effective charging algorithms are required to support the accounting function within the overlying telecommunications management network (TMN) and the study of such algorithms has thus been supported within the user-oriented ACTS programme by projects such as CANCAN[1] and CASHMAN[2].

The Value Chain

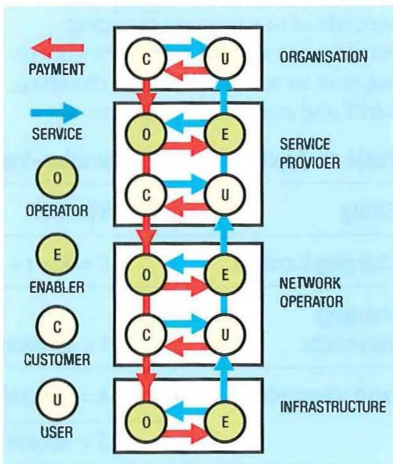
The wealth of possibilities for charging for ATM-based services calls for a structure within which such schemes may be developed and appraised. After a look at the value chain, we shall then take a closer look at how an ATM charging scheme may be structured. This will be followed by a discussion of some typical charging parameters and of how they may be classified.

In the example in Figure 1, we can consider four groupings: an organisation, a service provider, a network operator and an infrastructure provider. Of particular interest here is the payment links between the organisation and its service provider and between the service provider and the network operator.

The organisation could be a factory producing some commodity. Staff (the users) within the organisation use telephones and fax machines. The finance department of the organisation is the customer of the service supplier whose operating section it pays. If the 'organisation' is a domestic consumer, both customer and user may be the same person, or, the customers may be the parents, and the users their children! In this example, the customer may deal with the network operator directly, without going through a service provider.

The service provider gives the organisation access to the network operator and the services the operator provides. The enabler (part of the service provider) will provide the access by the user to the service provider, and may get paid by the operator party (part of the service provider) for so doing. The network operator actually runs the communications networks and the services hosted thereon. The infrastructure grouping provides the hardware and infrastructure of the network and its access lines.

Figure 1 – Example of a value chain



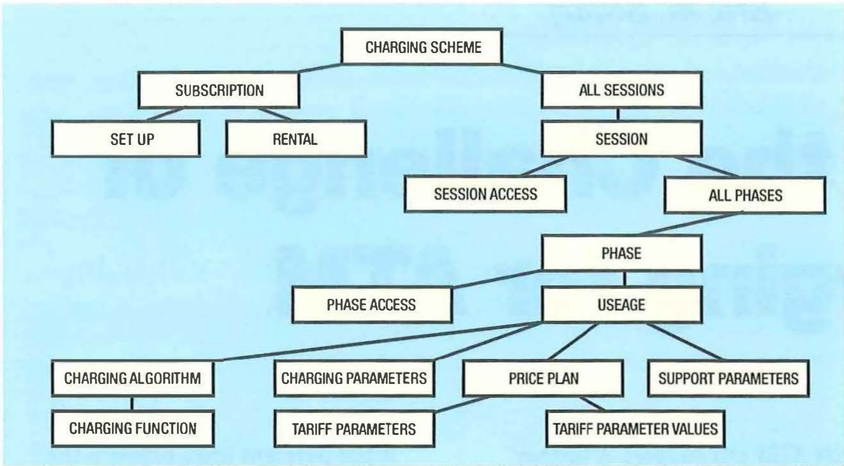


Figure 2—Charging scheme structure

Thus the enabler party within the infrastructure grouping will install the hardware and systems needed by the operator, including the hardware and systems needed to access the network.

Charging Schemes: Structure and Parameters

The subscription part of a charging scheme[3] consists of a set-up and a rental charge (Figure 2). The former is usually a flat charge and is sometimes termed the *joining fee* for a service. The recurring charge is often termed *rental* and is often a function of time. There may be several *sessions*[4] spread over the rental period. Each session may consist of several *phases*, each with its own set-up and usage charges. In this context, a phase[5] is a continuous period during which none of the following four parameters associated with an ATM connection change: quality of service (QoS), interruption of service (could be considered as a degraded QoS), time segment and traffic descriptor.

The usage part of a phase consists of a charging algorithm, one or more charging parameters, a price plan and, possibly, one or more support parameters. A *charging algorithm* consists of one or more charging functions in each of which the charge appears as a function of the charging, tariff and support parameters described below. The *charging parameters* quantify the 'amount' of service used. The price plan consists of *tariff parameters* that price the individual amounts per unit, and of tariff parameter values. *Support parameters* may be needed to complete the description of the charging scheme. An example, covering the phase access and usage part of a duration-based charging scheme is given in Table 1.

Charging Parameters are associated with network usage. Examples include volume, duration, effective bandwidth and contracted peak or sustainable cell rates. Charging parameters can have contracted values such as mean or peak allowable cell rates. These contracted values are known before a connection starts to operate and can thus be regarded as static, since the values do not change over at least the lifetime of a given phase of a connection. Other charging parameters can be measured and are regarded as dynamic (see below). Thus while the number of cells (volume of cells) *actually* passed by a connection may not be known before a connection starts to operate, this number may theoretically be measurable.

Support Parameters may be needed to complete the description of a charging scheme, but are not charged directly. Thus extra volume above some minimum agreed volume may be charged; however, the extra volume requires the measurement of the total volume which is not charged directly. Another example is instantaneous cell rate or bandwidth. This may not be charged, provided it does not exceed some agreed peak cell rate, itself a charging parameter; however, the instantaneous cell rate must be measured.

Tariff Parameters put monetary value onto service usage. Examples are Euro/cell, Francs/Mbit, and Dollars/second. The values of Tariff Parameters are influenced by market as well as technical considerations. As stated earlier, for a particular instantiation of a charging scheme, the collection of tariff parameters together with their values constitutes the *price-plan* for that scheme.

A further useful classification is in terms of static and dynamic parameters. *Static Parameters* are usually known at the start of phase and may also be known at the start of a session. Static parameters may also be fixed at predetermined times. Examples are contracted peak or sustainable cell rates. Static parameters are usually some function of:

- the objective ATM parameters as defined by the ATM Forum (for example, peak, sustainable and minimum cell rates, maximum burst size, cell delay variation, maximum cell transfer delay and cell loss ratio[6], and
- QoS classes.

Static parameters are usually based on assumptions about source statistics and are used by Connection Admission Control (CAC) to reach an accept/reject decision and to set up the Usage Parameter Control (UPC) (for example, leaky bucket) parameters.

Dynamic parameters are those that are *measured* while the connection is in progress or on the release of a connection. These parameters are used to establish the session-usage charge and examples include volume and duration. Dynamic parameters are usually measured at the point at which a user's traffic *enters* the network, but could also be measured at the receiving end.

Table 1: Typical sample of duration-based charging scheme

Entity	Symbol	Typical unit
charging function	$C = K \times t + S$	Dollars
charging parameter	t = duration	Second
tariff parameters	K = constant	Dollars/second
	S = access charge	Dollars

What Makes a Good Charging Scheme?

How do we assess the merits of a particular charging scheme? For the customer, ease of use and profitability are important considerations, while the operator may be concerned, not just with profitability but with the effect of a scheme on network performance and also with the ease with which a new charging scheme can be introduced[3,7].

Operator viewpoints

When introducing a new charging scheme, simplicity in terms of the required hardware and software will be important. How easily can the charging scheme be integrated into the existing environment comprising the host computer and the network itself? Are the measurement requirements complex or constrained by speed considerations? Does the implementation of the charging scheme depend on how it will interact with the CAC and UPC functions? Will the charging scheme entail high customer care and customer billing costs because of its complexity and the need to educate the customer about the basis for charging? Can the charging scheme be expanded easily to accommodate the changing aspirations and requirements of users, customers, network operators and service providers? To which service categories may the charging scheme be applied?

The use of a charging scheme should not adversely influence network performance. A charging scheme and its associated on-line displays require computation and network bandwidth resources. This may reduce the bandwidth and computational resources available to the other traffic carried by the network and may also influence network performance. Computation in real time may be needed to obtain usage parameters such as time and duration of connection, and mean and peak cell rate; computation may need to work with the UPC function. In addition code and data sizes and dynamic memory requirements related to the computation may need to be considered. Bandwidth may be required to request and obtain measurement information from, and to send results to, other locations in the network. How well does a charging scheme foster the sharing of network resources by a multiplicity of uses? Does the use of the charging scheme risk causing network traffic instability or congestion? The formulation of a charging scheme should help the operator to predict the effect a user will have on the network. Thus a band-limited scheme based on duration charging (for example, as used for POTS) gives the network operator information about the *maximum* requirement a user will make on bandwidth resources.

The ultimate purpose of a charging scheme is usually to generate revenue and earn profit for the supplier or operator. The scheme should therefore facilitate the identification and modification of the relevant parameters. Such mechanisms include peak-load pricing, discounts for high usage, ability to recover incremental costs, price discrimination and ability to response to elasticity in demand. A usage-sensitive charging scheme can help a network operator (NO) or service provider to relate the price he charges for a service to the use that the service makes of network resources. This can help the NO to set prices that will ensure he recovers his costs. The NO will usually appreciate a charging scheme that enables him to predict effectively the revenue generated by a customer's use of a particular charging scheme. The advantage to the NO is not only an accurate revenue forecast, but also a useful indication as to the values to apply to the appropriate tariff parameters. Uncertainty in establishing appropriate tariff parameter values could lead to values being set too high with a consequent customer rejection, or it could lead to the opposite extreme with consequent loss in profit. Indeed, this uncertainty could also lead to violent swings in the tariff parameter values, before a steady state is reached! Can the scheme be fooled by the user? How adaptable is it to changes in traffic conditions? Additionally, a scheme that depends on the content of a call can give the operator flexibility in adjusting prices.

Customer viewpoints

One customer survey (the CANCAN User Forum[8]), indicated that the criteria deemed as most important to the customer are the knowledge required, predictability, ability to audit the customer's bill, and whether or not the scheme is resource based.

How much does the user need to know about the workings of the charging scheme? Does the user need to pre-declare detailed traffic statistics, the complexity of which may not be commensurate with his mathematical knowledge? How much experience does the user need in order to maximise his benefit when employing the particular scheme?

The charging scheme should represent value to the customer in terms of quality of service and price. Knowledge of the nature of (see above), and experience in using, a particular charging scheme could help customers to maximise their gains when employing the charging scheme. This experience is fostered by factors such as: usage sensitivity, predictability, conformance to contract and ease of auditing the bill. Customers and users may like a scheme that gives them the possibility of tuning and re-negotiating prices for network resources according to demand. The prices to be paid for dynamic charging schemes are additional complexity, bandwidth requirement and computational overhead; these can increase the net costs to the customer. How secure is the scheme? For example, consider a charging scheme that samples bandwidth at regular intervals; an unscrupulous operator could ensure that traffic is only measured when traffic bursts occur. To what extent can the charge be estimated before start of the call? Is the tariff known before a call?

Some Candidate ATM Charging Schemes

We shall now turn our attention to some possible charging schemes for ATM. They all involve a need to measure charging parameters related to either time, cell volume, or both.

Three term charging

In this scheme we have a charge C composed of a flat rate term D, a duration-dependent term and a volume dependent term. In the duration dependent term E is the charge per unit time and t is the duration or lifetime of the connection. In the volume dependent term F is the charge per cell, and v is the total number of cells passed by the connection. The equation for this scheme is as follows:

$$C = D + E \times t + F \times v$$

While the tariff parameters D, E and F are likely to be contractual, the scheme, as presented, contains no contractual obligation regarding either the total number of cells to be passed by the connection or the maximum rate at which these may be passed. Hence the charging scheme, as it stands, can offer the customer surprises in the bill regarding the resultant total charge for the connection. In practice, the operator has to rely on mechanisms such as usage parameter control to limit a user's bandwidth. Such action provides a means of predicting the worse case effect a user might have on the network and can help to prevent network congestion and instability

Flat-rate charging

The flat-rate charging (parameter D) is very simple and has negligible computational and bandwidth overhead. It can thus easily be integrated into a computing environment and into the network. However, since it is not usage sensitive, there are no inducements in the scheme itself for the user to control his traffic volume. However, flat-rate charging is easy to explain to a customer who might find a flat rate charge—if is not excessive—quite attractive on account of its simplicity and potential to be profitable. It is a trivial task to audit the bill that the customer receives.

Duration charging

Duration charging, the second term, is familiar to customers since it is widely used in traditional telephone networks, albeit with bandwidth limiting. From the operator viewpoint, duration charging is very simple to set up and to explain to customers. It is best suited to CBR traffic. The bandwidth and computational overheads for running the scheme are low. Nothing is said in the contract about the traffic other than its duration and the bandwidth limiting imposed by the network. The customer gets charged for a maximum bandwidth, even if he or she does not use all of it. Hence the scheme can be profitable to the operator and it may also give the operator some flexibility in determining E, the price per duration unit. E could be based on some bandwidth other than the peak value.

Volume-based charging

In communications systems the commodity is information. One can try and quantify this by relating the amount of information conveyed by a connection to some function derived from counting the number of cells or bits actually sent in that connection. Volume-based charging may be attractive to customers since they are charged according to what they use. This is akin to the way customers are charged for mains electricity. It has been suggested[9] that volume-based charging could especially suit users of available bit rate (ABR) category services (for example, those transferring files or library information overnight) where access and transfer times are not important but space priority (no cell loss) is very important. Versions of volume-based charging are Walker's scheme[10] for wholesale charging and Viero's scheme[5] involving traffic sampling. Volume-based charging also features in schemes based on effective bandwidth and in dynamic charging schemes that are considered below.

An attractive feature of the three-term scheme is that it is simple enough to be implemented easily in silicon provided that the cell volume counters are able to accommodate the maximum expected cell rate; present technology may limit this to about 100 Mbit/s.

Effective bandwidth

The effective bandwidth of a variable bit rate (VBR) source aims to represent all the statistical characteristics of the source by only one parameter. The VBR source can then be charged as a CBR source. The concept of effective bandwidth is useful where a link has a common grade of service (GoS) parameter for each of the traffic connections and carries mainly VBR traffic, where each VBR connection expects multiplexing gains.

Possible candidates for this GoS parameter are cell-loss ratio (CLR) and probability of saturation. Irrespective of the basis for its calculation, the value of the effective bandwidth of a source usually lies *between* the source's peak and mean bandwidths. The long term average CLR is a natural choice, since cell loss is a phenomenon that users can readily appreciate and the CLR can be measured and thus can be audited after the call has finished. For bursty traffic, the probability of saturation (POS) is the probability that some arriving cells might be lost. While effective bandwidth based on the POS is easier to calculate, when compared with that derived using the CLR, it can be a very conservative estimate[5]. However, the EB based on the POS can be calculated[11,12] for a source, independently of the other sources that are going to mix with it, where all sources share a common resource such as an ATM link

The effective bandwidth is important for several charging schemes including Kelly's and Botvitch's schemes. Kelly's charging scheme[13] is based on the use of probability theory and requires knowledge of the statistics of the ATM cell traffic generated by the user's source. Kelly's algorithm assumes the user has a source for which he knows the peak bit rate as well as the probability distribution for the mean bit rate M, but not M itself. The network operator (NO) uses this information to construct a graph with axes for effective bandwidth $E(M)$ and for the mean bandwidth M of the user. A price line—

$$P = a + bM$$

—may then be drawn on the same graph and moved across it in a vertical direction. P is a price proportional to the actual effective bandwidth of the user, and a and b are proportional to charge per second and charge per volume per second respectively. A remarkable result is that the declaration m is derived from the point at which the price line P just touches the curve $E(M)$, where $E(M)$ is the effective bandwidth derived the user's source statistics with M as the unknown. The NO assumes that the user, in order to minimise the charge levied on him, declares a mean volume rate m. The NO charges the user extra for any excess effective bandwidth the user uses above $E(m)$. The user may well end up paying the NO for bandwidth above the agreed $E(m)$ value, and the NO himself can not predict the value of the mean bandwidth the user might actually use—there can be surprises for both user and operator! On the other hand, the user may use less than the declared $E(m)$ and may not get a corresponding price reduction. The question of maintaining a record of the various values of $E(M)$ over the lifetime of a connection for customer validation has not yet been addressed for Kelly's scheme. The NO must use an on-line effective bandwidth estimator[5] to determine the effective bandwidth actually used by the user.

The Botvich Effective Bandwidth Scheme[9] is based on the following formula:

$$K = \Sigma(T_i B_i) / \Sigma(T_i m_i)$$

where each connection i has a duration T_i, an effective bandwidth B_i and a mean bandwidth m_i. Σ indicates summation over all i terms. K is the aggregate factor of proportionality for i known connections. A new VBR connection with mean rate m will be charged in the same way as a CBR source whose rate is $B = K \times m$. The network has to measure on-line the mean rate of each new connection but the effective bandwidth is now derived from off-line measurements that are used to derive the factor K.

Envelope Charging

The Envelope Charging Scheme, due to Griffiths[5,9,14], offers no surprises to user or operator, while at the same time allowing for some burstiness in the user's ATM cell stream. The user is given a statistical envelope that he must not exceed. Poisson statistics are chosen because of their wide applicability and their linear additive property. The Poisson envelope can be enforced by a Poisson statistical filter, which can be implemented by a suitable combination of standard UPC components (two leaky buckets and a small buffer with negligible delay). Charging can be done on the basis of connect time, which facilitates validation; a record could also be kept of the number of occurrences of leaky bucket overflows (that is, contract violations by the user). Of all the charging schemes considered above, the envelope method is closest in concept to the original telephone duration-based charging scheme, while at the same time recognising the statistical multiplexing offered by ATM.

Dynamic charging

Dynamic charging may be defined in terms of dynamic pricing and on-line contract renegotiation[15]. On-line Contract renegotiation, in turn, can mean in-call renegotiation or spot-market re-negotiation.

Dynamic Pricing

The values of the tariff parameters of the charging scheme can vary with time. Prices varying slowly in response to inflation or competitive pressures or prices varying according to some peak/off-peak scheme are largely predictable by the customer. On the other hand, the NO may need to vary prices rapidly (for example, in response to the need for network congestion control), in order to maintain a customer's QoS, and this may lead to surprises on the customer's bill.

In-call contract renegotiation

In-call contract re-negotiation is usually a customer-controlled activity and may be viewed as a means of optimising charges and as a means of adapting the traffic contract to the bandwidth requirements of the application, possibly irrespective of the associated charges. Parameters whose values can be renegotiated may include the peak cell rate or a QoS attribute. The Distributed Pricing Scheme by Murphy at al. [16] for in-call contract renegotiation acts as a CAC mechanism and bandwidth allocation mechanism and is viewed as offering the possibility of eliminating both the CAC and UPC functions. Users (sources) compete for network resources according to their individual bandwidth requirements. For this reason, users may not be able to predict the final charge for any particular call. Network resources, such as signalling and processing influence the effectiveness of the renegotiation process[15,17]. Renegotiation with fixed phase lengths appears to be robust with respect to the exact phase length chosen, but advantages to both user and network operator vanish quickly with increasing renegotiation cost. Renegotiation with variable phase length is considerably more complicated than renegotiation with fixed phase length, but results in better transmission and renegotiation resource usage. These results are valid even if, before the call commences, there is no detailed information available on the bandwidth requirements that will arise during that call. The customer may be reluctant to renegotiate to reduce his bandwidth requirement if he perceives that any advantage thus gained is offset by the effort of the renegotiation process itself. Recently, Bigham et al.[18] have applied a system of distributed intelligent agents to the task of resource management and charging in ATM networks.

Spot Market

The telecommunications spot market is important for service providers and large corporate customers, since it will simplify buying and selling of telecommunication services, possible commodities being bandwidth/destination, time/destination and volume/destination[19]. Band-X, a London based Web exchange[20] and Rate Xchange in San Francisco[21] are leading the way in establishing spot markets for telecommunications bandwidth. A mathematical model of a telecommunication spot market has been developed[15]. Price movement in the spot market is modelled in a static fashion for discontinuous trading but this work provides a pointer to modeling the spot market in a dynamic fashion using stochastic models to represent continuous trading. A study of market-based call routing in ATM networks has been reported by Gibney et al.[22], who achieved adaptive pricing and real bidding by using distributed intelligent co-operating agents.

Conclusions

A charging scheme structure has been developed to include concepts such as subscription and usage as well as tariff, charging, static, dynamic and measured parameters. Operator viewpoints include the ease of introducing a charging scheme, its effect on network performance and its ability to generate revenue. A customer will view a charging scheme in terms of ease of understanding and profitability. Duration-based charging, especially envelope charging, looks attractive for DBR services, whereas VBR services could benefit from charging based on effective bandwidth. Purely volume-based charging may be more suited for ABR and UBR traffic. Dynamic charging is the focus of much research and could be attractive for in-call contract renegotiation as well as for spot-market renegotiation. The cost of renegotiation is an important consideration.

Acknowledgements

The author would like to acknowledge the work of colleagues in the CANCAN consortium (ACTS Project AC014). The project gave rise to a number of the concepts described in this paper.

References

1 CANCAN Consortium, Contract Negotiation and Charging in ATM

Networks, ACTS Project AC014, 1995-1998).
2 CASHMAN Consortium, Charging and Charging Schemes in Multi-Service ATM Networks, Contract Negotiation and Charging in ATM Networks, ACTS Project AC039, 1995-1998.
3 SCHARF, E. M.; and GRIFFITHS, J. M. Charging for ATM—Case Studies by the CANCAN Project, Third IFIP Workshop on the Management and Design of ATM Networks, London, 26-27 April 1999.
4 McGIBNEY et al. Contracts for ATM Services: A Structured Analysis. IEE Colloquium on Broadband Charging, IEE, London, 12 November 1996.
5 CANCAN Consortium, *ATM* Charging Schemes—Review of ATM Charging Schemes and their Performance Issues, CANCAN Deliverable D5, AC014/QMW/-/DS/P/005/b1, Document 312ds-b2.doc, October 1996.
6 LAMBARELLI, L. (Editor), ATM Service Categories: The Benefits to the User. White Paper, The European Market Awareness Committee, The ATM Forum, Brussels, May 1996.
7 SCHARF, *E. M.* First Steps towards the Development and Assessment of Effective ATM Charging Algorithms. IEE Colloquium on Broadband Charging, IEE, London, 12th November 1996.
8 CANCAN Consortium, Phase A Field Trial & User Forum Report, documents, 007ds-b2.doc and 07ads-b2.doc, October 1996.
9 CANCAN Consortium, Final Report on Static Charging Schemes and Their Performance, CANCAN Deliverable D9a, AC014/QMW/DS/P/350/a7, 1997.
10 CANCAN Consortium, Assessment of a Selection of Charging Schemes, CANCAN Deliverable D9b, AC014/LUN/DS/P/ 009/B2, May1997.
11 HUI, J. Y. Switching and Traffic Theory for Integrated Broadband Networks, 1st. Ed., Kluwer, Boston'90.
12 KELLY, F. P. Effective Bandwidths at Multi-Class Queues. *Queuing Systems Theory and Applications*, Vol. 9, No. 1-2, pp 5-15, October 1991.
13 KELLY, F. P. Tariffs, Policing and Admission Control for Multiservice Networks. 10th UK IEE Teletraffic Symposium, BT Laboratories, Martlesham Heath, April 1993.
14 GRIFFITHS, J. M.; and CUTHBERT, L. G. Very Selective ATM Filter, *Electronics Letters*, vol. 32, issue 7, 28 March 1996, pp. 637-8.
15 CANCAN Consortium, Dynamic Charging Schemes: Network Implications, CANCAN Deliverable D16, AC014/QMW/-/DS/I/370/A7, Document 370in-a7.doc, July 1998.
16 MURPHY, J.; and MURPHY, L. Bandwidth Allocation by Pricing in ATM Networks. IFIP Transactions C-24: Broadband Communications II, pp 333-351, North-Holland, 1994.
17 GRIFFITHS, J. M.; and CUTHBERT, L. G. The Impact of the Cost of Renegotiating ATM Rates, Proc. 5th IFIP ATM Workshop on Performance Modelling and Evaluation of ATM Networks, Ilkley, UK, 21–24 July 1997.
18 Bigham, J. *et al.*, *Resource Management & Charging Using a Multi-Agent System*, 3rd IFIP Workshop on the Management and Design of ATM Networks, Queen Mary and Westfield College, London, April'99.
19 *Communications Week*, 3rd Nov. 1997.
20 Band-X sets up spot market in telephone minutes, Financial Times, Monday 29 June, 1998.
21 http://www.RateXchange.com.
22 GIBNEY, M. A. et al, Market-based Call Routing in ATM Networks using Adaptive Pricing and Real Bidding, 3rd IFIP Workshop on the Management and Design of ATM Networks, Queen Mary and Westfield College, London, April 1999.

Biography

Dr. Eric M. Scharf
Queen Mary and Westfield College

Eric Scharf is a lecturer in the Department of Electronic Engineering at Queen Mary & Westfield College, University of London. Since 1989 he has been involved with European Union funded projects on ATM technology. In particular he has worked in the areas of network control and measurement as well as charging. He has also been involved with the Distributed Summer School on Broad Band Technology, held annually from 1993 to 1996. His interests are communication networks and computing, and he has written and co-authored several papers in these areas. He lectures in the areas of programming, electronics and microprocessors.

Kees Hol

'Bit by Bid by Bit'

Demand and Supply of Bandwidth through Electronic Auctions

Due to the liberalisation of the telecommunications market, the explosive growth of the Internet and its commercial use, traditional business models are not sufficient anymore to provide a representative description of the new telecommunications world. The rise of electronic bandwidth auctions, at which bandwidth services such as international leased lines, voice minutes and IP traffic are traded via a virtual market on the Internet, are an addition to the existing trade mechanisms for international interconnections. For the existence of an electronic bandwidth auction it is essential that at least one of the three parties, be it the supplying party, the demanding party or the auctioneer, gain from it in comparison to traditional means. Buyers and sellers of bandwidth benefit from lower search costs, larger market scope, and greater market transparency and efficiency.

Introduction

For many years, incumbent operators have dominated the market for international leased lines and long-distance telephony. Bilateral agreements, based upon the commonly used accounting rate system, were applied to charge these services. The accounting rate system is now complemented or partially replaced by agreements negotiated at Internet auctions. Pioneers like Band-X, RateXchange, InterXion, Arbinet, and Min-X are exploiting the Internet by creating virtual markets for the demand and supply of bandwidth services like international leased lines, long distance telephony and Internet protocol (IP) traffic.

The introduction of electronic bandwidth auctions has created a situation where bandwidth (bit) is auctioned (by bid), in an electronic way via the Internet (by bit). This article describes the traditional business of international telecommunications, related trends, and characteristics of electronic bandwidth auctions, such as the situational context, the processes, and the role and added value of a bandwidth auctioneer.

Traditional Accounting Rate System

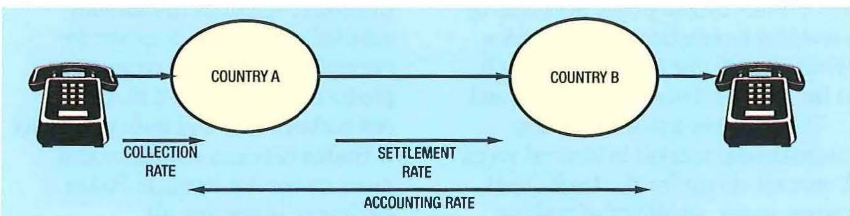
For the provisioning of international leased lines and long-distance telephony, operators need to interconnect their networks. Interconnection enables customers of a specific operator to communicate with subscribers connected to the networks of other operators. Sea cables, land cables, and satellite links connect the international gateways of operators. In the past, incumbent operators have erected international consortia for the joint implementation of these international links. Commonly, each one of the two involved incumbents 'owned' half of the leg, a so-called *half-circuit*; that is, from their gateway up to a virtual midpoint.

Incumbent operators have met, and still do meet, regularly at meetings, like the Global Traffic Meeting (GTM), for the worldwide market, and the Tariff Group for Europe and the Mediterranean Basin (TEUREM), for the European market, to discuss international interconnection. Here, operators negotiate with each other on financial and technical aspects of an international link, covering applied accounting rates, share of revenues and alignment of networks. These agreements usually remain confidential between the two operators, and are not subject to public disclosure: 'The traditional conveyance of international traffic was realised by a cartel-like "club" of corresponding national monopolist operators on the basis of bilateral agreements'[1].

The International Telecommunication Union (ITU) plays an advisory role as a third party, in the making of such bilateral agreements. By means of issuing recommendations on, for example, quality of service and billing methodology, the ITU aims to standardise international telecommunications. These recommendations are formally non-binding, but compliance is aimed for.

Figure 1 gives an explanation of the functioning of the accounting rate system used for the allocation of revenues from international calls. Two operators agree on an accounting rate for making calls back and

Figure 1—Accounting rate system.



Kees Hol
KPN, P.O. Box 30150
2500 GD The Hague,
The Netherlands.
Tel.: +31 70 3327283
Fax: +31 70 3438920
E-mail: c.hol@kpn.com

forth. In the case of a call from country A to country B, the operator in A pays the operator in B the settlement rate (often 50% of the accounting rate) to terminate the call in country B. The operator in country A charges its subscriber with a collection rate to cover at least the settlement rate. Periodically, incumbent operators check the balance of incoming and outgoing traffic. An operator with more outgoing traffic should pay the other operator the balance of outgoing and incoming traffic.

New parties, attracted by high margins, have entered the market for long-distance telecommunications. These new players are not bound by the existing accounting rate regime, and regulations have created possibilities for them to interconnect to existing operators. The emergence of competition in this market segment has forced incumbents to reduce their prices for long distance telecommunications. According to Cave and Michie[2] such a drop in prices was to be expected, as the past high prices of international telephone calls could not be explained by their costs. The accounting rate system is now put under pressure and its future existence is uncertain.

Trends in Bandwidth Trade

Alternatives for accounting rate system

Liberalisation has opened the market for long-distance telecommunications, and new ideas to bypass the traditional accounting rate system were adopted. Mechanisms like least-cost-routing and refile make use of differences between the international rates of different countries. For example, it may be cheaper to route a call from country A to C via B, rather than route it directly from A to C. Differences in international collection rates resulted in the rise of call-back services. A call-back service reverses the origin and termination sides of a telephone call so that the caller is charged with the lower rate as applied within the country of the called person. Furthermore, by means of International Simple Resale, new players bypass public networks by conveying international traffic over a private leased line and terminating it at the public network on the other end.

The Internet has affected the international market in several ways. A current threat for the traditional system is the capability of making telephone calls via the Internet, which implies a totally different pricing mechanism. Another threat is the use of the Internet as a source of information disclosing long-distance prices, for example, via a spot rate market.

Bandwidth as commodity

In addition to new methods of implementing an international service, market players have extended their transmission capacity by the laying of sea and land cables and by enhancing the used transmission technology. Currently, consortia and companies like TAT-14, FLAG, Global Crossing, Oxygen, and Gemini are laying miles of high-bandwidth optical-fibre cables throughout the world. More flexibility concerning pricing, contract duration and contract types, like rental, lease, future and option constructions, are available now.

The book *The Death of Distance* by Cairncross[3] postulates that prices for long-distance calls will drop radically due to the overwhelming capacity available on the international and long-distance networks. Furthermore, the costs for providing such networks have fallen as well. As an example, in the period 1986 to 1996, the capacity on the transatlantic routes between North America and Europe (both cable and satellite capacity) exploded from 100 000 to more than 2 million voice channels. 'Eventually, it will cost no more to telephone from Hollywood to London than to telephone to nearby Beverly Hills'[3].

Telecommunication services, and in particular bandwidth services, have become a normal part of life; services are widely available, the number of services has increased greatly, and prices for many services have dropped rapidly. In a sense one can say that bandwidth has become a commodity.

Electronic Auctions

The widespread Internet enables users to create new and innovative commercial opportunities. One of such opportunities is the implementation of an electronic auction. Electronic auctions are already available in various formats: for example second-hand consumer products are auctioned at eBay (www.ebay.com), and bulk electricity is traded between suppliers and users on the Amsterdam Power Exchange (www.apx.nl).

An electronic auction is a specific type of auction that makes use of an electronic infrastructure, in order to form a virtual market where products or services are traded by means of an auction mechanism. McAfee and McMillan define an auction as 'a market institution with an explicit set of rules determining resource allocation and prices on the basis of bids from the market participants'[4].

Bids and offers

The trade on an auction consists of bids and offers submitted by its users. A bid is a proposal to buy a product or service at a quoted price, according to specific terms and conditions. An offer is similar to a bid but it is now a proposal to sell something at a quoted price. An offer (bid) may be followed by one or more bids (offers), or it may be accepted immediately.

Auction types

Electronic auctions may be clustered by the motives for having such auction[5]. For the electronic trade of bandwidth, the use of an auction for price determination, efficient allocation, and attention and visibility of the bandwidth service is of most importance. For product price determinating several mechanisms are available. In an English auction bidders compete with each other by successively increasing the price until only one bidder remains. In a Dutch auction, the bidding process starts with an initial price, which is lowered every certain time-interval until the first bidder accepts the quoted price. In the case of *straight sales*, the first approved bidder gets the deal at the listed price. An auction may also be characterised by its accessibility. At open auctions, bids and offers are made publicly, while the bids and offers at closed auctions are made privately and remain secret[6].

Klein distinguishes potential objects traded on auctions into three broad categories: commodities, perishable products, and products with a limited availability[5]. For bandwidth, the first two characteristics are applicable. The value of perishable goods will drop to zero at some known point in time, as is the case with unused bandwidth.

Electronic Bandwidth Auctions

Since liberalisation, the number of operators in the long-distance market has increased, and thereby the potential number of relations and interconnection agreements to be made. The number of relations can be expressed by the formula:

$$\frac{n \times (n-1)}{2}$$

with n corresponding to the number of operators. For example, assuming only 15 different operators (incumbents or new operators that need to interconnect to each other) already implies more than 100 bilateral relationships, which will be complicated and costly to manage. As it is now very difficult for all of the operators to find out about each other, it is evident that intermediaries appeared to reduce this clew of interconnections, see Figure 2. Furthermore, as bandwidth is perishable and as it may be regarded as a commodity, bandwidth services are well suited for trade on an electronic auction.

Intermediaries such as bandwidth auctioneers run a location for the physical interconnection of bandwidth auction users (often done via a switch or router), supported by a virtual trade floor on the Internet. Auction users need to have their own physical link or leased line to this location. On the virtual trade floor, an Internet web site, buyers and sellers are brought together. Customers with unused bandwidth or excess capacity on a link from the auction to another specific destination, may offer such services on the virtual market or may search on this market for parties that have indicated to be in need of such capacity. Both potential buyers and sellers are capable of submitting a bid or offer on the virtual trading floor. In this way, a transparent and efficient market is constructed.

Market players

Currently, at least the following players are active in bandwidth auctions via the Internet:

- Arbinet (www.arbinet.com and www.acgn.com),
- Band-X (www.band-x.com),
- InterXion (www.interxion.com),
- Min-X (www.min-x.com), and
- RateXchange (www.ratexchange.com).

Nearly all of these companies are between one and two years old. Their main offices are located in the United Kingdom (Band-X), the Netherlands (InterXion), and the United States (others). However, some already have extended their offices to Asia, Australia, South America and several countries in Europe. Presently their customer focus is on the wholesale market of operators, resellers, ISPs etc. To avoid potential conflicts with their customer base, they generally stay out of the retail market.

Products and services

The most common services delivered by bandwidth auctioneers are the facilities for the trade in international leased lines and voice minutes. For voice minutes the auctioneer's switch will be used for routing the call. IP services, and in particular Voice over IP, are more recent. Depending on their strategies, some auctioneers also deliver their customer supplementary services like negotiation support, clearing of transactions, monitoring of the quality of service, billing, and invoicing. For the provided services, the bandwidth auctioneers commonly charge their customers with a certain percentage of the deal or traded number of voice minutes.

In addition, some auctioneers also offer customers equipment space by means of so-called *carrier hotels*. Customers may use this space to install network equipment for their own purpose.

Users

In order to be able to trade on the virtual trading floor, potential customers should register themselves for a membership and, in some cases, they have to sign a paper contract in advance. The awarding of a membership gives the user the capability to browse the bandwidth auctioneer's web site for details on the latest bids and offers. Furthermore, the member is allowed to place a bid or offer. Initially, the trading parties are anonymous; however, after having shown interest in a bid or offer, the bandwidth auctioneer may bring the parties involved together. In other cases, the auctioneer will function as an intermediary, with the parties remaining anonymous.

Modelling Electronic Bandwidth Auctions

Kambil and Van Heck[7] describe a generalised model on exchange processes within electronic auctions, see Figure 3. These exchange processes, split into basic trade processes and trade context processes, are identified as being present in all transactions within electronic auctions. Basic trade processes include the search for product or customers, the valuation of offered products, the related logistic activities, the payment and settlement methods, and the authentication procedures. The trade context processes comprise processes that reduce the risks of trading (that is, predetermined product representations, legal arrangements, influence mechanisms within the institutional context) and a system for dispute resolution. The communications and computing activities within electronic auction environments are the glue between the two different types of processes. Due to the general applicability of this model, here it is used for the modelling of exchange processes within electronic bandwidth auctions.

Search

A party looking for a potential buyer for its offered product or service, or a

Figure 2—Relations between operators via an intermediary.

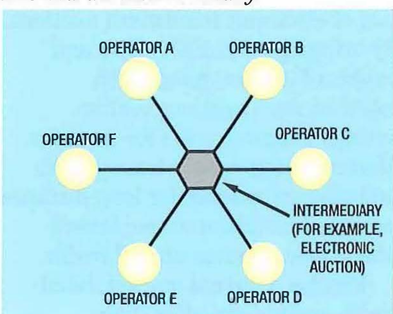
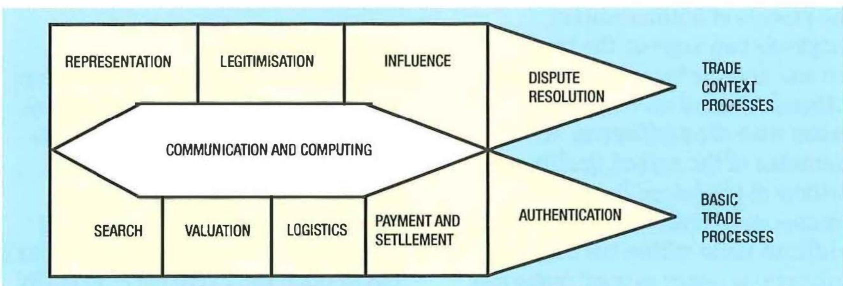


Figure 3—Exchange processes within an electronic auction.



party looking for a selling party which is able to fulfil its own product demands, faces search costs. An electronic bandwidth auction enables a reduction in search costs for bandwidth, as relevant players are present at the virtual trading floor. The measure of success in a match depends on the number of registered users willing to trade and on the number and quality of submitted bids and offers.

Valuation

The valuation of bandwidth can be defined either by the initiating supplying or demanding party. This valuation may be subject to a price negotiation. To inform their users, some auctioneers provide indices to keep track of bandwidth and voice minute prices. These indices may be based upon specific international routes, or on prices settled on the trade floor. The use of an index makes it possible to design future and option constructions for bandwidth. Users can protect themselves against unpredictable variations in future bandwidth prices. So far, auctioneers have not yet implemented such constructions.

Logistics

As bandwidth is an item which is moved electronically, the logistics involved within bandwidth trade are limited to the physical connection of a user's link to the location of the auction's switch or router. Users need to have their own copper, fibre optic or wireless connection or needs to hire a leased line from a third-party operator. The management and configuration of switched or routed links may be considered a logistic process as well.

Payment and settlement

The payment and settlement of closed deals may be done directly between the buyer and the seller, or the bandwidth auctioneers can offer this as an added-value service. In this way, the auctioneer acts as a kind of clearing house.

Authentication

The process of authentication comprises two aspects: the traded services and the trading parties. The authentication of services is concerned with the verification and guarantee of the agreed quality and features of the leased lines, voice minutes or IP link. This aspect is still a difficult issue within the telecommunications sector as good measures of quality are not available or are hard to guarantee. Common measures for quality are answer seizure ratio (ASR), post-dialling delay (PDD) and voice compression ratio; however, these measures may be insufficient to monitor a link. As the physical route of a service may pass through several operator networks, the applied voice compression ratio may differ along the line and a ratio worse than that agreed may be hard to detect. Some auctioneers guarantee the buying customers the agreed level of quality, but they also have strict demands on sellers offering bandwidth, concerning this issue. The availability of a standardised bandwidth unit may improve the authentication of traded services.

The authentication of trading parties is concerned with the integrity of the customers. In principle, the auctions are available to all Internet users, however, the auctioneers authenticate their customers by means of a compulsory registration, or the signing of a paper contract. In some cases the auctioneer keeps track of a (public) black list of suspicious users.

Communications and computing

The Internet and the web page form the communication and computing fundament for the electronic bandwidth auction. The Internet is the communication path to the auctioneer, while the essential computing power is at the web server site of the auctioneer. Additionally, other (administrative) applications may be used to ease the transactions.

Representation

Representation is concerned with the description of the traded services. To express the type of demanded or supplied service, characteristics like the two end-points of a link, its transmission speed, the compression ratio, and contract duration may be specified. Although these terms are familiar, no global definition is available for a standard representation of such service on an electronic bandwidth auction.

Legitimisation

In order to validate a trade between two parties, the bandwidth auctioneer needs to apply some legitimisation method to formalise the deal. Initially, the two parties are kept anonymous; however, after having shown serious interest in each other's bid or offer, the auctioneer generally brings the two players together for further negotiation and contract making. In general, auctioneers do not publicly disclose any information about the traded deals.

Influence

To ensure fair trade, the auctioneer influences the trade by acting as a gatekeeper and as a guard. To reduce any risk, new entrants to the bandwidth auction have to register themselves or have to sign a contract in advance. Moreover, auctioneers may keep track of a (public) blacklist to exclude unreliable parties.

Dispute resolution

In case of dispute, the bandwidth auctioneer may act as an intermediary, depending on the type of issue. Some auctioneers make themselves responsible for a certain quality of service, avoiding a potential dispute between two of its customers. However, mechanisms for arbitration or legal proceedings, if required, are to be defined within the contract between the parties involved.

Conclusion

Due to liberalisation in the telecommunication market, the traditional way of pricing long-distance telecommunication services is under pressure. The entry of new long-distance providers made this market highly competitive and new service provisioning alternatives, like least-cost-routing, refile, callback services, and International Simple Resale, undermine the accounting rate system, as applied by the incumbent operators. Several consortia and companies have intensified the laying of high-bandwidth optical-fibre cables throughout the world. Due to the aforementioned, and the progress in transmission technology, bandwidth has become a commodity.

Liberalisation has turned out to be the main driver for the change in long-distance telecommunication services. Supported by the rise of the Internet, it has created new commercial opportunities like the introduction of electronic bandwidth auctions. By bringing potential sellers and buyers of (excess) bandwidth together, such auctions realise savings in search costs for its users. Moreover, it creates a transparent and efficient market for long-distance services like international leased lines, voice minutes and IP traffic.

Besides a virtual market, bandwidth auctioneers offer their registered members various supplementary services like negotiation support, clearing of transactions, monitoring of the quality of service, billing, and invoicing. In order to ease the trade, to guarantee quality of service, and to enable dealing in future or option contracts for bandwidth, the introduction of some kind of standardised bandwidth unit is recommended.

References

1 VLAAM, H. DE, Cross-Border Interconnection. *Br. Telecommun. Eng., Jul. 1998,* **17**, pp. 37-42.
2 CAVE, M.; and MICHIE, J. Developing Competition in International Telephone Service. In Telecommunication, New Signposts to Old Roads, KLAVER, F., and SLAA, P., pp.7–18, Amsterdam, IOS Press, 1992.
3 CAIRNCROSS, F. The Death of Distance, p. 27. The Orion Publishing Group Ltd., London, 1997.
4 MCAFEE, R.; and MCMILLAN, J. Auctions and Bidding. *The Journal of Economic Literature.* 1987, **25**(2), pp. 699-738.
5 KLEIN, S. Introduction to Electronic Auctions. *International Journal of Electronic Markets.* Vol. 7 (1997), No. 4, pp. 3-6.
6 RECK, M. Trading-Process Characteristics of Electronic Auctions. *International Journal of Electronic Markets,* **7**(4), pp. 17-23.
7 KAMBIL, A., and HECK, E. VAN, Reengineering the Dutch Flower Auctions: A Framework for Analyzing Exchange Organizations. *Information Systems Research,* 1998, 9(1), pp. 1-19.

Biography

Kees Hol
KPN (Royal Dutch Telecom)

Kees Hol is a student at the Business Administration department of the Open University in the Netherlands. He is working on a graduation thesis about telecommunications bandwidth sold through electronic auctions. In 1993, he obtained a degree in Electrical Engineering at the Delft University of Technology. He has worked for KPN Research and Siemens, and recently he joined KPN as an international account manager responsible for the sales of Internet services via satellites.

Cornelis H. Hoogendoorn

Multiple Networks, One Access

The past few years have seen the rapid rise of the Internet to become a second universal network alongside the telephone network. Already the Internet traffic (measured in bits) exceeds that of the telephone network. Network operators are responding by making large investments in data networks and by looking for synergy opportunities through converging the networks. Futhermore, as competition increases, not only within a certain kind of network but also across heterogeneous networks, offering converged voice/data services is seen as a means to increase market share. However, even though data has overtaken voice traffic, voice services are still the main revenue source for carriers and should therefore not be disrupted by convergence. As the cost of bit transport in backbone networks continues to drop, the access represents an increasing part of a carrier's overall investment. It is of cardinal interest to maximise the return on this investment by turning today's access (the telephone line) into a single high-bandwidth line for voice, data and video services. This coincides with the interests of network users, who are best served by a single point of access to a converged set of services which use the full capabilities of the underlying networks. The paper explores the implications of these trends on the evolution of access, services and the networks themselves.

Cornelis H Hoogendoorn:
Siemens AG, Information and Communication Networks, ICN WN CS SE, D-81359 Munich, Germany
Tel. +49 89 722 22204
Fax. +49 89 722 27284
E-mail:
cornelis.hoogendoorn@icn.siemens.de

Introduction

The past few years have seen the rapid rise of the Internet to become a second universal network alongside the telephone network. Already the Internet traffic (measured in bits) exceeds that of the telephone network, and this trend looks set to continue. However, even though data has overtaken voice traffic, voice services are still the main revenue source for carriers, estimated at between 80% and 90% of the total. For this reason data network operators and equipment suppliers are increasingly focusing on the provision of voice services over data networks. Traditional network operators are responding to this trend by making large investments in data networks and by looking for synergy opportunities through network convergence. Futhermore, as competition intensifies, not only within one kind of network but also across heterogeneous networks, operators are looking for differentiating factors to increase market share.

A second trend is the dramatic drop in the cost of bit transport in backbone networks, fuelled by technology innovations and giving rise to the 'Death of Distance'. Distance-independent charges are of course well-known to Internet users, but are increasingly becoming a fact of life in the PSTN as well, where long-distance tariffs are being slashed and distance-dependent tariff structures are drastically simplified or even eliminated. This has been recognized by the Voice-over-IP community, where a large degree of consensus exists that the window of opportunity for rate arbitrage is rapidly closing. Cheap telephony with reduced quality voice will not justify widespread deployment of voice-over-IP. As tariffs become comparable, as well as low in absolute terms, users will demand at least today's voice quality and convenience. Competition will therefore shift from price wars to service differentiation. Users will be attracted by converged voice and data services which increase convenience in their daily communication and which save time and effort. Carriers offering this will be rewarded with increased income from services and additional call minutes, as well as with increased customer loyalty or 'stickiness' in Internet terminology. In fact, services may be given away free for increasing stickiness of other services[1].

As the cost of bit transport in backbone networks continues to drop, the access represents an increasing part of a carrier's overall investment. It is therefore of cardinal interest to carriers to maximise the return on this investment by turning today's access (the telephone line) into a single high-bandwidth line for voice, data and video services. This coincides with the interests of network users, who are best served by a single point of access to a set of converged voice-data services which use the full capabilities of the underlying networks. On the other hand, users generally will neither know nor care how the various media streams comprising the services they are using are transported in the network; for example, by time division multiplexing (TDM), Internet protocol (IP) and/or asynchronous transmission mode (ATM). Service and access convergence will therefore happen much faster than convergence of the networks themselves.

Service Convergence

Service providers of all types are seeking solutions for delivering high-profit, differentiated voice and data services. The incumbent carriers, however, have an additional concern. They already have huge amounts invested in traditional TDM switching, an investment that cannot be cast aside simply for the sake of new technology. Incumbent carriers realise that to remain successful, they must somehow preserve the best of their existing networks, while simultaneously adopting newer technologies to, in the first instance, support data services. In other words, they must find a way to leverage their existing circuit-switched investment through interworking and combination with packet-based data networks.

Emerging carriers, while seeking newer packet-based technologies, clearly understand that their networks must be capable of interfacing with the existing public switched telephone network. Another key concern for these new carriers is determining which of the thousands of features available through today's networks their customers will need. These new service providers, must look for ways to construct their networks that allows them to provide their customers access to the rest of the world while at the same time delivering all of the services and features that their customer base has come to rely upon.

Both incumbent and new carriers need a strategy that offers an economical path to ubiquitous service delivery and growth. To address competitive pressures and achieve their long-term goals, service providers require networking solutions that:

- allow for the rapid introduction of new competitive features;
- extend the revenue-generating life of existing networks;
- provide a cost-effective way of bridging their TDM and ATM/IP domains;
- take advantage of the existing line-side revenue producing features;
- deliver a future-proof solution; and
- operate in a multi-vendor environment.

For the service provider, one of the primary drivers for network evolution is revenue generation, whether from new or existing streams. The packet-based networks of the twenty-first century, capable of combining voice and data traffic, will enable a new set of services that could not be supported by the existing TDM circuit switched networks alone, for example:

- *Single-Stage Dialling Voice-over-IP (VoIP)*: Dial an IP call with a simple access code;
- *Virtual Private Networking*: Provides business customers with cost-effective, secure private networks using shared network resources;
- *IP-based Centrex*: Creates mixed VoIP and wireline Centrex groups;
- *Voice over DSL*: Delivers multi-line presence using a single copper pair;
- *Unified Messaging*: Retrieves both voice mail and e-mail from either PC or telephone;
- *E-Mail Notification and Voice Delivery*: Notifies users over the telephone when they have received e-mail, and receives messages over the telephone;
- *Subscriber Controlled Input*: Allows subscribers to conveniently change their telephony feature usage profile over the Web whenever and as often as they like;
- *Click to Dial*: Web pages with *call-me* buttons allow surfers to trigger a PSTN call setup over the Web;
- *IP Call Conference*: Set up a large telephone conference through the Web and receive notification of who is currently speaking;
- *Web Service Subscription*: Select a subscriber service over the Web and receive it immediately;
- *Voice/Data Collaboration Service*: Allows colleagues in different locations to collaborate on document/presentation development while communicating in real-time;
- *Multimedia Conversations*: School students discussing homework assignments, teenagers discussing (digital) pictures taken at yesterday's party;
- *Multimedia Call Centres*: New multimedia centric call centres allowing an end user to see a product and talk to a customer service representative over a single connection;
- *On-line Billing*: Check current call charges over the Web;
- *High-bit rate Internet Access*: Via xDSL; and
- *Remote Access Services*: Provides dial-in access offering for corporate accounts.

The major impact of any network solution on a service provider's bottom line is not the underlying technology, but rather the flexibility to deliver such network services. Solutions that support differentiated and innovative services, allow for the rapid introduction of new services and, most importantly, enable interoperability with the existing public voice network. What should be pointed out here is the different ways in which services have been provided in voice and in data networks. In the traditional voice-telephony networks, services and features are 'inside'—part of the network itself—usually implemented in a proprietary manner by the equipment vendor and owned by the network operator. This is the case irrespective of whether the service logic is implemented in the telephone switches themselves or in centralised computers (as in intelligent networks). In universal data networks (Internet), the services are 'outside'—implemented in computers (servers) at the edge of the network—based on industry standard (open) platforms, and most frequently not owned by the network operator but by independent service providers. This paradigm, with its promise of fast feature delivery and the ability to capitalise on the creativity of thousands of independent service suppliers, will also be applicable to future converged voice-data services. Traditional voice switches will therefore make use of commercial platform technology to allow client-server access to the large base of voice services, thus avoiding extremely costly re-implementation and preserving the stability and performance which is the hallmark of the telephone network. The same approach can make these voice services and features available in other networks, (for example, voice-over-IP) with the following advantages:

- supports today's complete telephony feature set,
- permits the fast introduction of new revenue generating features,
- provides for differentiated features,
- carrier-class, highly reliable, feature-rich solution,
- supports feature interworking,
- offers open interfaces
- provides cost-effective TDM to packet network integration, and
- provides TDM features to IP and ATM access devices.

Access Convergence

Although the access area has seen tremendous development over past years, it is still the case that for most residential users and small businesses access to both voice and data services is via the telephone access line, not least because of the enormous embedded base of such lines, making wholesale replacement by other technologies very costly. Carriers must therefore be offered means to squeeze the most additional revenue out of this huge investment and to offer competitive high-speed access to data services. Here the various forms of digital subscriber loop (DSL) technologies come into their own.

xDSL solutions require, as the network prerequisite, copper twisted-pair cables and thus offer incumbent operators a good way of upgrading their network towards higher bandwidths. Universal asymmetric digital subscriber line (UDSL), symmetric digital subscriber line (SDSL), and asymmetric digital subscriber line (ADSL) need no upgrade of the copper networks at all, but only equipment at the central office and at the customer premises. The plain old telephony system (POTS) and some integrated services digital network (ISDN) traffic can be transmitted together with the broadband signals using frequency-division multiplexing. At the customer premises either a network termination or a PC card has to be installed. Additionally ADSL needs a splitter for separation of POTS and xDSL signals. UDSL (also known as G.lite) is well suited for residential mass market due to its splitterless nature, which allows installation at customer premises without truck roll out. SDSL is intended for upgrading ISDN lines.

At the central office side a digital subscriber line access multiplexer (DSLAM) is used which contains the modems and a concentrator to reduce bandwidth needed within the core data network. Looking at stand-alone systems, for every subscriber who wants to have broadband service, the copper cable from the main distribution frame has to be connected to a splitter to be able to add the broadband signal and has then to be rerouted to its narrowband line card. This requires skilled personnel as well as additional infrastructure and space for splitters and broadband equipment. The alternative to a stand-alone DSLAM is a switch-integrated xDSL line card which offers narrowband telephone service and broadband xDSL service at the same time. No external splitter is required. Offering broadband service to a subscriber now only requires the exchange of a line card. As xDSL is mainly intended for an upgrade of existing telephony networks, such a solution with minimum impact with respect to mechanical rearrangements, service interruption and introduction of new management procedures and systems has clear advantages.

The benefits of an integrated voice/data access are obvious:

- It provides the basis for converged voice/data services, by providing parallel access to PSTN and data networks, as well as interworking between them. Service convergence is enhanced by integrated authentication, operations, administration and maintenance (OAM) and billing.
- It is fast and flexible, permitting services to be changed without the need to change the access, thus providing a short time-to-market for new services. Additionally, it offers the subscriber unrestricted choice between circuit-switched and/or packet-based services.
- It is economical, as it provides for the reuse of existing highly reliable infrastructure and, with suitable design, minimised installation costs through plug and play provisioning.

Conclusion

Network convergence comprises many aspects. It has been argued here that service and access convergence will occur much faster than network convergence, as the former provides most utility to the end user and most revenue opportunities to the network operators. Backbone network convergence will most likely occur over a much longer time frame, as its benefits are not directly visible to the end user. For the network operators, on the other hand, long-distance traffic accounts for a declining proportion of revenue. Expansion in the backbone network will therefore be justified by the demand for new data services with rapidly increasing bandwidth demand, rather than the early replacement of the voice network infrastructure. Both of course will continue to profit from the continuing advances in transport technology such as wave division multiplexing. This leads to the conclusion that he network of the future will be a network of cost-optimised heterogeneous networks with full interworking capabilities, providing users with a rich, seamless set of converged voice/data services via a common access.

Reference

1 Tolli, Dominick (AT&T). Presentation given at the 'Voice on the Net' conference, Las Vegas, 13–16 April, 1999.

Biography

Dr. Cornelis H. Hoogendoorn
Siemens AG, Information and Communication Networks

Dr. Hoogendoorn studied Electrical Engineering and Computer Science at the University of Stellenbosch, South Africa, Stanford University, United States of America, and the University of Cambridge, United Kingdom. He joined Siemens in 1987 after serving as Chairman of the Computer Science Department at the University of Witwatersrand in South Africa. He holds a variety of patents for his work on the architecture of switching systems. Dr. Hoogendoorn is currently Senior Vice-President responsible for Systems Engineering, Carrier Switching Networks, in the Information and Communication Networks division of Siemens AG.

Gaetano Vespasiano and Maria Stella Iacobucci

Simulation of Asymmetric Digital Subscriber Line

Impact of line characteristics on system performances

The growing demand to transmit high-speed digital data down copper pairs has resulted in a wide variety of transmission systems that have to co-exist on the same or adjacent binder groups. Asymetric digital subscriber line (ADSL) is a technology that can be used to convert the copper cable access line into a high-speed digital link and avoid overloading the circuit-switched public-switched telephone network (PSTN). This paper addresses the problem of maintaining spectrum compatibility between ADSL and other services that may use different transmission technologies.

Introduction

ADSL is a technology that allows broadband access from residences or small offices to the local exchange. It has the potential to supply residential and small business users with all types of new broadband services, from educational to financial, and customers would essentially need only a new ADSL modem, allowing up to 8 Mbit/s downstream (to the end user) and up to 1 Mbit/s upstream.

The American National Standards Institute (ANSI) T1E1.4 Standards Committee has selected the discrete multitone (DMT) as the standard modulation scheme for asymmetric DSL[1].

A multitone modulation scheme, such as DMT, has the flexibility to optimise the transmission capacity and the power spectrum over more than one (disjoint) frequency band.

The results presented here are for the specific application of DMT to carry ADSL payloads of over 8 Mbit/s from the network to the customer. Spectral compatibility between ADSL, high bit-rate DSL (HDSL), and integrated services digital network (ISDN) basic-rate access (BRA) systems are considered in two different Italian network scenarios—urban and rural areas.

Gaetano Vespasiano
Tel: +39.0862.336508;
Fax: +39.0862.336542;
E-mail: gaetano.vespasiano@ssgrr.it
Maria Stella Iacobucci
Tel: +39.0862.336616;
E-mail: mariastella.iacobucci@ssgrr.it
Scuola Superiore 'G. Reiss Romoli', via G. Falcone 25, 67100 L'Aquila, Italy.

ADSL System

Figure 1 shows the ADSL network model. ADSL shares the existing telephone line with the plain old telephony system (POTS); at each end there is a set of filters that split the line by frequency: a low-pass filter passing telephone signals and a high-pass filter passing ADSL signals above 25 kHz. The high bit rate digital signals do not overload the existing telephone network because they are connected to a broadband network; that is, an asynchronous transfer mode (ATM) network.

Figure 2 shows the ADSL spectrum allocation. ADSL systems are planned to have two different spectrum allocations: frequency division multiplexing (FDM) between upstream and downstream, and echo cancellation with frequency overlapping.

ADSL Simulated System

Using DMT, the channel bandwidth is divided into N independent subcarriers, of bandwidth W, and n_i bits are assigned to each positive frequency sub-channel.

In the encoder, each set of n_i bits is mapped into a complex subsymbol, which forms the quadrature amplitude modulation (QAM) constellation for that subchannel. An N-point fast

Figure1 – ADSL network diagram

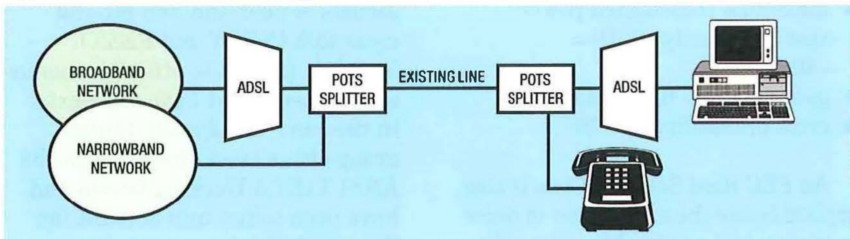
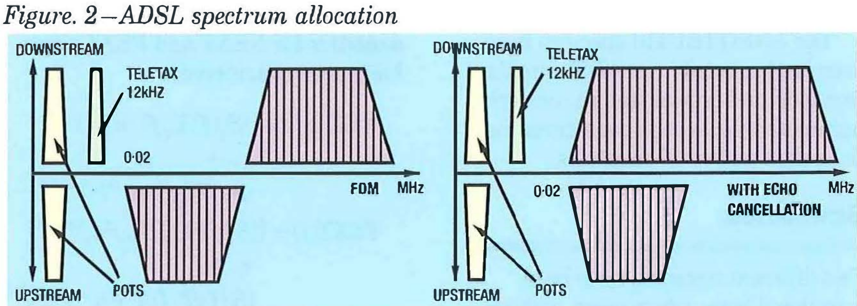


Figure. 2 – ADSL spectrum allocation



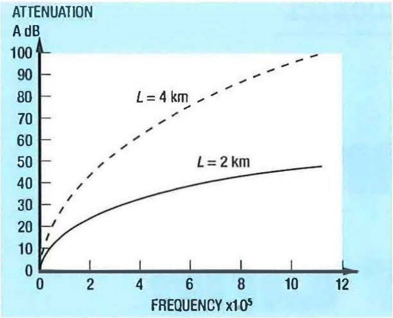


Figure 3 – Cable attenuation versus frequency

Fourier transform is used to obtain the modulated signal. As the sub-channels are not independent for finite N, a cyclic prefix is used to remove the intersymbol interference (ISI) between the sub-channels[2,5].

The overall bit rate R_b is given by

$$R_b = \sum_{i=1}^{N} n_i W$$

The power of each QAM signal tone is equal to P_i and the total transmitted power is:

$$P = \sum_{i=1}^{N} P_i$$

The multitone modulation allows R_b to be maximised through an optimal choice of power and bits per symbol for each tone, under the restriction that the symbol error probabilities for all tones be equal and the total transmitted power P be limited[5]. The ANSI T1E1.4 standards committee selected the discrete multitone modulation for the implementation of asymmetric DSL, with the following parameters:

- W = 4 kHz;
- maximum n_i = 14 bits;
- maximum transmitted power spectral density (PSD) = –40 dBm/Hz;
- gain margin > 6 dB; and
- error probability $P_e < 10^{-7}$.

An FEC Reed Solomon Code is also applied before the modulation in order to reach a higher signal-to-noise ratio (SNR) at the receiver side[1].

The ANSIT1E1.413 specifics have been implemented in an ADSL simulation tool[3] and typical Italian network parameters have been considered for the urban and rural scenarios.

Scenarios

Two different scenarios have been considered here: urban areas, with an average copper loop length (L)=1000 m and rural areas, with an average copper loop length (L)=2000 m; in both cases a copper wire diameter of 0·4 mm has been considered. The cable length assumed for the implementation is twice the average value in both environments, because in the Italian network more than 90% of cables length is below 2 × (L). The copper pairs attenuation is plotted in Figure 3 for both the urban and rural cases, showing the proportional dependence on the square root of the frequency and the cable length:

$$A_{dB} = \alpha\sqrt{f} \times L$$

Interference Model

Models of near-end and far-end cross talk (NEXT and FEXT) interference due to others transmission systems that have to co-exist in the same or adjacent binder groups have been developed in the ANSI T1E1.4 Working Group and have been taken into account for the simulation[4].

The following power spectral densities for NEXT and FEXT noise have been considered:

$$\text{NEXT}(f) = \left((S_1(f)X_N f^{3/2} n_1^{0.6})^{1/0.6} + \ldots + (S_I(f)X_N f^{3/2} n_I^{0.6})^{1/0.6}\right)^{0.6}$$

$$\text{FEXT}(f) = \left((S_1(f)H^2(f)X_F f^2 n_1^{0.6})^{1/0.6} + \ldots + (S_I(f)H^2(f)X_F f^2 n_I^{0.6})^{1/0.6}\right)^{0.6}$$

where $S_i(f)$ is the PSD of the i^{th} interfering system, n_i the number of i^{th} type interfering transmission systems, $H(f)$ the channel frequency response and I the number of different interfering systems. The frequency is in MHz.

The crosstalk attenuation X can be approximated with a Gaussian random variable, but constant values have been put in the model, meaning that the 90% of X realisations fall under the chosen value.

Figures 4(a), 4(b) and 4(c) show the PSDs of ISDN, HDSL, and ADSL upstream (taken into account in NEXT) and ADSL downstream (taken into account in NEXT) and ADSL downstream (taken into account in FEXT).

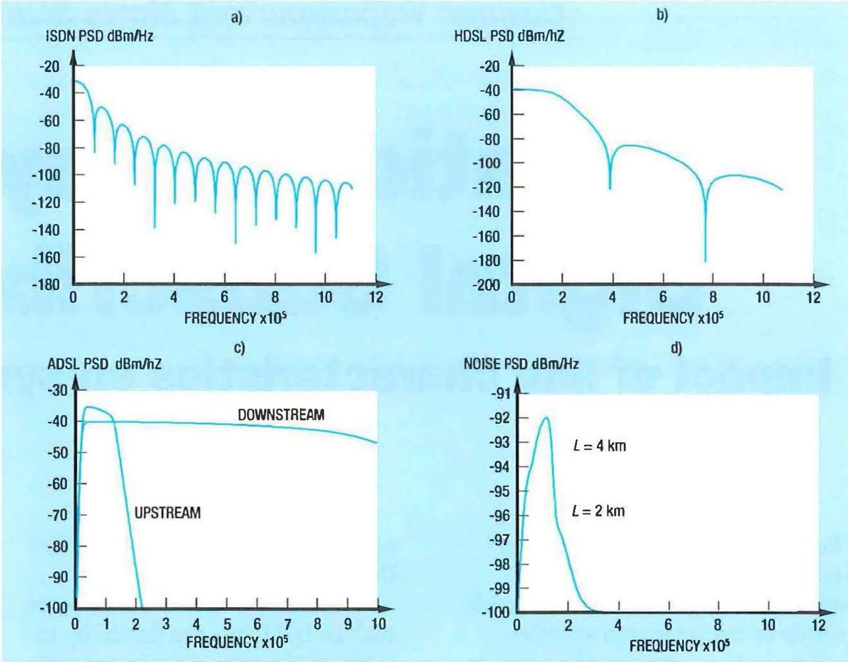


Figure 4 – Power spectral densities of: (a) ISDN BRA; (b) HDSL; (c) ADSL upstream and downstream; (d) NEXT+FEXT+ additive white Gaussian noise (AWGN)

Simulation Results

The crosstalk configuration for both urban and rural cases has been chosen for 30 ISDN, 5 HDSL and 30 ADSL and an additive white Gaussian noise floor equal to −100 dBm/Hz has been considered.

Table 1 shows the downstream gross and net achievable data rates with an ADSL system with echo cancellation (lower band edge of 26 kHz). The net data rate is obtained removing the overhead channel bit rate, the cyclic prefix overhead and the FEC redundancy.

Figures 4(d), 5 and 6 show, respectively, the total noise PSD (meaning NEXT+FEXT+AWGN), the measured SNR at the receiver end, and the bit allocation in frequency for both urban and rural cases.

Figure 4(*d*) shows the total noise contribution for the two scenarios. It can be seen that there are no differences between the urban and rural cases because the FEXT noise contribution is negligible at the ADSL frequencies. Figure 5 shows the received SNR; in the rural case SNR is lower because of the higher cable length; the bit allocation directly reflects the SNR behaviour (Figure 6).

The depicted case shows that in urban areas more than 90% of copper pairs allow the installation of an ADSL system with a net bit rate reaching more than 8 Mbit/s (Table 1).

In rural areas, because of the longer cable length, a lower bit rate can be reached, depending on interference and cable characteristics. With the interference configuration chosen in this study, more than 90% of copper pairs in rural areas permit, with an ADSL system, up to 4Mbit/s. Those results apply when considering only the copper pairs between the central office and the subscriber locations.

Figure 5—SNR at the receiver end side

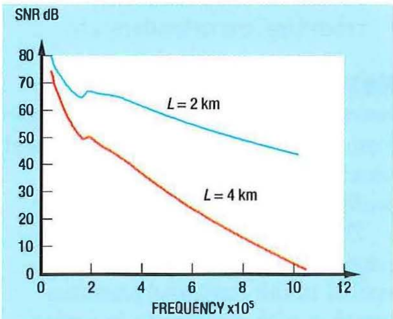


Figure 6—Bit allocation for urban and rural case

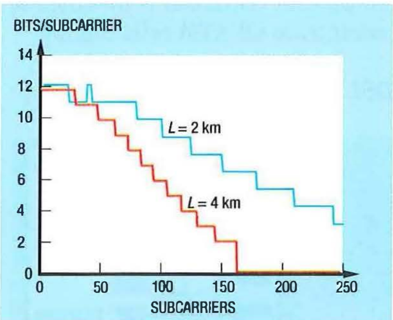


Table 1: Bit rates in the two cases

	Gross Bit Rate	Net Bit Rate
Urban case	> 8Mbit/s	~ 8 Mbit/s
Rural case	< 4·8 Mbit/s	< 4·1 Mbit/s

A typical loop is usually terminated by some additional wiring at the central office and subscriber locations. These additional wires can have lengths that can vary from a few metres to several hundreds, and they tend to have worse performances than the twisted pairs used in the loop plant. Moreover, some of the access network twisted pairs are bridged tapped, and the signals reflected from bridged taps cause an echo noise that, taken into account in the channel model, lowers the ADSL capacity. Between the NEXT and FEXT interference, HDB3 is incompatible with ADSL because it makes the system bit rate fall down; ISDN has a worse impact on ADSL upstream than downstream; not too many HDSL systems have been put into the ADSL cable because of their influence on system performances.

Conclusions

This paper presented the results of a study to determine how ADSL technology can meet the growing demand for multimedia services in two typical scenarios: urban and rural areas. It has been shown how cable characteristics and interference impact on ADSL performances, considering the twisted pairs used in the loop plant and typical interference.

Acknowledgements

The authors would like to thank Rosaria Persico of CSELT (Telecom Italia Group) for many useful interactions on many aspects of this work.

References

1 T1E1.4/98-007R4 Standards Project for Interfaces Relating to Carrier to Customer Connection of Asymmetrical Digital Subscriber Line (ADSL) Equipment, June 12, 1998.
2 Chow, J. S.; Tu, J. C.; and Cioffi, J. M. A Discrete Multitone Tranceiver System for HDSL Applications. *IEEE Journal on Selected Areas in Communications*, **9**(6), Aug. 1991.
3 Contessa, M. Sistemi per la trasmissione digitale ad alta velocità su rame. Laurea Thesis on Electronic Engineering, L'Aquila, Dec. 1998.
4 Cook, J.; Magnone, L.; and Persico, R. Working Text For Appendix On Crosstalk Combination Method. ANSI T1E1.4 meeting in Arlington, 20–23 April, 1999.
5 Chow, P. S.; Cioffi, J. M.; and Bingham, J. A. C. A Practical Discrete Multitone Transceiver Loading Algorithm for Data Transmission over Spectrally Shaped Channels. *IEEE Transactions on Communications*, 43(2/3/4) 1995.

Biographies

Gaetano Vespasiano
Scuola Superiore G. Reiss Romoli

Gaetano Vespasiano was born in Catignano, Italy, in 1959. He received his degree in Electrical Engineering from the University of L'Aquila in 1984. In 1986 he joined the Scuola Superiore 'G. Reiss Romoli' where he teaches on optical fibre communication systems and broadband access networks. He has been active in the field of optical communications for 12 years and has published many papers and two books. Current research programmes include: WDM transmission for an all optical network, high-speed data transmission using copper pairs, comparison of alternatives for the broadband access network.

Maria Stella Iacobucci
Scuola Superiore G. Reiss Romoli

Maria Stella Iacobucci was born in L'Aquila, Italy, in 1972. She received the Dr. Ing. Degree in Electrical Engineering from the University of L'Aquila, Italy, in 1995. After graduation she was engaged in research on communication theory and coding from the Infocom department of the University of Rome 'La Sapienza'. Since 1997 she has been with SSGRR (Scuola Superiore Guglielmo Reiss Romoli), the post graduate school in telecommunications of the Telecom Italia holding group in L'Aquila. Her current research interests include digital communication theory and coding, performance evaluations, spread spectrum systems.

Koen Berteloot

ADSL Roll-Out in Belgacom

A Technical and Operational Perspective on the Physical Layer

This paper describes technical and operational aspects of the roll-out of an asymmetric digital subscriber line (ADSL) network in Belgacom. Firstly, the typical archetecture of the ADSL end-to-end network is reviewed, followed by the typical backbone interfaces and their impact on the processes. Next, the use of microfilters at the customers premises, to ease and speed up the provisioning and avoid the need for an engineering visit, is discussed. Finally , a typical European issue is considered: the use of ADSL over ISDN by means of the shifted ADSL spectrum, taking into account the impact on the spectral management in copper cables.

Introduction

Belgacom started a trial of asymmetric digital subscriber line (ADSL) technology in January 1998 in some main Belgian cities including Brussels, Mechelen and Antwerp. 1000 pilot-users were connected via ADSLs to the asynchronous transfer mode (ATM) backbone. A fast-Internet package, called *Turboline* was offered using a downstream bandwidth of 500 kbit/s and an upstream bandwidth of 100 kbit/s.

During this trial, the technology proved to be stable and reliable for broadband services. The procedures have also proved to be suitable for a commercial roll-out. A survey of the pilot users showed 80% of them recommended the service.

The commercial phase started in April 1999, offering two principal products:

- *Turboline-service profile:* This is a residential customer-oriented package that targets heavy Internet surfers with its particularly fast Internet package. (The fast-Internet market in Belgium has several players like Telenet with its Pandora service using cable-modem technology.) Turboline-service offers no guarantees on the delivered bandwidth. It is based on a 'best-effort' situation; however, the delivered packages have a peak-bandwidth of 1 Mbit/s downstream and 100 kbit/s upstream. The price of the package depends on the maximum volume of downloaded data offered by the service, or it can be on a flat-fee basis plus a charge for each extra downloaded Mbyte.
- *Turboline-PRO service profile:* This package targets business-customers, and also offers a fast Internet service but with a guaranteed bandwidth. At the customer premises, an ADSL modem with router functionality connects multiple users. Within the profile of 1 Mbit/s downstream and 512 kbit/s upstream, there is a guaranteed bandwidth of 256 kbit/s downstream and 64 kbit/s upstream. As in the Turboline package there are various volume-based prices on offer or a flat-fee option. In this commercial phase different Internet service providers can be used for the ADSL connection.

The Turboline-service itself contains the following main features:

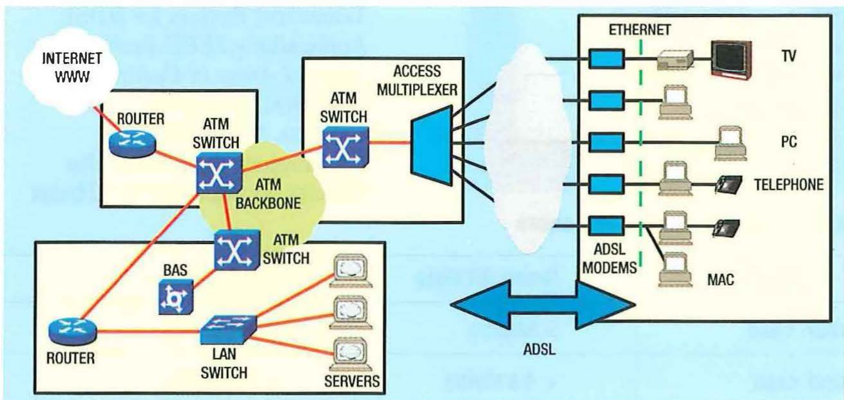
- home installation of ADSL modem and splitter, configuration of the fast Internet access as well as free customer support and repair;
- a basic service including: mail, newsgroups, web, chat, etc.; and
- an extended service including: downloading of software and data, audio- and video-on-demand, education, entertainment, videoclips, movietrailers, etc.

Network Topology

Figure 1 illustrates how the end-to-end network topology on ADSL is configured for the Turboline service.

The access side of the network consists of an ADSL modem installed in the customer premises together with a separate low-pass filter. On the local exchange side a digital subscriber line access multiplexer (DSLAM) is installed to concentrate all ATM cells coming

Figure 1 – End-to-end network topology on ADSL



Koen Berteloot:
Belgacom NTS TPM/TRE 7T29
Bd. E. Jacqmain 177
1040 Brussels, Belgium
Tel: +32 2 202 9697
E-mail: koen.berteloot@belgacom.be

from different clients and to map them on to a dedicated interface to the ATM backbone. This can be an STM-1 interface, an E3 interface or multiple E1 interfaces, depending on the expected load on the specific DSLAM.

The Belgacom ADSL network can be considered as a 'client' of the national ATM backbone. The ADSL using client is authenticated by a broadband access server (BAS). The client gets the IP address via this server to make the connection to, say, the Internet.

Additional servers are installed to offer local content; for example, on-line gaming, audio and video-on-demand, etc.

The DSLAM Interfaces to the Backbone and Their Impact on the Processes

Depending of the interface used, we can separate the typical architecture in to two parts:

Direct connection to the ATM backbone

In this case the interface of the DSLAM is directly connected to the ATM switch via an STM-1 or E3 interface (Figure 2). (Multiple E1 access is also possible.)

The physical link between the DSLAM and the ATM switch can be made using an optical connection for the STM-1 interface or an E3 leased line.

Subtended configuration of a DSLAM to a 'HUB' DSLAM

The second option is to connect a DSLAM to the network by 'subtending' this DSLAM with an interface to another DSLAM that is connected to the ATM backbone. The subtended DSLAM has no direct connection to the network (Figure 3). Typically this would be used with the multiple E1 interfaces.

In this case the connection between the two DSLAMS can be made by means of multiple (typically up to four) E1 leased lines. The HUB DSLAM connection to the ATM network usually uses an STM-1 (a synchronous transport module) or E3 leased line.

STM-1 interface

The STM-1 interface is an optical interface to the ATM backbone. When there is no direct connection to the SDH network by means of an add/drop multiplexer (ADM), the maximum acceptable optical attenuation imposes distance limitations.

So, when the decision is made to equip a DSLAM in a central office with this type of interface, the position of the central office to the ATM equipment has to be considered. When the distance is too long, an extra ATM concentrator can be placed in the neighbourhood (in the same central office) of the DSLAM. The STM-1 connection is placed between the DSLAM and the concentrator. Then, from the interface of the concentrator, we enter the ATM backbone. This interface need not be the same capacity as the STM-1.

The disadvantage is that an extra investment has to be made for the installation of the concentrator.

A big advantage of the STM-1 interface is the flexibility of the available bandwidth. Taking capacity (in terms of maximum installed clients per DSLAM) into account, there are almost no restrictions on selecting a certain bandwidth profile. Moreover, the 'simultaneous usage' and 'burstiness rate' of the IP traffic it possible to concentrate more clients into the available bandwidth.

The optical connections between the interfaces require skilled optical teams. This has also impacts on the repair process where, in case of optical faults, extra reminders have to be given to the responsible optical teams to check out connectivity.

When in-band monitoring of the DSLAM is carried out by centralised management, an optical fault can result in all communication with the DSLAM being lost. Troubleshooting requires an extra engineering visit(s) to the central office involved.

E3 (PDH 34 Mbit/s) Interface

The E3 interface is via a coaxial connection to the ATM backbone rather than an optical connection. There are typical options: in the first case, where the ATM backbone interface is situated in the same central office as the DSLAM, the connection is made using a coaxial internal connection in the building. Extra attention has to be paid to the quality of the cable used and the maximum distance between the two connection points, to prevent increased attenuation and possible loss of quality due to bit errors.

In the second case, where the ATM backbone access is not in the same central office, a 34 Mbit/s leased-line used between the two central offices.

The construction of a DSLAM connection is more complex than for a STM-1 interface because there has to be a synergy between the DSLAM installation team and the leased-line provisioning team to achieve an acceptable cycle time. End-to-end connectivity tests are mandatory to ensure the quality.

If the same client concentration is used on the DSLAM as for the STM-1 connection, extra attention has to be added to the used client profile to prevent an overload of the available

Figure 2—A direct connection to the ATM-backbone

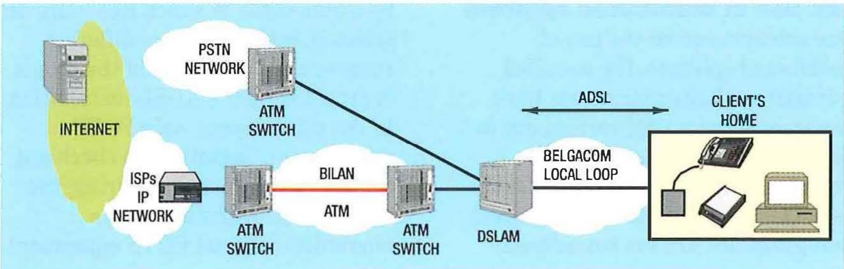
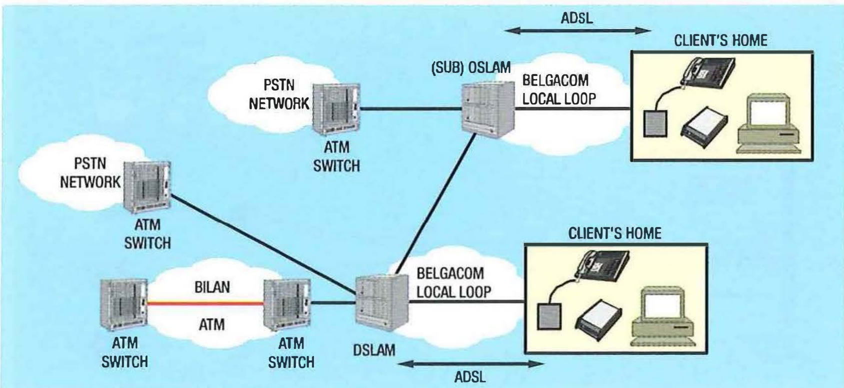


Figure 3—Connecting a DSLAM to the network by 'subtending'



bandwidth occurring. This can happen when, for example, ATM layer minimum bitrate guarantees are given to the clients.

With regard to the management of the DSLAM, especially for the repair process, it is important to recognise the difference between a fault at theDSLAM-interface level or at the leased-line level to prevent wasted time during troubleshooting. Hence, a good process flow is mandatory.

The N×E1 Interface (PDH Multiple 2Mbit/s)

This interface has two possible architectures. First we can have a direct (multiple) leased-line connection to the ATM backbone, similar to the E3 interface solution. Second, there is the 'subtended' architecture, as described above. In case of the subtended configuration, the leased-line connections are not made between the involved DSLAM and the ATM backbone interface, but between the so-called *subtended DSLAM* and the HUB DSLAM situated in an other central office.

An important difference with the above mentioned interfaces is the fact that multiple leased-line connections can be used between the two DSLAMS. This gives extra flexibility due to a degree of scalability of the available bandwidth.

On top of the E1 layer there can be an inverse multiplexer (IMA) running. This means that the DSLAM, in the case of multiple E1s, does not 'see' the separate E1 leased-line links but an overall bandwidth. Thus the ATM traffic is fitted on to the available bandwidth. The result is more economic use of the available bandwidth.

The use of the IMA facilities has also an impact on the repair process. In fact, the IMA group makes the interface connection dynamic. This means that the traffic is adjusted dynamically when the bandwidth changes; for example, when there is a leased line error.

When a different routing is used for the selected E1 leased lines for the DSLAM connection, a failure of one of the leased lines does mean that the DSLAM goes out of service, but the traffic is redistributed among the E1(s) still in service. Trouble shooting from a central management station is still possible via the in-band monitoring of the remaining E1(s).

Because a subtended DSLAM is connected to a HUB DSLAM, a fixed part of the interface bandwidth of the HUB DSLAM is used for the sub DSLAM. This is important when setting up the network architecture using the maximum possible subtended DSLAMS, connected to a hub, taking into account the interface bandwidth of the hub and the used client profiles.

ADSL Cabling at Customer Premises: Use of Distributed Splitters

Initial set-up

By installing ADSL on the client side, a low-pass filter is placed between the network entry and the customer premises equipment (CPE). This makes it possible to transmit simultaneously both the existing POTS services and the new ADSL services over the same copper pair. An additional in-built high-pass filter in the ADSL modem takes care of the fact that there is as little as possible interference between the two spectra.

The disadvantage of such an installation is that, in most cases, extra cabling has to be included in the house cabling, besides the normal configuration of the client PC. So the installation requires the visit of a telecommunications company installer. Thus, there is a search for alternative installation solutions to make the engineering visit more effective while saving time or even a situation where a visit is not necessay—there is a need for a self installation plug-and-play solution.

The use of distributed splitters

One solution can be the use of distributed splitters, the so called *microfilters*. Laboratory tests have been made and a trial carried out to check the technical and process impact on the installation. (Note: the use of G.Lite solutions or splitterless configurations are not considered.) We only considered the possible use of microfilters on full-rate ADSL.

The microfilter can be delivered in different configurations depending on the needs of the telecommunications company and the network situation. Here we consider the 5-pole solution and the in-line RJ11 solution.

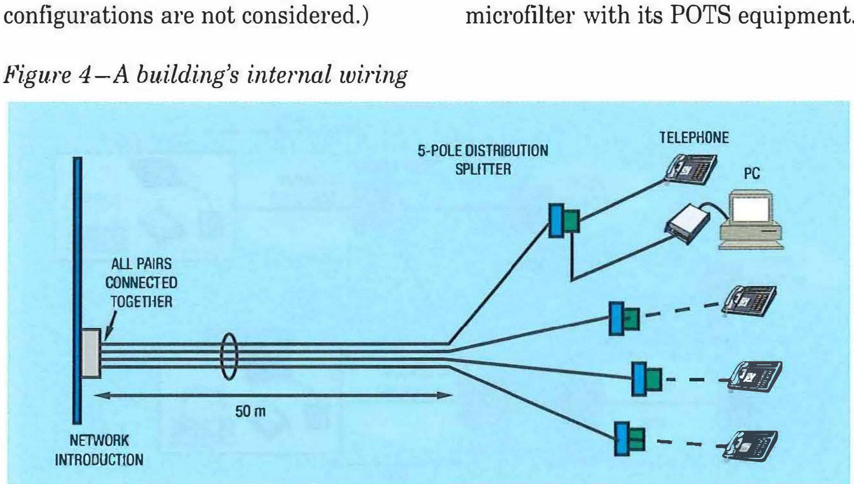
The 5-pole microfilter can be connected into the normal 5-pole wall socket between the wall socket and the connector of the CPE. The RJ11 connection of the microfilter is connected to the ADSL modem. The RJ11 in-line version has two RJ11-connection points (male and female) and is fitted in serial on the cable to the CPE. The house cabling determines which type of microfilter can be used. Both types have the same electrical functionality.

The installation set-up is different from the conventional splitter set-up: the microfilter is installed on *each* connection point in the home rather than at the network entry to the home. The number of microfilters used in an installation depends on the number of available wall connections in the house.

Also the in-house wiring becomes included in the system because the microfilters are connected to the wall sockets. This means that building's internal cabling has to be taken into account for the overall performance of the ADSL modem. Hence, bad or extensive cabling can influence the performance negatively. The laboratory set-up illustrates the impact on performance (Figure 4).

The internal wiring is simulated by a 50m cable of which five pairs are selected, each going to a different room in the house. One of the rooms is provided by the ADSL installation. In the other rooms, only POTS equipment is installed. To check out the impact of the house wiring, the test started by installing one microfilter with its POTS equipment.

Figure 4—A building's internal wiring



Then one items of POTS equipment was consecutively installed on each of the remaining wall sockets.

The results were that with the second microfilter, there was a performance loss of 9.3% compared with the initial maximum performance. With the third microfilter, there was an additional loss of 10.2%; the fourth microfilter caused 2.6% additional loss.

These results mean we can take into account the performance loss due to the impedance changes caused by the additional POTS equipment and microfilters on the line. Additionally, there is a bridged-tap effect when a wall socket is not used and no microfilter is installed in it.

Note that the performance loss is encountered on the downstream traffic, due to the higher frequencies used in this spectrum. The laboratory tests present the worst case situation since pairs wires are rarely together for a distance of 50m in the same sheath in a house.

Considering the technical impact, we can conclude that the internal wiring on the client side can influence the overall ADSL performance and reduce the maximum usable distance from the central office to the customer. The rate of that impact is unknown and depends on the quality and the cabling situation in the house.

A technical trial is necessary to check if there is internal wiring that does not cover the use of microfilters for full-rate ADSL. This trial would also indicate if it is feasible to create a self-install package with microfilters included in the package; for example, consider the number of helpdesk calls requiring installation assistance. In this trial it is also possible to determine the proportion of used 5-pole/in-line microfilters.

It is important for the self-install option to have a good user installation manual.

The use of microfilters on full-rate ADSL also impacts the provisioning process.

When an order is taken, it is strongly recommended that as much data as possible about the cabling at the customer premises is collected. Then a decision can be made as to the feasibility of installing microfilters without too much risk. One determining factor can be the number of wall sockets in the home. Special cases, such as alarm centrals or PBX installations, can result in the need to install a conventional splitter instead of a set of microfilters. In the case of corporate clients, who already have an in-house structured cabling for other high-speed services, it is essential to use conventional splitters.

Use of ADSL Over ISDN: Implementation in the Network and Technical Feasability

The European telecommunications market is promoting ISDN as a means of accessing the Internet. It is possible that this client segment will eventually consider an upgrade to ADSL for high-speed Internet access. It is unlikely that the client would downgrade from ISDN to POTS while installing ADSL.

So, for the European market, ADSL over ISDN is taken into consideration.

The fact that ISDN uses a larger frequency spectrum (2B1Q: 80 kHz / 4B3T: 120 kHz), has impact on the use of ADSL technology in the higher frequencies (Figure 5).

In the illustration mentioned above we can see that the used ADSL spectrum in the ISDN variant is shifted to the higher frequencies. This also means a reduction of the downstream bandwidth due to a reduction in the number of carriers.

Moreover, we have to take into consideration additional crosstalk, especially when a conventional ADSL over POTS line is collocated with ADSL over an ISDN line.

The most significant crosstalk comes from the part of the downstream spectrum of the ADSL over POTS on the upstream part of ADSL over ISDN. This makes the coexistence of both solutions in the same binder difficult.

Also the shifted spectrum makes both technologies incompatible. This requires additional hardware at both the central office side and at the customer premises.

A solution, offered by vendors is the universal splitter. This is an adapted filter that can be used to filter out POTS, ISDN 4B3T and ISDN 2B1Q. It makes eventual up or downgradeability possible without the need for an engineering visit or hardware changes.

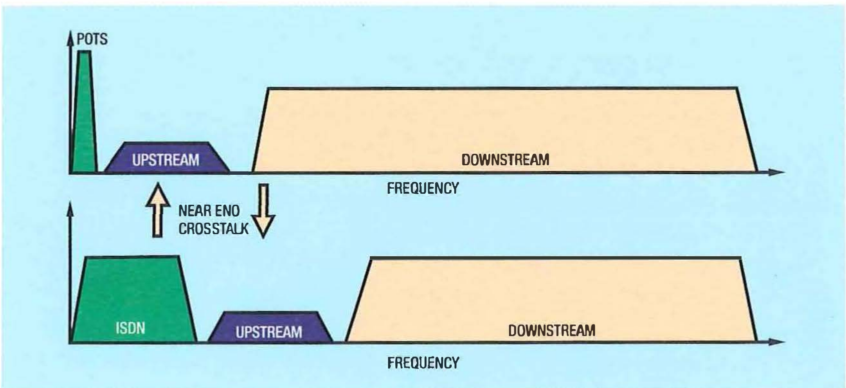
So two provisioning solutions are possible. The first is a universal solution where only ADSL over ISDN hardware with a universal splitter is used. This can be a solution for networks with a high penetration grade of ISDN. The universal splitter makes it possible to downgrade to ADSL over POTS without any hardware changes. The same hardware is used for clients who continue to use their POTS services without upgrading to ISDN. Then an eventual upgrade to ADSL over ISDN does not require any additional hardware at that time.

The fact that this solution only uses one type of equipment makes it more manageable. Also the additional near end crosstalk issue disappears.

A disadvantage is that even the POTS user has a decreased downstream performance due to the shifted spectrum, nor is there compatibility with G.Lite later.

The other solution can be the coexistence of ADSL over POTS with ADSL over ISDN in the same binder or cable. Therefore, pair-selection guidelines have to be adapted to make the additional NEXT acceptable. This can be a solution for networks where the ISDN penetration is rather low. A disadvantage is that new hardware is introduced at the central office and at the customer premises. Thus an ADSL-over-POTS client who wants to upgrade to ISDN, while keeping the ADSL service, has to change his/her hardware to make this possible. However, the ADSL-

Figure 5 – The frequency spectra for the POTS and ISDN cases



over-POTS clients can benefit from the full downstream capacity.

Note that, eventually, the NEXT can be reduced by disabling the interfering carriers in the profile of the ADSL-over-POTS client. At that time, we have to remember that the ADSL-over-POTS client gets the same downstream reduction of an ADSL-over-ISDN client.

It is obvious that the presence of ISDN in the network can cause additional provisioning difficulties when we take into account the deployment of ADSL on the same copper pair. One important point in the future deployment of both technologies is the flexibility and manageability of the whole technology, in particular the spectrum management in the copper cable network.

Biography

Koen Berteloot
Belgacom

Koen received his degree in Electronics at the V.H.T.I Kortrijk in 1992. He joined Belgacom in 1994 in the transmission department where he was responsible for the operational part of optical-line systems and the zonal optic-cable network in Brussels. In 1998 he was involved in the selection of ADSL equipment during a pre-commercial trial. Since then he has been responsible for the physical layer and transmission part as well as the set-up of the construction and repair-process of the ADSL project.

Milan Meninger

Synthesis of HFTP Multilayer Cell Access Network

This paper describes the multilayer cell architecture of hybrid-fibre twisted-pair (HFTP) access network offering optimised design for both narrowband and broadband services, simultaneously. The method described is suitable for network conceptual considerations and is demonstrated on five examples of networks offering services from the analogue plain old telephone service (POTS) up to broadband services supported by digital very-high-speed digital subscriber line (VDSL) technology up to 52 Mbit/s.

Introduction

The recent development era of telecommunications is characterised by great efforts to better utilise existing telephone symmetrical user lines, currently estimated at 700 million installed across the world.

High-frequency spectrum components have greater attenuation and are more exposed to cross-talk interference than lower frequency components. This phenomenon reduces the maximum length of access digital sections operated at higher bit rates.

The range of a broadband (BB) digital access section can be up to ten times shorter than a POTS access section. In other words, the range of BB user lines is approximately only one order of magnitude greater than are dimensions of urban elements (that is houses, roads etc.) compared with the POTS range that is two orders greater. Narrowband (NB) access equipment is usually located in the centre of gravity of its *user area* to minimise the cost of user lines. BB access digital sections are relatively very short and BB access equipment is likely to be distributed geographically homogeneously across the network, independent of the user density.

Investment and operational expenses of BB access network increase with the bit rate used. As not all the users will require the highest offered bit rate the access network will be decomposed into layers consisting of access cells of different dimensions for different bit rates. The mutual relationship between these cells must also be optimised to reach efficient economical parameters for the whole network.

Definition of Multilayer Access Cell Network Architecture

Model of multilayer network is based on four following architectural elements.

Access Cells (AC) form the elementary architectural elements of access network (AN) and cover the area from which all users of the same service are connected to a common access equipment belonging to a specific technology class. Each AC is characterised by the location of its central access equipment, called the *cell centre* ($CC_j^{(i)}$) and the cell territory ($T_C^{(i)}$), where the upper index i denotes specific technology class and the bottom index j specifies the ordinal number of the AC.

Access Layer (AL) represents the architectural element one level higher than AC and consists of the association of ACs of the same class (Equation 1).

$$AL^{(i)} = \bigcup_{\forall j} AC_j^{(i)}$$

AL is characterised by the set of relevant cell centres ($CC_j^{(i)}$) and the association of relevant AC territories $\bigcup T_{Cj}^{(i)}$

As an AL is composed of ACs, so the access network is composed of NL access layers, equipped with NL significantly different technologies, and the feeder layer (FL), providing interconnections of CCs to the Network Centre (Equation 2).

$$AN = \bigcup_{i=1}^{NL} AL^{(i)} \bigcup FL$$

Network Centre (NC) represents the geographical association of service node interfaces (SNIs) and completes the description of the whole network model.

Reliable operation of the network requires all the geographical distances, denoted *GD{.}*, between users located at internal cell points and the corresponding CC to comply with the following relationship (Equation 3),

$$GD\left\{CC_j^{(i)}; r_j^{(i)}\right\} \leq \rho^{(i)}$$

where $\rho^{(i)}$ denotes the range of user lines equipped with i class technology.

The described network model is shown in Figure 1.

Optimisation of Multilayer Cell AN Architecture

Optimisation of AN from the *Life Cycle Cost* view can be modelled by the following relationship (Equation 4),

$$\int_{LC} \frac{B(t)}{M(t)+I_M(t)+E(t)+I_E(t)+O(t)}\,dt \ldots \text{max}$$

where *B(t)* represents time course of revenues from provided services and components in the denominator which denote costs of transmission media and their installation, costs of access equipment and their installation and network operation and maintenance costs, respectively.

Milan Meninger:
SPT Telecom, a.s.
E-mail: milan.meninger@spt.cz

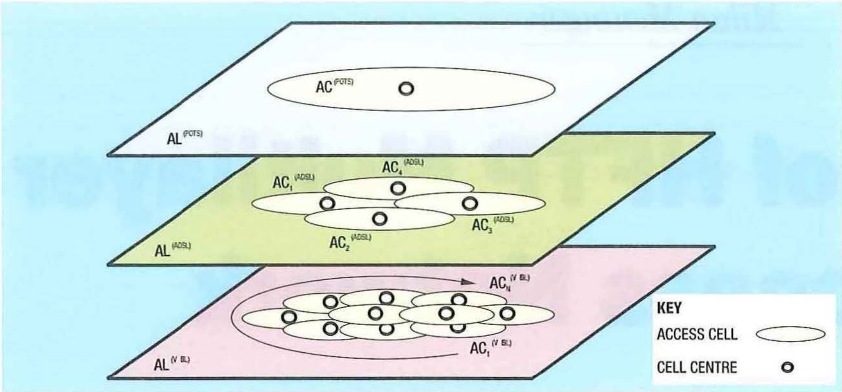


Figure 1 – Example of geographical decomposition of the access network into POTS, ADSL and VDSL access layers and the relevant access cells

Practical experience indicates that operational and maintenance expenses have small correlation with specific classes of transmission media and technology. Therefore, on this level of abstraction, the *O(t)* factor can be neglected in next considerations. Finally, the rest of the denominator of Equation 4 can be decomposed into two separate parts, the first dealing with equipment aspects and the second dealing with transmission media aspects of the network model.

Equipment aspects

General experience shows that the asymptotic cost of technically matured telecommunication hardware is mainly determined by its physical volume which results in the following simplified condition for minimisation of the network summarised equipment cost (Equation 5),

$$NEL\left\{\bigcup_{i=1}^{NL}\bigcup_{\forall j}\left\{CC_j^{(i)}\right\}\right\} \dots \min$$

where the function *NEL{}* denotes the number of entries of its set argument.

Physical realisation of Equation 5 can be interpreted as the association of CCs belonging to different layers but located in the same geographical location combined in one common multilayer CC. In technological terms, the association of CCs represents multiplexing of bearer channels into one aggregated stream.

Media aspects

The primary objective of HFTP access network architecture is an upgrade of existing copper POTS access network for future BB services. Therefore, the following considerations dealing with transmission media and their installation can be limited only on the optical FL.

The general rule says that the cost of a cable installation is higher than the cost of the cable itself, therefore the cost of cable infrastructure can be roughly approximated by the length of cable routes. The minimum cost of the FL infrastructure is given by the condition (Equation 6),

$$\sum_{T_{LV}}\bigcup_{i=1}^{NL}\bigcup_{\forall j} GD\left\{CC_j^{(i)}; NC\right\} \dots \min$$

There are two main phenomena influencing coverage efficiency of a multilayer cell AN territory with ACs.

Cell overlapping

As stated above, the theoretical operational area of an access equipment is the shape of a circle, but full network territory coverage by ACs results in the overlapping of the outer parts of adjacent ACs. The coverage efficiency can be expressed by the *cell overlapping factor* (Equation 7).

$$\eta_C^{(i)} = \frac{T_C^{(i)}}{\pi\rho^{(i)2}}$$

where $T_C^{(i)}$ denotes the net operational non-overlapping territory.

In spite of the best efficiency for hexagon cell shape, square cell shape was selected for further modelling due to its identical parameters in orthogonal directions.

Layer overlapping

The problem of ACs overlapping is even more critical in a multilayer AN where different ALs can differ in an order. This phenomenon can be quantified by the *layer overlapping factor* (Equation 8)

$$\eta_L^{(i)} = \frac{T_L^{(i)}}{\sum_{j \ni AL^{(i)}} T_{Cj}^{(i)}}$$

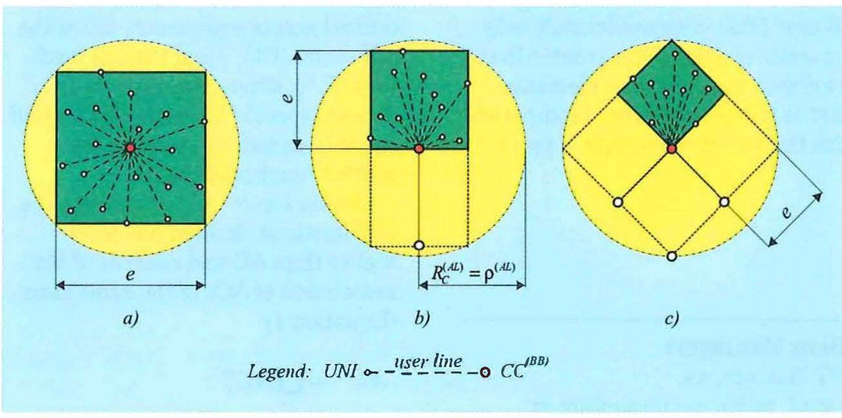
Overlay Broadband Infrastructure

Old cable networks were originally designed optimally for POTS applications with radial tree topology and later selection of places for CCs in order to enable injection of BB signals into POTS user lines will not be always simple. Figure 2 demonstrates how the position of CC in AC influences the total AC territory, despite the range staying identical in all cases. To increase probability for finding a realisable solution, the worst case of CC position in its AC will be used in the following considerations; that is, in a corner of the AC.

For the modelling of FL topology a segment decomposition approach is adopted. According to this method the territory around NC is decomposes into four 'V' segments in which the same topological paradigm is repeated, see Figure 3. (Note: The number of segments used is not unique, but the impact of the number of segments on AN infrastructure parameters is not analysed in more detail here.)

Quantitatively, the network cable infrastructure occupying *k* levels of ACs can be described by the total

Figure 2 – The impact of CC position AC dimensions



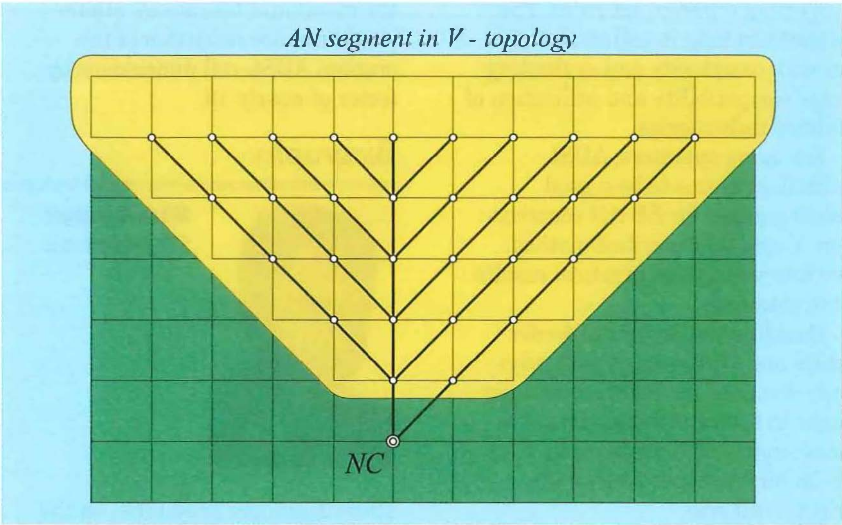


Figure 3 – Paradigm of overlay V-topology

length of cable trenches (LT_{Vk}) and by the total territory (T_{Vk}), defined by Equations 9a and 9b.

$$LT_{Vk} = 4(k-1)\left[(k-1)\sqrt{2}+1\right]e$$

$$T_{Vk} = 4k^2e^2$$

where e denotes the length of the AC edge.

Synthesis of Overlay Network

So far, the majority of AN matters have been solved only from the viewpoint of one isolated network layer. The design of the multilayer cell access network is solved in the following four steps:

1 *Arrangement of the set of involved technological classes* All NL involved access technologies are arranged in the descending order according to their ranges; that is, $\rho^{(NL)}$ corresponds to the shortest one.

2 *Calculation of relative ranges* All ranges are converted into relative form related to the shortest one (Equation 10),

$$\rho_R^{(i)} = \frac{\rho^{(i)}}{\rho^{(NL)}}$$

From this step it follows that the range is $\rho_R^{(NL)} = 1$.

3 *Calculation of transformed ranges* Relative ranges $\rho_R^{(i)}$ are transformed into integer $\rho_T^{(i)}$ values. This step supports collocation of some CCs of different ALs and this way also supports fulfilling requirements of Equation 5. The shortest range of applied technologies becomes the basic module AM of the FL web. To guarantee necessary transmission parameters of all ALs it is 'useful' to select all the transformed ranges $\rho_T^{(i)}$ shorter than their parent relative ones (Equation 11).

$$\rho_T^{(i)} \leq \rho_R^{(i)}$$

The term 'useful' is used for the reason that some exceptions of this rule are tolerable to the extent of the 'geographical range margin' (see Note under Table 1).

4 *Correction of Layer overlapping* So far, design of individual ALs was independent, and therefore individual total layer territories can exceed the basic $T_C^{(POTS)}$ territory by $MS^{(i)}$ AM modules (Equation 12)

$$MS^{(i)}AM = M^{(i)}\rho_T^{(i)} - M^{(NL)}\rho_T^{(NL)}$$

where $M^{(X)}$ denotes the number of applied $AC^{(X)}$ levels.

The $MS^{(i)}$ parameter determines a zone of $AC^{(i)}$'s which shall be shrinked toward the NC. The shrinking zone may be distributed between up to $MS^{(i)}$ cell levels to increase their signal to noise margin. As the last step, the Layer overlapping factor can be derived from equation 8. (Equation 13.)

$$\eta_{LV}^{(i)} = \frac{8\left(\dfrac{e^{(POTS)}}{\rho_G^{(i)}}\right)^2}{\left[\sqrt{NEL\{CC_V^{(i)}\}}+1\right]^2}$$

Table 1 Ranges of existing digital access technology classes on 0·4 mm copper pairs

AL reference code	Access Technology	Maximum bit rate downstream /upstream (Mbit/s)	Theoretical range (km)	Geographical range (km)
$AL^{(1)}$	BA ISDN/ POTS	0,16/0,16	$\rho_H^{(1)} = 4{\cdot}3$	$\rho_G^{(1)} = 3{\cdot}04$
$AL^{(2)}$	ADSL	6/0,6	$\rho_H^{(2)} = 2{\cdot}8$	$\rho_G^{(1)} = 2{\cdot}00$
$AL^{(3)}$	$VDSL^{(13Mb)}$	13/1,6	$\rho_H^{(3)} = 1{\cdot}5$	$\rho_G^{(1)} = 1{\cdot}06$
$AL^{(4)}$	$VDSL^{(26Mb)}$	26/3,2	$\rho_H^{(4)} = 1{\cdot}0$	$\rho_G^{(1)} = 0{\cdot}71$
$AL^{(5)}$	$VDSL^{(52Mb)}$	52/6,4 or 13/13	$\rho_H^{(5)} = 0{\cdot}3$	$\rho_G^{(1)} = 0{\cdot}21$

Note: In Table 1 are specified theoretical and geographical ranges. The latter one represents 0·707 shortening of user lines due to rectangular bends, a margin for multisystem interference and other degradation factors.

Application of the Method on existing xDSL technologies

Derived method will be demonstrated on existing xDSL technologies, parameters of which are summarised in Table 1.

Results of analyses of ANs supporting different maximum bit rates are summarised in Table 2, Table 3 and Figure 4. Brief figures gathered in Table 2 and Table 3, are completed by Figure 5 which converts the light grey columns in the tables into an idealised network plan of AN infrastructure supporting the maximum bit rate of 26 Mbit/s. The collocation of CCs of different network layers can be seen in Figure 5. Notice the merging of $CC^{(3)}$s and $CC^{(4)}$s.

Table 2 AC dimensions for different combinations of services

Access Layer	Cell Edge (km)			
$AL^{(1)}$	2·23	2·50	2·25	2·15
$AL^{(2)}$	1·34	1·50	1·50	1·07
$AL^{(3)}$	0·74	0·50	0·75	-
$AL^{(4)}$	0·45	0·50	-	-
$AL^{(5)}$	0·15	-	-	-

Table 3 Layer overlapping factor for different network layers

Access Layer	Layer overlapping factor (-)			
AL(1)	1·071	1·352	1·096	1·000
AL(2)	0·619	0·781	0·663	0·578
AL(3)	0·979	0·445	1·000	
AL(4)	0·786	1·000		
AL(5)	1·000			

Conclusion

The method described in this paper is not for precise network planning but rather for conceptual modelling of access network infrastructure determined for xDSL technology supporting different bit rates. The method can help in estimation of network complexity and evaluating range compatibility and utilisation of planned technologies.

For many operators, ADSL technology seems to be a good starting phase for BB AN construction. Using the described method, two interesting and practical results were obtained.

Doubling the density of feeder cables and ONUs in network originally designed for ADSL technology might in future increase the maximum supported bit rate up to 13–26 Mbit/s, depending on the original cell size.

On the other hand, the modelling showed that reconstruction of AN for the maximum bit rate 52 Mbit/s would require reduction of the original ADSL cell dimensions by factor of nearly 10.

Figure 4 – Density of feeder cables and ONUs in broadband HFTP Access Network

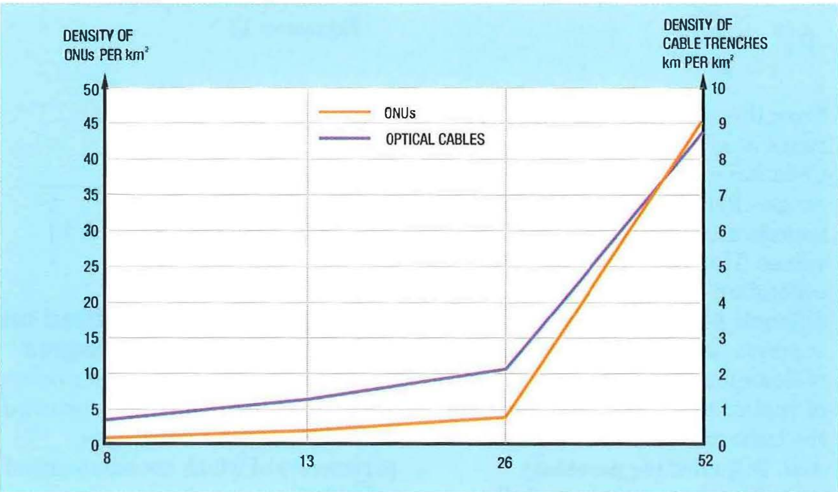
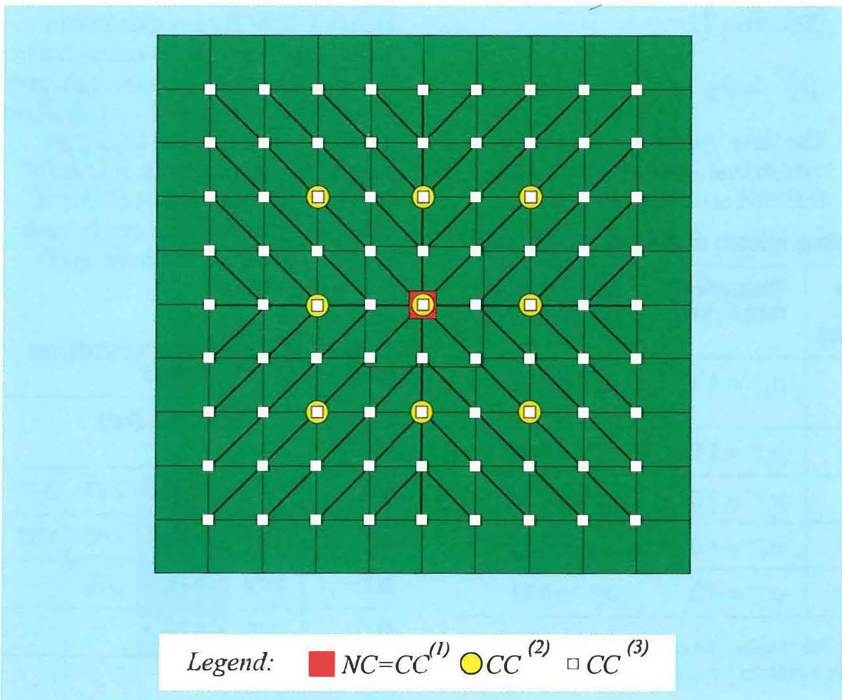


Figure 5 – Idealised plan of four layer AN



Biography

Milan Meninger
SPT Telecom, a.s.

Milan Meninger graduated on the Czech Technical University in Prague in 1968 and at the same university obtained his scientific degree C.Sc. (similar to a Ph.D.) in 1979. From 1968 to 1979 he worked in the research and development laboratory of Tesla, the Czech telecommunication manufacturer, where he dealt with design of digital lines. Between 1979 and 1995, he was working at the state PTT research and development centre. Since 1995, he has worked for SPT Telecom, a.s., the major Czech telecommunication operator, where he is responsible for specification and selection of access technology. He is also active in STC TM6 of ETSI. Milan is also the author of 14 patents and many technical papers.

M. Ali Salman and Terry Hodgkinson

The Future of IP–PSTN Interworking

As the Internet continues to grow at a phenomenal rate, there is an emerging desire for it to be able to support voice communications, sometimes referred to as* voice over IP *(VoIP). Although the Internet is some way from being able to support carrier-class voice quality, interworking of PSTN and VoIP is currently receiving much attention. The approach being adopted at the moment is the use of a media gateway box, which enables carrier PSTN standards to be interworked with VoIP protocols. In the longer term, however, and depending upon the actual traffic mix and volumes, such an approach may lead to an expensive and unscalable global telephony service. The particular issue addressed by this paper is the need to understand the technical and economic pros and cons associated with interworking different telephony service standards rather than developing a new single global standard.

The aspects this paper concentrates on include: the current approach of VoIP provisioning based on gateway/gatekeeper interworking and the limitations presented to the network in terms of cross-network signalling. It will compare the cost and complexity to other evolving networks such as public and corporate PSTN and mobile networks. The paper will examine how the current growth trends, which are being fuelled by data traffic, may influence future network architectures and protocols and hence the VoIP implementation options.

Introduction

Traditionally, voice services, or *telephony services* as they are more commonly known, have been carried over circuit-oriented networks. These networks have been evolved and optimised on the assumption that telephony would essentially continue to remain the predominant service carried. In contrast, the Internet, which is based on connectionless networking principles and uses the TCP/IP (transmission control protocol/Internet protocol) protocol suite, has been evolved on the assumption that it would always carry a very diverse range of services. It was inevitable, therefore, that the Internet would eventually aspire to provide voice services, and it is no surprise that this has now happened.

This capability to support voice services is commonly known as *voice over IP* (VoIP). VoIP is currently in its infancy and its real value will only become apparent once the new wave of Internet voice-related services have matured. It is predicted that as VoIP standards become more established voice-related services will become increasingly prolific. The accuracy of this prediction will have to be judged by the passage of time. But assuming it to be true, it is likely that there will be a virtuous circle of improving the network to support new voice services which will inspire newer services that will stimulate further network improvements, and so on. The final outcome of this will eventually be a network that is optimised for supporting a diverse range of applications of which VoIP will be a sub-set. This is the complete opposite to how the traditional telephony network has evolved.

However, for the foreseeable future, the dominant issue will almost certainly continue to be interworking VoIP with traditional telephony services. It is the aim of this paper to indicate why this will be so and to highlight the interworking issues that will be encountered. It will then look at ways for improving/eliminating the need for interworking, and use these observations to identify how to seamlessly remove all need for interworking in the 'ideal' future network.

What is changing?

Traditional telephony networks are typically described by the following four key points:

- based on homogeneous technology end-to-end;
- homogeneous service; that is, telephony;
- other services spoofed to look like telephony; that is, use of modem for carrying data; and
- well understood traffic characteristics and usage patterns.

However, with the explosive growth in the Internet, user expectations and demands will change. For the future the typical network is more likely to be described by the following three key points:

- based on heterogeneous mix of technologies end-to-end;
- heterogeneous mix of applications/services; and
- unpredictable mix of different types of traffic statistics and usage patterns.

The increasing heterogeneity and unpredictability will demand that networks should be made as independent of services and applications as is practicable. The current plans for application programmable interfaces (APIs), capable of opening up the

Dr. M. Ali Salman:
British Telecommunications plc
Room 136, B29,
BT Adastral Park, Martlesham Heath,
Ipswich, IP5 3RE
UK
Tel: + 44 1473 642509
Fax: + 44 1473 643906
Email: ali.salman@bt.com
Dr Terry Hodgkinson:
British Telecommunications plc

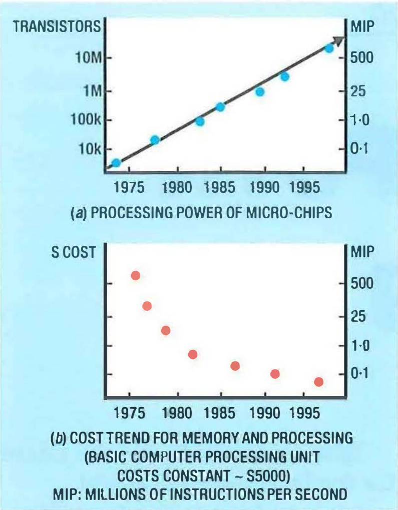


Figure 1—Moore's Law (source: http://www.intel.com/intel/museum/25anniv/hof/moore.htm)

network to various applications, may help, but the key feature will be to migrate all application/service-related features into the host machines. Traditionally, the cost and the extremely limited processing power of the host have prohibited this. However, this is unlikely to be the case in the future as memory and processing power costs continue to fall. This is especially true for processing power despite the fact that the processing power continues to increase according to Moore's Law, see Figure 1 (data from Intel web site). The network simplifications offered by migrating all application/service specific features into host machines will be considered in more detail later.

Typical Network Scenarios for VoIP

The incumbent public switched telephone network (PSTN) will have to be able to interwork with new VoIP data networks if it is to offer its users global connectivity. In other words, if someone using a telephone wants to speak to someone using a PC with voice-service capabilities, global connectivity cannot be claimed if the PSTN prohibits such communication. To offer global connectivity, special interface equipment is needed, namely *signalling gateways*, *media gateways* and *media controllers*: these are discussed in more detail later. These gateways effectively provide the necessary features needed to support PSTN and VoIP interworking. A further reason for interworking is the emergence of different mid-band and broadband network services, carrying mainly data, requiring access through both PSTN and the Internet.

Figure 2 shows the main two scenarios of interworking in VoIP:

- core network, and
- standard PSTN access interworking with the IP network.

Interworking the core networks has different requirements from interworking in the access networks. Depending on the size of the network, and how extensive the VoIP is going to be, the proliferation of VoIP interworking is likely to be in the access networks due to the fact that core networks only deal with aggregated traffic.

Interworking IP and standard PSTN equipment requires the installation of additional servers, gateways and control equipment. These gateways and the features they support are described below.

Gateway Features for Core Interworking

The interworking of IP and PSTN has made it necessary to increase the complexity of the signalling protocols in order to provide call and bandwidth control across the two networks. Figure 3 depicts the current signalling architecture for network elements interfacing PSTN and IP networks.

User access to the PSTN is carried by Signalling System No. 7's (SS7's) user-to-network signalling interfaces (UNI), through to the signalling point on the IP network side. ISUP information is carried between the two networks to provide features for network-to-network Interfacing (NNI).

As for deployment, the biggest market for VoIP in the short-term is expected to be in the intranet and virtual private network (VPN) domains where bandwidth provisioning is more easily managed and controlled. The issue of quality of service (QoS) for PSTN voice over the IP network will remain the major barrier to VoIP deployment in the public network.

Signalling gateway

This is the box that performs the mapping of SS7 signalling formats

Figure 2—High-level architecture for current VoIP

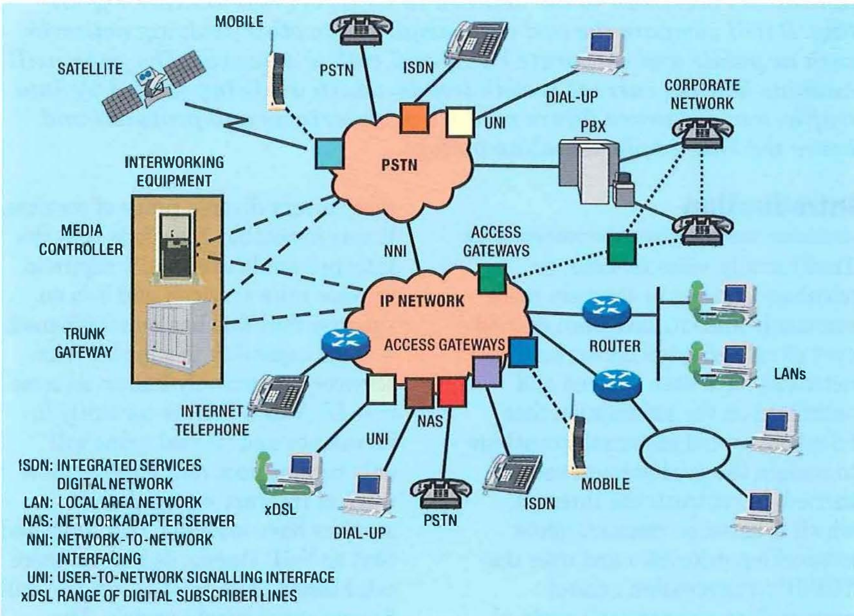
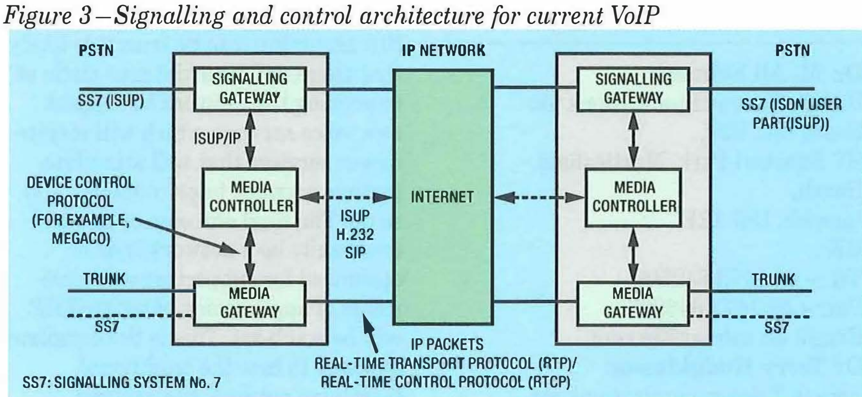


Figure 3—Signalling and control architecture for current VoIP



into IP compatible formats. SS7 is a global common-channel signalling standard defined by the International Telecommunications Union (ITU-T). It defines the protocols by which PSTN network elements exchange information over a digital signalling network for both fixed and cellular networks. The signalling network is used primarily for the following functions:

- call set-up and clear down;
- efficient use of voice circuits;
- intelligent number processing—local number portability, (0800) services etc.;
- intelligent services—call forwarding, call waiting, three-way calling etc.; and
- authentication and roaming services for mobile communications.

ISDN user part (ISUP) protocols are used to set-up, modify and release (among others) trunk circuits that carry voice and data between the calling parties. Signalling connection control part (SCCP) provides global title translation to associate an address (for example, dialled 0800 number) with a destination signalling point code and an application unique number. SS7 messages can be encapsulated (*tunnelled*) within IP and used by the signalling controller for routing within the network. The media gateway looks up the SS7 contents and establishes the required transcoding scheme to be employed.

The PSTN is still a cornerstone for the current VoIP architecture. The SS7 signalling used in the PSTN is also taken as the de facto signalling that new VoIP equipment will have to interwork with. Therefore, the claim that 'VoIP will cause the death of the PSTN' is still far from accurate. The PSTN, its core switches and many of its access technologies will still be used for the foreseeable future.

Media controllers (MCs)

Gateway controllers perform the majority of the call control and intelligence, to aid call progression within the network. For example, H.323 gatekeepers perform telephone-number-to-IP-address mapping and can act as platforms to run services for VPN applications. Media controllers implement one or more of a set of developing protocols: H.323, SIP, MGCP etc.

The main functions provided by these elements include *address translation*, *connection control* and *bandwidth management*. The use of these boxes has proven most useful in the corporate and leased lines and scenarios where bandwidth is more easily controlled.

Signalling between media controllers can be carried using ISUP with a special extension for IP, using the H.323 protocol with extension for PSTN or using session initiation protocol (SIP)[4] with extension for PSTN. This is an active area of discussion within the Internet Engineering Task Force (IETF), where the standards have not yet been fully agreed.

Protocols such as ITU's H.323 are commonly used in media controllers to provide mapping and various call set-up control functions between network servers. The IETF is proposing a new equivalent 'lighter weight' SIP to perform similar functions.

Media gateways

Gateways are end points in the network providing ports that connect to IP and circuit switched network users. Currently, each customer needs two ports, in order to be able to send and receive. Gateways can also carry out translation functions between video and audio codecs etc. Gateways are mainly used to establish links between PSTN terminals and IP or ISDN-based terminals (among others). Media gateways provide basic functions necessary for analogue access (UNI) and network-to-network (NNI) for core interworking. These functions include:

- *IP encapsulation/de-encapsulation* To traverse from PSTN to the IP network, ISUP is carried over IP between the signalling gateway and the media controller. This uses the ISUP information to issue control signals to the media gateway using protocols such as MGCP. Control functions may include connection creation, connection clear down, notification information etc. The media path that connects one gateway to another is used to transfer voice data from one end box to another using protocols such as RTP/RTCP protocols.
- *Encoding of analogue signals* This includes the digitisation of analogue signals and the packetisation of the resulting data.
- *Supervisory tones* The network user needs continued feedback on the progress in the call set-up stage. Dialling tones, busy tones, line-down tones etc. are generated according to the status of the call. Tone codes are provided by the switches which connect the various gateways.

In brief, the combination of the media gateway and controller provide the key features that have traditionally been placed in the network. In order to establish interworking between two different networks, encoding is essential due to inherent analogue access. Encapsulation is also needed due to interworking two different signalling standards (for example, SS7 and IP). Supervisory tones are generated in the gateway to compensate for the local exchange functionality. These features have been replicated in order to facilitate the communication between the various network systems. However, the ultimate price is paid in making the network service dependent. This will limit future service deployment and the flexibility in enriching current service features.

Network Futures

Basic network functions

Migrating certain network features into the host machine will enable the network to be service independent. This requires that there is minimal interworking and that supervisory tones are generated in the host machines. The key issues for PSTN and VoIP interworking are those of IP packetisation, common addressing, supervisory tones and signalling interworking. The impact of migrating these features into the host is designed to reduce the need for the gateway and its features, thus eliminating the need for interworking. This section will give an outline on how this could be achieved.

IP packetisation

Assuming that eventually all host machines will send a digital stream into the network rather than an analogue one. The need for IP encapsulation within the network will not be a network requirement. Moving these functions to the host side will enable their use by new applications, without much impact on the network functionality. This would also eliminate the encapsulation function from the media gateway.

Encoding

For the same reason as above, encoding can now be done by the host, although host-to-host communication may take place to determine what to use. This does not, however, involve the network. Signalling and messaging functions such as those carried out by H.323, SIP, TCAP etc. can now reside in the host, eliminating the need for them to be in the MC, or in other network nodes.

Service specific supervisory tone handling

These functions can be moved to the host side, enabling new applications to use a richer set of tones. These could now be audible, visual, text or any other form. Historically, tones were placed in the network nodes because that was the only way to use them. Modern server-based applications could generate the tones, as configured by the end user. Moving supervisory tones to the host eliminates another major function that media gateways provide. However, a 'unified' set of supervisory codes remain to be provided by the network, and may be carried by signalling messages.

Signalling interworking

Current developments in IP and other network technologies require a fresh look at providing a global 'lightweight' signalling system. Instead of 'interworking' technologies, there is now a clear need for an 'internetworking'[2] signalling solution that spans across all technologies. The new architecture will provide a network level intelligence that is built around 'general-purpose' servers that are independent from the network switching fabric. The future network can be based on heterogeneous technologies but still needs to be controlled by a 'global' homogeneous signalling system, sparing signalling mapping that occurs at every interface point in the network. Current network trends show that every new technology tried to be the universal solution for today's problems. This does not work, requiring us to think of ways to 'internetwork' different technologies rather than 'interwork' them.

It is, however, understood that it is more difficult to design a 'universal' signalling system, given existing network complexity. This remains essential if the network 'future proof', 'cost' and 'complexity' are to be kept under control.

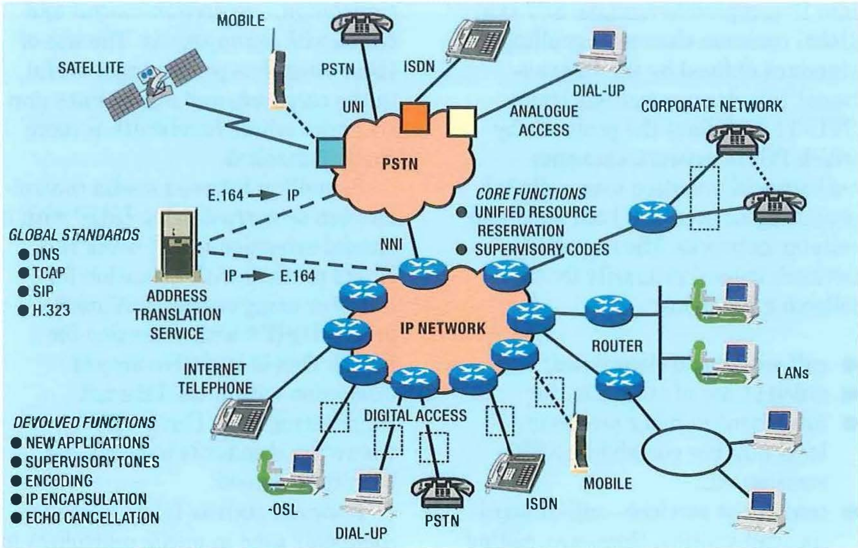


Figure 4—An example of universal 'multi-service' heterogeneous network

Basic signalling functions

Comparing IP protocol functionality with that of SS7 and that of asynchronous transfer mode (ATM) reveals that there are close similarities between the various stacks. The major differences are in the layer at which certain functions have been assigned. There is enough evidence that these signalling stacks may converge to provide one globally accepted standard. Salient features of such a standard include:

- *Messaging for call control and service association* A 'global' addressing scheme that is capable of locating network nodes and servers. Messaging systems such as SS7's SCCP, DNS and SIP all attempt to achieve similar functions. TCP-like functionality (error checking, re-transmission etc.) can be moved to this layer in the network. MTP2 in SS7 signalling will not need to do re-transmission as it is done at the TCP level. Applications will choose between the various messaging systems, resulting in the need to keep several systems at the user host machine. Different services can also create their own formats, but all will be encapsulated through one layer (the IP layer). IP does not make assumptions about the underlying technology.
- *Resource allocation* A network-wide reservation system will provide resource allocation functions. RSVP-like signalling may be used to provide reservation control for the generated traffic. Packets could also carry enough information to the network about their resource requirements on a hop-by-hop basis. Re-transmission of failing packets will have to be handled correctly, by higher layers.

Resulting network

It is quite plausible that by migrating functions such as re-transmission and error corrections to the TCP layer and functions such as address resolution protocols to the session layer we could provide service features that are not affected by the underlying network. IP can be used as the sole 'internetworking' protocol carrying all traffic (encapsulating or otherwise) to 'technology independent' lower layers. This 'functional separation' will allow any service feature to be carried by any underlying network technology using a 'single' universal signalling standard. Necessary information may be carried by packets to enrich the way this is finally implemented. Figure 4 depicts the way some of these changes may be achieved:

The future network (as depicted in Figure 4) has mainly devolved functionality between the network and the host. Functions such as IP encapsulation, encoding, tones etc. are now performed by the host eliminating the need for gateway devices such as gateways and MCs. Hosts can now connect directly to a router network. Duplication of 'messaging' functions at the user host may become inevitable, in the same way that computer applications now share a desktop PC. Interworking various network technologies will not be an issue any more.

Cost of Interworking PSTN and VoIP

Current arrangements for VoIP include the use of at least three networks (PSTN, IP and the SS7 signalling). This implies the use of at least three sets of standards, interworked to provide a voice service. The complexity in terms of network elements, numerous protocol stacks, signalling and stove-piping arrangements puts hurdles in the way of building a cost-effective network. When considering the broadband data network solutions, the number of interworking networks can mount to dozens. If not addressed, complexity may be the biggest factor in hindering the expected take-up of VoIP.

A simpler network signalling arrangement will result in savings in money and complexity. The internetworking proposals in this paper should help towards bringing in reasonable saving through eliminating intermediary devices and introducing a 'unified' signalling stack. This will encourage re-use of resources within the network, and simplify resource reservation across the network. The penalty that will be paid, is the duplication of software addressing and messaging schemes. We foresee that several of these applications (for example, SIP, H.323, DNS) will form a global standard and will co-exist in the host station.

The future network will serve all existing telecommunications applications and will be simple enough to provide seamlessly interconnected networks, majoring on IP and capitalising on existing network infrastructure. The move towards a multi-service network will yield added technical and economic benefits. The network operator should not worry about extra operation, management and investment every time a new network technology is introduced. This architecture introduces a single user box (for example, a router) to provide network access, including QoS, security, and reusability of resources.

Conclusion

Digitisation at the host into IP packets removes the need for gateways to encapsulate into IP packets. Use of a single global addressing scheme removes the need for layer 3 address translations. The telephone number can still be used but this would now have become a layer 4 address and effectively it is an alternative form of uniform resource locator (URL).

Having removed the need for IP encapsulation and address mappings, the only feature now provided by a gateway is signalling interworking. However, because layer 4 signalling messages can be evolved independently of the network—for example, DNS, SIP, TCAP—and other features of user parts not associated with resource reservation, they can be ignored from the point of view of network interworking issues. SS7 becomes recast in a use as 'fit-for-purpose' collection of protocols that are defined independently. For example, interworking DNS and SIP is not an issue for the network and they can co-exist completely independently. The basic reason for duplication is that this communication is purely server–server interaction, and does not require the network to have knowledge of the messages carried.

Because the fundamental principles of IP networking make no assumptions about underlying technologies, the technology-specific aspects of SS7 standards become redundant and can be discarded. All that remains is resource reservation and supervisory code information. By adopting a single, end-to-end globally accepted resource reservation protocol and supervisory messages, all need for signalling interworking is removed. Service specific tone generation would be located within host machines with the appropriate mapping between the supervisory message and the tone/visual display now being under the control of the host application. At this point the gateway becomes redundant and PSTN telephony and VoIP have converged to become the same.

References

1 COMER, DOUGLAS E. Internetworking With TCP/IP. Vol 1: Principles, protocols, and architecture. Prentice Hall, Inc., Upper Saddle River, New Jersey 07458, USA.

2 PETERSON, LARRY L.; and DAVIE, BRUCE S. Computer Networks, A Systems Approach. Morgan Kaufmann Publishers, Inc., San Francisco, California, USA.

3 HALABI, BASSAM. Internet Routing Architectures. New Riders Publishing, 201 West 103rd Street, Indianapolis, IN 46290, USA.

4 SCHULZRINNE, HENNING; and ROSENBERG, JONATHAN. Signalling for Internet Telephony. Department of Computer Science technical report CUCS-005-98.

Biographies

Dr. M. Ali Salman
British Telecommunications plc

Dr. M. Ali Salman gained his first degree in Computer and Communications from Essex University in 1984. He obtained a Ph.D. in adaptive systems from Loughborough University in 1988. He joined BT Laboratories in 1989 to work on traffic and SS7 management, utilising object-oriented knowledge-based systems. He joined the Intelligent Systems Unit in 1992 to develop network design solutions. In 1996, in the Network Technology Centre, he investigated issues of Internet protocols and contributed to the VoIP investigations. He recently moved to the Internet and Data Networks Centre focusing on IP solutions for future core networks.

Terry Hodgkinson
British Telecommunications plc

Terry Hodgkinson joined BT Laboratories in 1975, and since then has been involved in many aspects of optical communications and data networking research. From 1975–92 his research covered various aspects of optical transmission, particularly coherent optical detection methods, fibre polarisation effects and fibre amplifier performance modelling. Since 1992, his research interests have concentrated on data networking, and IP and ATM in particular, and how they might converge. Currently, he is leading a team researching the issues beyond IP/ATM networking as they are known today. The key aim is to identify new networking philosophies/technologies/architectures for achieving very efficient and flexible future networks. He has been awarded a D.Sc. Higher Doctorate and jointly awarded the ECOC '82 prize and the 1984 Rank Optoelectronics Prize. He has written 66 conference/journal papers and three book chapters and filed 10 patents in the fields of optical communications and data networking. He is a Member of the Institution of Electrical Engineers (IEE) and a Chartered Engineer.

David Blair

Offload of Dial-Up Internet Traffic from the PSTN/ISDN

Internet customer growth has risen rapidly in the past few years. A significant portion of users are connected to Internet service providers (ISPs) using the public switched telephone network (PSTN)/integrated services digital network (ISDN) for access. As Internet traffic is data traffic with longer average holding times than voice calls, this traffic may adversely affect other traffic in the network. Internet traffic itself may also experience congestion, and cause complaints to service providers.

In the long-term, switch suppliers plan to provide integrated offload technology as part of the core or remote subscriber unit. This will allow Internet traffic to be diverted directly into a data network at the earliest possible point before entering the PSTN.

In the short-term, a number of Internet traffic offload measures will be required by operators to prevent problems arising in their PSTN/ISDN networks. This paper presents the results of a study conducted by Telecom Eireann to determine the most suitable Internet traffic offload strategies to adopt and technologies available in order to address this short-term problem.

Introduction

Internet usage and Internet traffic is growing at a rapid rate due to a combination of many factors. As most customers access their Internet service providers over the PSTN, which was engineered for short holding time voice calls (three minute average) rather than the long holding time Internet data calls (15 minute average), this poses several short-term problems and challenges for network planners. The principal ones are:

- to provide effective quality of service for Internet users without degrading the quality of service of the fixed voice network;
- to optimise routing between the sources of Internet traffic and the Internet service providers' (ISPs) points-of-presence (POPs); and
- the need to phase in new Internet offload technology to help avoid future PSTN congestion.

Adding to the problem is the introduction of 'pay-as-you-go', low tariff, or absolutely free Internet services. These services drive forward the market for the Internet, leading to further growth in dial-up Internet traffic on the PSTN. Some users will migrate up the value chain to asymmetric digital subscriber line (ADSL)/very-high bit-rate digital subscriber line (VDSL), asynchronous transfer mode (ATM) and other broadband services which will modify growth slightly. However, the composition of forces is such that it continues to drive forward growth, and this will require technical answers in the short to medium term. Carriers have to invest more in their local networks to handle the strain of the increased dial-up Internet traffic.

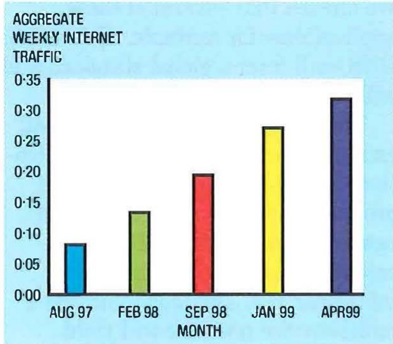
Internet Traffic Statistics

Internet traffic statistics reflect the latest growth trends. Internet traffic data for all ISPs in Ireland was recorded for the week ending 29 April 1999, as well as local PSTN traffic for the same week. Data was recorded in hourly intervals over the 24-hour period and included ISP, date and hour, number of calls and average call durations in the hour.

Growth in traffic

Figure 1 shows the growth in aggregate weekly Internet traffic hours from August 1997 to April 1999. This shows the underlying growth has increased four-fold in this 20-month period.

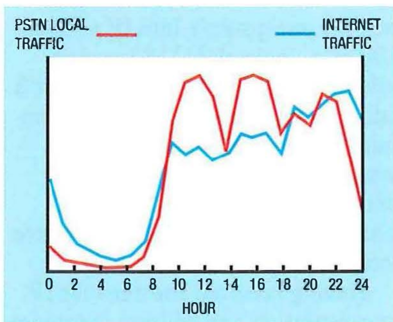
Figure 1 – Weekly Internet traffic 1997–1999



Internet versus PSTN—daily traffic profile

The daily traffic profile of all Internet compared to PSTN traffic for Friday 23 April 1999 is shown in Figure 2. The large morning and afternoon PSTN traffic peaks are mirrored by very slight morning and afternoon Internet peaks. Internet traffic displays the highest peaks in

Figure 2 – Internet versus PSTN traffic



David Blair:
Telecom Eireann,R332A, Ardilaun House, St. Stephens Green West, Dublin 2.
Tel: +353 1 7015344
Fax: +353 1 4780788
E-mail: dblair@telecom.ie

the evening from 2000–2100, and from 1800-1900 when tariffs are lower and PSTN traffic has tailed off.

Average holding times

A comparison of Internet and PSTN average holding times for a weekday (Fri 23/4/99) is shown in Figure 3. The average Internet call holding times during business hours is about 11 minutes, but rises sharply after 18.00 hours, and can average over 20 minutes during off-peak times. This trend shows that when Internet tariffs are low at off-peak times, Internet call holding times rise significantly, which increases pressure on local networks.

Trend in proportion of Internet to PSTN local traffic

The average proportion of Internet to local PSTN traffic is now 13·5%. The trend in growth is shown in Figure 4 (September 1998 was lower due to seasonal factors.)[1]

Statistics trends

These statistics show growing Internet traffic, a growing proportion of Internet to PSTN traffic in the local network, the likelihood of longer holding times as low cost or free Internet services become available. These statistics confirm a growing short-to-medium-term problem exists, unless significant Internet traffic offload measures are taken.

Figure 3—Average holding times

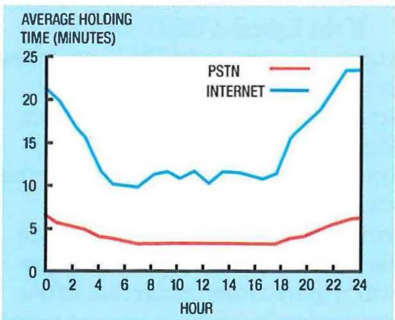
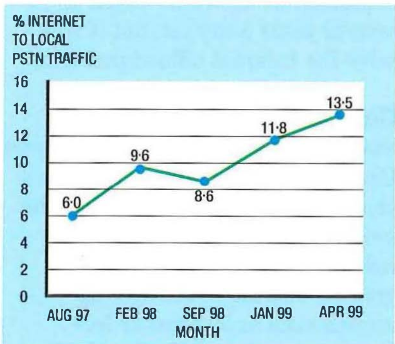


Figure 4—Trend in Internet traffic



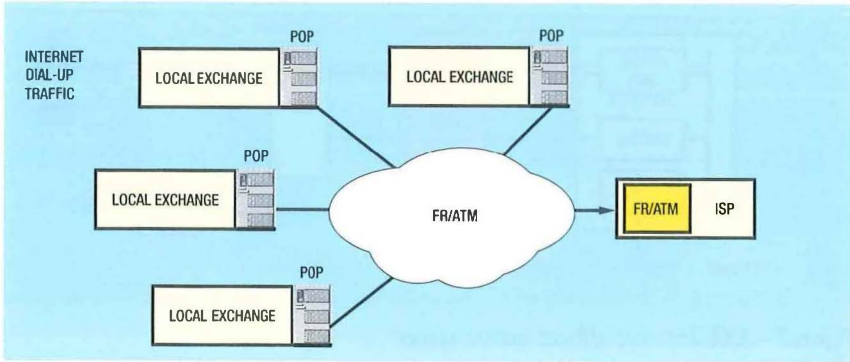


Figure 5—Decentralised POPs

Internet Traffic Offload Options

Options available to offload Internet traffic include:

- long-haul out-of-area primary rate accesses (PRAs) to ISP POPs;
- decentralising POPs by locating at tandem or local exchanges;
- Signalling System No. 7 (SS7) gateway solutions;
- switch supplier solutions; and
- xDSL and ATM.

Decentralising POPs

POPs could be located at local/ tandem switches showing the highest Internet traffic, or at all local switches, with access to the ISP over frame relay, ATM or other data networks as shown in Figure 5.

SS7 signalling gateway

An SS7 signalling gateway solution has advantages[2,3], and is favoured by ISPs. Internet traffic can be carried on a trunk instead of ISDN primary-rate interfaces, which are costly, while signalling and call management is done over the existing SS7 network. Cisco, Nortel[4], Ascend, and Stratus now offer SS7 signalling gateway products. Figure 6 shows the Cisco SS7 gateway based on an SC22XX signalling controller working together with NAS 5X00 access servers[5]. The signalling controller provides connection to the SS7 network and network access servers (NASs) while the NASs terminate the integrated services user part (ISUP) trunks. Network event monitors (NEMs) manage the signalling controllers SS7 functionality, and authorisation, authentication and accounting (AAA) and simple network management protocol (SNMP) servers provide security and network management. A backhaul router provides connection to the ISP backbone.

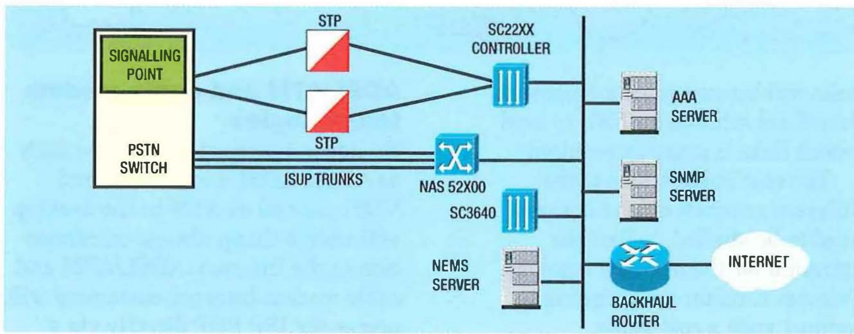
Switch supplier Internet offload solution

Ericsson now offer an Internet offload solution based on an ACC Tigris Internet access server[6] shown in Figure 7. This is a pre-switch solution where the access server is connected directly to the AXE group switch. IP traffic is offloaded to a data network before calls enter the PSTN. An access server for local/ transit switches, but not for remote switching units, is available.

Issues

POPs or Internet access servers can be located at local and tandem switches allowing diversion of IP traffic to a data network at this point. This limits the problem to the local switch. Remote subscriber units (RSUs) are still a problem, as offload technology for RSUs is not currently available. With the high percentage of IP traffic on many RSU links, these

Figure 6—Cisco SS7 Gateway solution



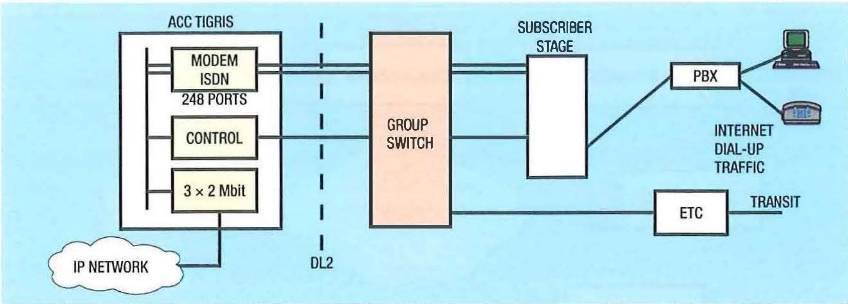


Figure7 – AXE Internet offload access server

Figure 8 – ATM multiservice network

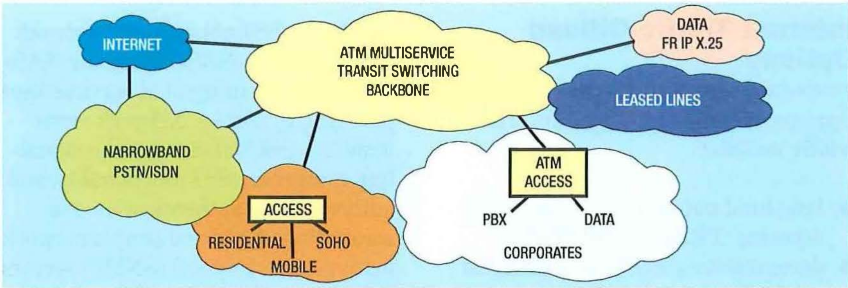


Figure 9 – IP only network: Internet access

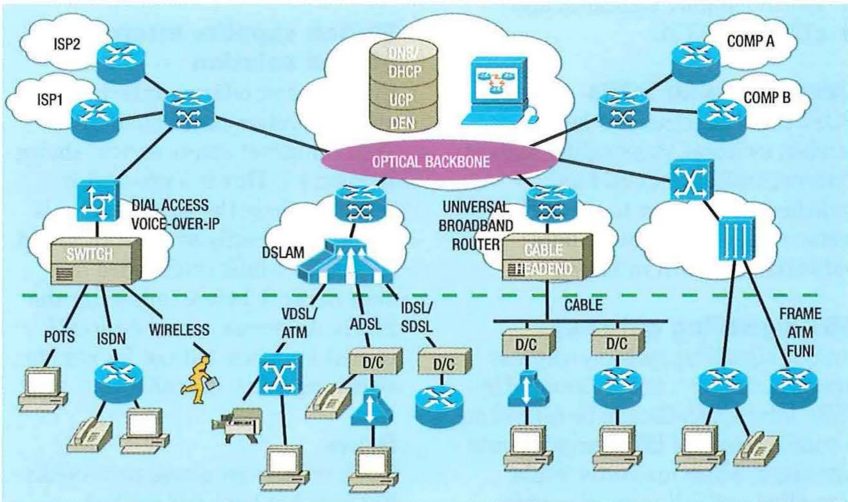
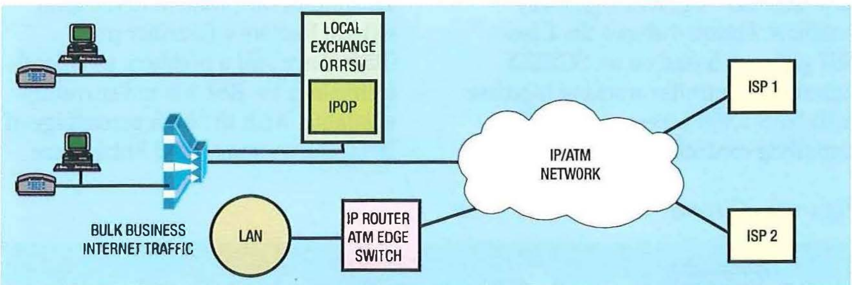


Figure 10 – Future Internet access



links will become quickly congested. An offload solution for RSU to local switch links is urgently required.

The cost implications of the different network offload options need to be studied, to find the optimum for the network topology. Telecom Eireann are planning to conduct such a cost study.

ADSL/ATM and cable modem technologies

Broadband access technologies such as ADSL, ADSL Lite, SDSL, and VDSL as well as ATM to the desktop will enable cheap always-on connection to the Internet. ADSL/ATM and cable modem Internet customers will access the ISP POP directly via a data network, and avoid the PSTN. Internet customer numbers using these technologies are likely to be limited in the short-term, and will not have any great impact on the dial-up Internet traffic problem. In the long-term however, the impact of these technologies is likely to be significant.

Future Vision

Network offload of Internet traffic is unlikely to be necessary in the future, as the networks envisaged will solve the problem. Strategists predict a universal routing/switching protocol supporting a vast range of services over a multiplicity of end-user terminals, with distributed intelligence across nodes, and an all-optical transport core. There are two main visions:

- *ATM multiservice network* for high quality end-to-end connectivity is ideal for corporate virtual private networks (Figure 8). This would use ATM edge and core switches, and provide IP switching over ATM.
- *IP-only network* (Figure 9) could be used to offer low grade, cheap connectivity, bulk dial-up traffic for cheap IP public services[7]. This would use IP edge routers and terabit routers directly on an optical transport layer. QoS would be provided by MPLS or Diffserv.

A hybrid ATM/IP network is likely to emerge, with ATM switches having a connected IP-terabit router.

If the hybrid ATM/IP network vision emerges, and ATM/IP switches provide the network core, Internet traffic would be carried as shown in Figure 10. Bulk business Internet traffic from corporate LANs would be connected by router or ATM edge switch into the ATM/IP network via the local exchange. Local exchanges will require either integrated POPs, data gateways, or upgrading to voice/ATM-IP nodes. Network-wide implementation of this vision is several years away yet, but it will solve the Internet offload problem.

Conclusion

Dial-up Internet traffic is causing a short-term network problem for the switched voice network. Statistics indicate rapid Internet traffic growth, and the forces driving growth such as low cost or free Internet service are fuelling higher

Internet traffic and longer holding times. Future solutions such as the hybrid ATM/IP network with combined voice and ATM/IP switches will solve the problem but are some years away, leaving us with a short-term problem. The general strategy is to offload IP traffic at the earliest possible point, and transfer it to a data network. The main offload technologies for the short-term include: remote PRAs, POPs or RASs at local exchanges connected by FR/ATM to ISPs, SS7 signalling gateway solutions, and supplier offload solutions. Telecom Eireann has started to locate POPs at the busiest Internet local exchanges, and connect remote PRAs to these. This Internet traffic is desirable in that it will generate future revenues. However carriers must be prepared to invest in their local networks which were designed for a different purpose, so they are able to handle this new surge in IP traffic. Detailed Internet traffic analysis, and careful cost modelling of the choice of Internet offload technologies will be needed to provide an optimum solution.

References

1 Blair, D. Impact of dial-up Internet access traffic on PSTN/ISDN dimensioning. European Network Planning Workshop (CNET), Les Arcs, March 1998.
2 Atai, Amir; and Gordon, James. Impacts of Internet traffic on LEC Networks and Switching Systems. Bellcore. New Jersey. USA.
3 Atai, Amir; and Gordon, James. Architectural Solutions to Internet Congestion based on SS7 and IN capabilities. Bellcore White Paper, Oct 1997.
4 Clay, K.; and Agnew, P. Internet and Network Access. INFORM '99, Nortel Networks, Las Vegas, Apr. 1999.
5 Cisco SS7 Gateway Solutions, http://www.cisco.com.
6 AXE Internet Access Server. Ericsson Datacomms/IP Product Information.
7 Shreve P. Building the New World. Cisco Systems Inc., 1998.

Glossary

ADSL Asymmetric digital subscriber line
ATM Asynchronous transfer mode
ETC Exchange terminal circuit
FR Frame relay
IAS Internet access server
ICR Internet call routing node
IFC Internet frame concentration
IPOP Integrated point of presence
ISDN Integrated services digital network
ISP Internet service provider
ISUP Internet services user part
NAS Network access server
NEM Network event monitor
PDU Protocol data unit
PMO Present mode of operation
PSTN Public switched telephone network
POP Point of presence
PRA Primary rate access (ISDN)
RAS Remote access server
RSU Remote subscriber unit
SDSL Synchronous digital subscriber line
SS7 Signalling System No. 7
STP Signalling transfer point
VDSL Very-high bit-rate digital subscriber line
xDSL the range of digital subscriber line technologies

Biography

David Blair
Telecom Eireann

David Blair is an executive engineer in the Network Strategy and Planning Department at Telecom Eireann. He graduated in 1979 with a B.Sc. (Hons) from University College Dublin, and received an M.Sc in Electronics and Microelectronics from Trinity College Dublin in 1981. After a period lecturing in the Dublin Institute Technology he joined Telecom Eireann in 1983. He worked in the Switching Department on switched network management, and SS7 signalling. From 1990 he has worked in the Strategy and Planning Department on switched network and SS7 signalling development, the introduction of ATM, and is currently working on switched aspects of Internet traffic routing.

Jan Verbeke and Hernick De Clercq

Architectures for Internet Traffic Off-Load from the PSTN/ISDN Network

This paper explains how telecommunications companies can cope with the increased load on public switched telephone network/integrated services digital network (PSTN/ISDN) resources due to the constant growth of Internet traffic. The key bottlenecks of the present mode of operation are identified and an overview is given of the possible post-switch Internet traffic off-load architectures for the redirection of Internet traffic from the PSTN/ISDN network towards a packet-switched data network. Finally, a migration strategy towards a given target post-switch Internet traffic off-load architecture is discussed.

Introduction

The constant growth of users with dial-up Internet access via the PSTN/ ISDN network continuously increases the load on the resources of the PSTN/ISDN network. This could ultimately give rise to congestion problems in this network that could affect other users.

Jan Verbeke:
Belgacom—NTS/TPM/DNW
Bd. E. Jacqmain 177,
B-1030 Brussels, Belgium
Tel: +32 2 202 85 42
Fax: +32 2 202 71 35
E-mail: jan.verbeke@belgacom.be
Hernick De Clercq:
Belgacom—NTS/NPE/INA
Bd. E. Jacqmain 177,
B-1030 Brussels, Belgium
Tel: +32 2 202 73 65
Fax: +32 2 202 73 10
E-mail:
hernick.de.clercq@belgacom.be

Solving these congestion problems would require the purchase and the deployment of additional equipment and/or the adaptation of modified engineering rules for both PSTN/ISDN switches and trunks. Off-loading Internet traffic from the PSTN/ISDN network towards a packet-switched data network may be an alternative and less expensive solution.

This paper explains how telecommunications companies can cope with the problems mentioned above. First we give a description of our current Internet protocol (IP) access network architecture. The key bottlenecks of the present mode of operation are identified. Next we give an overview of the possible post-switch Internet traffic off-load architectures, followed by an analysis of the key advantages and disadvantages for each architecture. Finally we discuss a migration strategy towards a given target Internet traffic off-load architecture.

Current Situation: IP Access via the PSTN/ ISDN Network

The following section gives a short description of the current IP access network architecture where points of presence (PoPs) of Internet service providers (ISPs) are connected to local exchanges with primary rate access (PRA) interfaces.

The route of an Internet call through the PSTN/ISDN network is shown in Figure 1. During Internet access a PSTN/ISDN connection is set-up between the modem of an Internet user and the remote access server (RAS) of the ISP. The Internet call is routed through the local PSTN/ISDN network, possibly via a local exchange-remote unit (LE-RU), towards ingress local exchange-basic unit (LE-BU).

From the ingress LE-BU, the Internet call originating exchange, Internet calls are routed via one (or more) toll exchange(s) towards an egress LE-BU, the Internet call terminating exchange. The egress LE-BU is connected to the RAS by means of PRA interfaces. The RAS is located in the PoP of the ISP. Alternatively, the RAS is connected (by means of a leased line) to the ISP data backbone network.

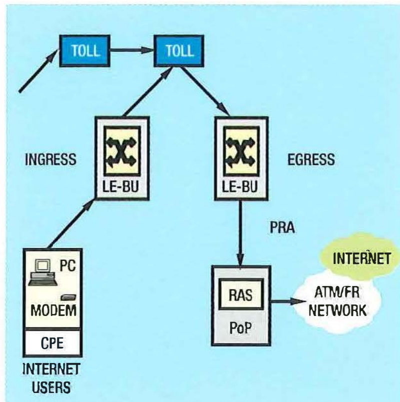
In the Belgacom case, the connection is to an asynchronous transfer mode (ATM) access switch of the ATM/ frame relay (ATM/FR) data backbone network. From there, IP packets are routed to the central router of the ISP towards the Internet.

Key Bottlenecks of the Current Situation

Due to the continuous increase of users with Internet access via analogue modems or via the ISDN, a constant growth of Internet traffic has been observed in the PSTN/ISDN network.

Internet traffic increases the load on the resources of the PSTN/ISDN

Figure 1—Internet access via PSTN/ ISDN network



network, because of the constant growth and particularly because Internet calls have a much longer call duration than voice calls for which the PSTN/ISDN network was designed. This could ultimately give rise to congestion problems in the PSTN/ISDN network that could affect other PSTN/ISDN users.

The problems occur in different areas:

- ingress exchanges (basic units (BU) + remote units (RU)) that connect subscribers;
- toll exchanges;
- egress exchanges that connect the ISP modem pools; and
- trunks between these exchanges.

Solving these congestion problems would require the purchase and the deployment of additional equipment and/or the adaptation of modified engineering rules for both PSTN/ ISDN switches and trunks.

Off-loading the Internet traffic from the PSTN/ISDN network into a packet switched data network may be an alternative and less expensive solution, because circuits set-up in the PSTN/ISDN network for Internet traffic are not actively used during the entire call.

Internet Traffic Off-load

In this paper the interception of Internet traffic to be redirected from the PSTN/ISDN network (which is circuit-switched and optimised for voice traffic) towards the ATM/FR network (which is packet-switched and optimised for data traffic) we will call *Internet off-load*.

This includes the use of existing features and capabilities of the PSTN/ ISDN network to obtain this goal, such as the use of intelligent network/ Signalling System No. 7 (IN/SSN7).

The following section provides an overview of the different Internet traffic off-load architectures considered.

Internet Traffic Off-load Architectures

Pre-switch off-load architectures

A first approach consists of intercepting the Internet traffic at the line side of the (ingress) switch and redirecting it out of the PSTN/ISDN network into a packet-switched network. These architectures are often called *pre-switch off-load* architectures. An example of such architecture is xDSL (the range of digital subscriber line technologies).

Post-switch off-load architectures

A second approach consists of the interception of Internet traffic after the (ingress) switch. These architectures are called post-switch off-load architectures. Internet traffic is allowed to pass through the (ingress) switch before being redirected out of the PSTN/ISDN network into the packet-switched network.

This paper focuses on post-switch off-load architectures.

Introduction of IN functionality and optimisation of Internet traffic routing

In this scenario, existing features and capabilities of the PSTN/ISDN network are used for optimising the routing of Internet traffic through the PSTN/ISDN network.

The introduction of IN functionality allows improved control of the routing of Internet traffic through the PSTN/ISDN network; that is, the telecommunications company can define to which egress exchange the Internet traffic is routed.

Internet traffic routing can be optimised following several approaches:

- *Definition of Internet areas:* In this case the PSTN/ISDN network is divided into a number of Internet areas. If one can arrange that the PoPs of the ISPs are located in an Internet area in order to reach the Internet users of that area, one is able to limit the number of interzonal links for Internet traffic.
- *Definition of egress exchanges:* If one can ensure that the PRAs of the RAS in the PoPs are connected to one (or more) well-defined egress exchange(s) within the Internet area under consideration, one can also limit the number of zonal and interzonal links for Internet traffic within a given Internet area.

This will ultimately lead to a significant reduction of the operational cost, because one has to manage a limited number of trunks for Internet traffic.

Internet off-load via the PSTN/ ISDN network

In this scenario, the number of PSTN/ISDN exchanges between the ingress exchange and the PoP is reduced, bypassing (intermediate) toll exchanges and/or the egress exchanges. Direct links can be provided between (important) ingress exchanges and egress exchanges.

Internet off-load via the ATM/FR network

In this scenario, the remote-access server is moved from the egress exchange to the ingress exchange, optimising the transmission network resources by the use of a packet-switched ATM/FR network.

Trunk interfaces between the PSTN/ISDN switch and the remote access server

So far the interfaces between the PSTN/ISDN switch and the remote access server have been PRAs. However, both switching- and data-network vendors have proposed an evolution towards trunk interfaces. In the latter case interworking between the RAS and the SSN7 network is required.

The SSN7 gateway functionality can reside in a standalone device but can also be integrated in the RAS.

Integration of a remote access server into the PSTN/ISDN switch

One switching network vendor has proposed an integrated/post-switch Internet traffic off-load architecture.

In a first step the vendors have proposed the functional integration of an external remote access server with PRA interfaces between the PSTN/ISDN switch and the RAS. In a second step the IP PoP is fully integrated in the PSTN/ISDN switch.

The key advantages and disadvantages of the post-switch Internet off-load architectures are summarised in Table 1.

Internet Traffic Off-Load Migration Scenarios

In the following section we will discuss a migration scenario towards a target Internet traffic off-load architecture. We will explain how Belgacom will cope with the problems mentioned above. We will show how Internet traffic off-load will be combined with the introduction of IN routing and a special number for Internet access to distinguish Internet traffic from other traffic.

The following network architecture evolutions have been identified:

Table 1: Advantages and Disadvantages of the Post-Switch Internet Off-load Architectures

Architecture	Advantages	Disadvantages
Introduction of IN functionality and optimisation of Internet traffic routing	Optimised routing of Internet traffic in PSTN/ISDN network.	Access numbers of ISPs have to be changed. PRAs of ISPs have to be connected to a limited number of well defined egress exchanges.
Internet off-load via the PSTN/ISDN network	Optimised utilisation of PSTN/ISDN network resources. Reduction of trunk interfaces.	PRAs of ISPs have to be connected to one (or more) (ingress) exchanges.
Internet off-load via the ATM/FR network	Optimised utilisation of transmission network resources. Packet switching.	ISPs have to move RAS towards ingress exchanges.
Trunk interfaces between the PSTN/ISDN switch and the remote access server	No need for PRA interfaces between PSTN/ISDN switch and RAS (that is, the possibility to connect RAS to (pure) toll exchange). Internet supplementary services due to interworking with IN/SSN7.	Increased complexity; the need for SSN7 gateway.
Integration of a remote access server into the PSTN/ISDN switch	No need for PRA interfaces between PSTN/ISDN switch and RAS. Internet supplementary services.	Less flexible than a standalone solution.

Phase 1: 078/17xxxx access;
Phase 2: off-load via PSTN/ISDN network; and
Phase 3: off-load via the ATM/FR network.

Phase 1: 078/17xxxx access

Belgacom will introduce a special Internet tariff linked with a special, non-geographic 078/17xxxx number. For this purpose Belgacom has divided the PSTN/ISDN network in a number of Internet areas. An ISP to which a 078/17xxxx number is allocated has to be present in an Internet area to reach the Internet users of that Internet area; that is, the PRAs of the RAS in the PoP have to be connected to one (or more) well-defined egress exchanges of the Internet area under consideration. Furthermore, IN functionality will be introduced to route Internet traffic. Internet traffic is separated from ordinary PSTN/ISDN traffic in so called *service switching point/advanced freephone service points* (SSP/AFS). The choice (number and localisation) of SSP/AFS points is optimised for the PSTN/ISDN network; for example, SSP/AFS functionality in toll exchanges and ingress LE-BUs with direct links to egress LE-BUs.

Internet traffic is routed towards the egress LE-BUs (the equivalent of an Internet call terminating exchange) as follows:

- Internet traffic from ingress LE-BUs with direct links to egress LE-BUs is routed via the direct link to the egress LE-BU towards the PoP (plus overflow via toll exchange); and
- Internet traffic from adjacent zones within an Internet area (or from ingress LE-BUs with no direct links to egress LE-BUs) is routed via SSP/AFS toll exchange and egress LE-BU towards the PoP.

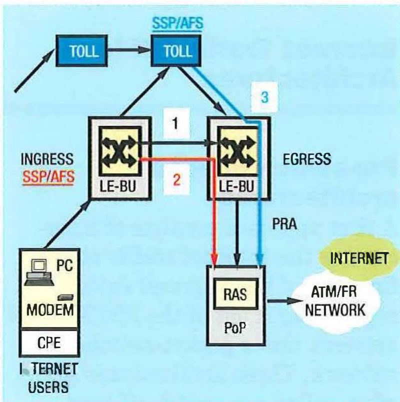
Phase 2: off-load via PSTN/ISDN network

As Internet traffic increases, congestion problems will move from the egress exchanges towards the ingress exchanges. In this case application of Internet traffic off-load via PSTN/ISDN is appropriate, as described in a previous section. This implies a reduction of the number of (intermediate) PSTN/ISDN exchanges in the route of Internet traffic from the Internet user towards the PoP of the ISP.

This can be done in several ways, as shown on Figure 2:

- If the Internet traffic from a given ingress LE-BU towards the PoPs of a number of ISPs connected to the same egress LE-BU is sufficient high, we can foresee a direct trunk between the ingress LE-BU and the egress LE-BU, bypassing the intermediate toll exchanges (see Figure 2, box 1).
- If the Internet traffic from a given ingress LE-BU towards the PoP of a given ISP is sufficient high, we can connect one or more PRAs of the PoP directly to the ingress LE-BU, bypassing the egress LE-BU for these PRAs. The PRAs should have a sufficient filling ratio (see Figure 2, box 2).
- If the Internet traffic from a number of ingress LE-BUs towards the PoP of a given ISP is sufficient high and if the Internet traffic is collected in a toll exchange, we can connect one or more PRAs of the PoP directly to the toll exchange, bypassing the egress LE-BUs for these PRAs (see Figure 2, box 3).

Figure 2 – Off-load via PSTN/ISDN network



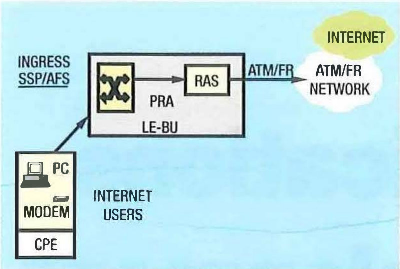


Figure 3 – Off-load via ATM / FR network

The solutions described above are more cost efficient because of the more efficient use of trunk interfaces on the PSTN/ISDN switches.

Note that because of the introduction of the 078/17xxxx number, the application of the Internet traffic off-load strategy stays transparent for the customers of the ISPs under consideration.

Key issues to determine whether an exchange (toll or local exchange) is suitable for Internet traffic off-load will be, among other things, the amount of Internet traffic to be off-loaded in the switch, the availability of SSP/AFS functionality in the switch, the availability of PRA interfaces in the switch and the impact of switching consolidation.

Phase 3: off-load via ATM/FR network

Suppose that a number of PRAs of the PoP of a given ISP are already connected to a given ingress LE-BU, so that the switching network resources can be considered to be optimised for these PRAs. Then the next step will consist of the optimisation of the transmission network resources. This can be done in the following way (see Figure 3).

The RAS (connected to the PRAs under consideration) is moved from the PoP of the ISP towards the ingress LE-BU. This supposes that the ISP is willing to leave the administration of the modem pool to the telecommunications company.

Additionally, modem pools located in ingress LE-BUs can be shared between different ISPs. The RAS is still connected to the LE-BU by means of PRA interfaces, but in this case the Internet traffic is off-loaded from the PSTN/ISDN network to the ATM/FR network at the ingress LE-BU.

The solution mentioned above is more cost efficient due to the more efficient use of transmission capacity (due to the use of statistical multiplexing) between the ingress LE-BUs and the PoP of the ISP which is now directly connected to the ATM/FR network.

The issue whether or not to integrate IP PoPs into PSTN/ISDN switches has to be seen as a final step towards a target Internet off-load architecture, only to be taken under consideration when the modem pools of the ISPs are already moved towards the ingress LE-BUs in areas where a large amount of Internet traffic has to be off-loaded and if the ISPs are ready to leave the administration of their modem pools to the telecommunications company.

Conclusion

Off-loading Internet traffic from the circuit-switched PSTN/ISDN network towards a packet-switched ATM/FR network may be an alternative, simplified, more flexible and less expensive solution than the present mode of operation. This includes the use of existing features and capabilities of the PSTN/ISDN network such as the use of IN.

In the discussion of the migration strategy towards a target Internet traffic off-load architecture, this paper has shown how Internet traffic off-load can be combined with the introduction a special number for Internet access to distinguish Internet calls from other calls.

As Internet traffic increases, the off-load point will move from the egress switches (Internet call terminating switches) towards the ingress switches (Internet call originating switches).

The key issue in determining whether an exchange is suitable for Internet traffic off-load will be the amount of Internet traffic to be off-loaded in the switch.

Glossary

AFS Advanced freephone service
FR Frame relay
ISDN Integrated services digital network
ISP Internet service provider
LE-BU Local exchange—basic unit
LE-RU Local exchange—remote unit
PoP Point of presence
PRA Primary rate access
PSTN Public switched telephone network
RAS Remote access server
SSP Service switching point
IN Intelligent network
SSN7 Signalling System No. 7

Bibliography

Atai, A.; and Gordon, J. Architectural solutions to Internet congestion based on SSN7 and Intelligent Network Capabilities. Bellcore.

Biographies

Jan Verbeke
Belgacom

Jan Verbeke graduated in 1990 from the University of Gent as a civil engineer in electronics. In 1995 he received his Ph.D. degree in electrical engineering with specialism in optical communications. He joined Belgacom in 1995 in the Research and Development department. In 1996, he joined the Backbone Network Strategy group. Since 1998 he has been working in the Data Networks group.

Hernick De Clercq
Belgacom

Hernick De Clercq graduated in 1990 from the University of Gent as a civil engineer in electronics. He joined Belgacom in 1992 at the transmission department. He was one of the key persons in the development of the SDH network architecture and the planning of the regional SDH network. Since 1998 he has been involved in PSTN/ISDN network architecture studies and does research on the Internet traffic evolution and impact.

Frank Wilson and Hill Stewart

Advanced Communications and Sustainable Development

A View from the ACTS Programme

Advanced communications technology and services (ACTS) provide the platform whereby the business, political, administrative and social communities communicate, organise and transact all of the important relationships and exchanges which form the socio-ecomonic life of our planet. The increasing concern with sustainability of physical, social and business resources as a basis for continued development at a global scale identifies a number of concerns which implicate advanced telecommunications technologies and services as a key determinant of sustainability 'best practice'. This paper attempts to summarise the current sustainable-development debate and to position recent developments in the ACTS programme of CEC DG13 as exemplars of positive responses to need—both in terms of demonstrations of current good practice and guidelines for future practice.

Frank Wilson:
Interaction Design Ltd, 13 Stonehills House, Welwyn Garden City, Herts AL8 6NH, UK.
Tel: +44 1707 338 751
Fax: +44 1707 373 901
E-mail: fwilson@i-d.co.uk
Hill Stewart:
Italtel Central Research Laboratories Castelletto di Settimo Milanese, 20019, Italy.
Tel: +39 02 4388 8030
Fax: +39 02 3351 0146
E-mail: hill@btinternet.com

Introduction

At the World Conference on Environment and Development in Rio 1992, sustainable development was defined as the principle that current generations should meet their needs without compromising the ability of future generations to meet their own needs. The aims of sustainable development are so clearly valid that there can be few who would not, in principle, subscribe to them. Achieving a level of socio-economic development which allows all to move forward, while those lagging in development are enabled to catch up, and at the same time avoiding destroying the ecological and material resource inheritance of generations as yet unborn, is a set of aims with which industry, business, government, science and non-profit organisations can agree, as indeed can most citizens[1]. This now widespread recognition that sustainable development is the responsible way forward is added to by the recognition that information and communication technologies (ICTs) are key drivers of socio-economic change and so have to be brought into focus as part of the equation. For example, Radermacher[2] points out that 'Modern information and communication technologies drive the worldwide economic system and the process of globalisation. In this process we see enormous growth worldwide with opportunities for overcoming poverty and promoting human rights but also with major threats to sustainability and to social justice. At the moment, our European citizens are witnessing that—in spite of economic growth and enormous wealth being generated by many companies and individuals—jobs flow away from us and unemployment is on the rise, social standards have to be given up and tax income from global companies is shrinking considerably'. Taken together, these two perspectives demand that any collaborative initiative proposing to develop and consolidate aspects of the future European framework must also consider how the need for sustainable development can be taken into account.

The CEC DG-XIII ACTS programme is the focus of the European Union's collaborative research effort in advanced communications. Its aim is to support the early deployment and effective use of advanced communications services. Experts from all over Europe are working together on around 200 projects covering a wide range of novel technologies, applications and services, and groups of ACTS projects work together to produce guidelines on strategically important themes. A set of these guidelines explores how information and communication technologies (ICT) can contribute to 'sustainable development', and specific projects have been developed in ACTS and elsewhere to refine and promote the results embodied in these guidelines.

Information and communications technologies clearly have considerable potential to contribute to sustainable development by, for example, helping to:

- monitor the global environment;
- promote 'dematerialisation' by achieving results with reduced material consumption;
- encourage 'immaterialisation' by providing added value which does not require material consumption (primary knowledge);
- rationalise industrial production and organisational processes and structures to increase effective-

ness, and to reduce redundancy, socio-economic imbalance, wasted effort, and environmental impact;
- substitute for travel through 'telework, teletrade and telecooperation' (Mitchell[3]) or to manage it more efficiently;
- match local production to global distribution;
- facilitate social and economic inclusion by offering opportunity to all regardless of language, age, physical ability, geographical location; and
- enable democratisation by increasing effective participation in political process requiring the aggregation of interests through representation mechanisms.

(Main themes adapted from Marsh[4])

The above themes are wide-ranging and cannot be covered in detail here. However, the following subsections provide key examples of ACTS trials promoting approaches to sustainable development, and selected ACTS guidelines addressing relevant themes. Together they are offered to support the reader in examining how advanced communications applications might support sustainability, and to expose further detail of selected themes for consideration.

ACTS Trials Promoting Sustainable Development

ACTS trials demonstrate effective use of new technologies but do not promote consumption of technology per se. Rather, they recognise that while technology is part of the solution, it is also part of the problem and we must beware of 'rebound' effects. For example, solid-state electronics vastly reduced the quantity of materials and energy consumed by a computer but it was so successful that it made it possible to put a computer on everybody's desk. On a global scale, today's computers consume far more materials and energy than their ancestors of the 1950s and 1960s. So, ACTS trials seek to show how new applications, or ways of working more effectively and efficiently, can be achieved by technological evolution for larger social benefit.

A good general example is the integrated broadband communications on broadcast networks project (IBCoBN). Here the participants developed solutions to allow existing (mature) cable circuits to be upgraded to full service networks so as to facilitate a new access network at minimal cost while preserving the investment and operational infrastructure of municipalities, cable companies and service organisations across Europe. As part of their trial they implemented applications such as remote care, whereby high-quality videotelephony was used as the access medium allowing 'on demand' attention and service between residential users (for example, the elderly) and service providers (for example, municipal authorities). Such technical developments along with their demonstrated applications show a number of interesting features for sustainable development. They show 'inclusion' of elderly people (and potentially many others), and also show rationalisation of service infrastructure whereby service staff save travel time to devote to service actions. This travel substitution also includes dematerialisation, and together these features show a wide range of benefits based on application of broadband to a remote service provision scenario. Further detail of this can be found in Descamps and Wilson[5] and Wilson and Descamps[6], as well as on the project web site at http://www.ibcobn.nssl.co.uk.

A powerful example of rationalisation (also including strong elements of dematerialisation, immaterialisation, and travel substitution) comes from the collaborative integrated communication for construction project (CICC). This project has developed a teamwork solution for distributed groups in the construction industry, and the approach has been trialed with success. The partners have produced advanced services including a 'people and information finder', and a three-dimensional project model. Furthermore, augmented reality and multimedia communications have been tested in a number of large distributed construction projects in Europe. This kind of project shows the application of virtual reality techniques to deliver distributed teamwork solutions where participants are allowed to feel part of a real enterprise despite distance. Architects have been able to really 'examine' the state of development of sites, and a number of companies have been able to present ideas, designs, and early 'build' work without travelling, yet the 'team' formation required for such a project has been fully supported.

Indeed, rationalisation of business process has been a major theme in ACTS trials where the new advanced communications technologies have bitten deep into the heart of large-scale commercial activities. The combination of rationalisation of process has commonly been supported by transport substitution, and dematerialisation and this is strongly evident in project trials shown by:

- PRINTIT—where a distributed publishing service toolkit has been tested by means of three pilot services: the distributed printing of a newspaper; the production of personalised catalogues; and a web kiosk printing service.

- TEAM—where on screen collaborative working for concurrent engineering has been demonstrated by a group of car manufacturers and their component suppliers. This has involved large-scale interworking of corporate users plus their SME supply chain. They have used a common set of tools developed to meet their specific needs and have produced results to support uptake of the approach generally.

- TECODIS—where a teleworking model for distributed software development has been established, modules have been implemented, and the system applied in trials and demonstrations where it was evaluated. The application has been shown in a real commercial environment and has used extensive European broadband connections.

- COVEN—where a general platform for collaborative virtual environments (CVEs), including computational service for teleworking and virtual presence has been demonstrated. The project provides the facilities needed to support future cooperative teleworking systems and to demonstrate the added value of networked virtual reality for both professional and home users.

While the above have mainly benefited the business community and addressed sustainable development issues relevant to them, an exception (along with the preceding IBCoBN project) is using mobile personal telecommunications innovation for the disabled in UMTS pervasive integration (UMPTI-DUMPTI) which aims to:

- develop proposals for meeting user requirements for applications and services;
- develop usability requirements, and non-conventional uses supporting social integration;
- specify common application programming interfaces to system services;
- develop applications and services using appropriate human-computer interfaces (HCI); and
- demonstrate the applications and services in a networked environment evolving towards a UMTS type network, using emerging technology and equipment, adapted where necessary.

This kind of approach to *inclusion* shows that even the most isolated communities (for example, the deaf) can be supported by a sustainable development approach.

The preceding examples show that ACTS trials include demonstration of dematerialisation, immaterialisation, rationalisation, travel substitution, social and economic inclusion. These have been demonstrated in realistic trials and show the way ahead, and other examples are also operational showing scope for environmental monitoring (through the many satellite projects), matching local production to global distribution (through extensions of the TEAM and CICC ideas), and democratisation (through initiatives such as EPRI-WATCH). Based on observation of this work, there are a number of sustainable development guidelines produced and in production, and the following section addresses these in detail.

ACTS Guidance for Sustainable Development

The ACTS programme, as part of its well-established concertation process, has supported a series of 'chains' whereby projects with shared interests work together to develop common guidelines. One such group of chains is concerned with generic access (GA) and includes a specialist sub-chain focusing sustainable development (GAD). The GAD chain is therefore a concertation mechanism for ACTS projects concerned with the potential impact of their work on sustainable development, considered in its economic, social and environmental dimensions. Work has evolved within the framework previously described in terms of dematerialisation, immaterialisation, rationalisation, inclusion and democratisation. In parallel, special attention is paid to the so called 'rebound effect', whereby sustainability gains in a specific dimension are offset by negative effects in the same or other dimensions.

The key objectives have been to define the key issues, to experiment with impact assessment approaches, and to study how sustainability can be developed as an opportunity and a business case. For this reason the target audience has been wide and has included service and product developers, industrial decision makers, research activities and standardisation bodies.

The guidelines have been formed within the framework described to include:

- Conceptual clarifications and transitional scenarios for sustainable lifestyles (Reflective Guidelines):

  —ICT, the Information Society and Sustainable Development (GAD A1);
  —Balancing between Information and Materials (GAD A2);
  —Sustainability as a Business Case (GAD A3); and
  —Socially Sustainable Best Practice of Workflow Re-engineering (GAD A4).

- Sustainability Issues in ICT Application Sectors (Technical Guidelines drawing directly on ACTS trials results):

  —Residential Broadband Access (GAD B1);
  —Impact Assessment of ICT Applications (GAD B2);
  —Impact of ICT for the Empowerment of people with Special Needs (GAD B3).

- Sustainability Impact Assessment (Developed in collaboration between horizontal actions and ICT trials):

  —Procedures for Assessing the Impact on Sustainable Development of Applications Using Teleservices (GAD C1); and
  —Assessing Sustainable ICT Application Scenarios (GAD C2).

From the above it can be seen that the GAD group has developed not only guidelines on specific sustainability issues, but also introduced techniques that developers can use to evaluate the medium-to-long-term impact of their applications. Cross-impact techniques have been developed to analyse the effects of possible technical solutions to the social and environmental problems of regions, nations or the world as a whole. These techniques are still at an early stage but they have already shown how there can be a positive environmental impact. In one example, the introduction of an intranet has made wine production economic in an area of Sicily where abandoned vineyards were creating a serious problem with soil erosion. In another example (concerning IBCoBN as described earlier) it has been shown how the region around Kortrijk in Flanders could gain significant general regeneration bounties by more widespread deployment of the full-service-network upgrade to mature cable television circuits (this work also supported the local analyses addressing Telnet Flanders).

The work of the GAD chain can be examined in detail at its web site at http://ww.epri.org/gad

Summary Comments and Future Developments

The ACTS programme has included a wide set of project trials, many of which directly address, or at least contribute to, the sustainable development debate. It is clear that sustainable development is a well-accepted part of the global socio-economic and socio-technical agendas. The work on the sustainable development guidelines in ACTS, based on the activities in the trials and a wider consideration, have helped to progress two distinct domains of discourse and have provided a useful fusion of themes emerging from them. The traditional 'ecology' focus has related sustainable development to the economic and social dimensions, while the 'information society' debate has examined the relationship of ICT to the same social and economic dimensions. The ACTS GAD work has unified these by exposing and developing a clear view of how advanced communications development has the potential to benefit and to unify the concerns present in social, economic and environmental spheres if dealt with appropriately, since, as Prof. Radermacher states 'Modern information and communication technologies drive the worldwide economic system and the process of globalisation'[2].

Stemming from, or at least benefiting from, the work of the GAD group in ACTS, a number of key actions for the future are emerging. First, the Alliance for a Sustainable Information Society (ASIS) is developing its charter and its working groups, including sustainable communities (AG1), dematerialisation of industrial and business processes (AG2), social inclusion (AG3), homes and workplaces of the future (AG4), and mobility and transport (AG5). At the same time the European Commission has launched its Fifth Framework Programme (see http:// www.echo.lu/fifth) which includes a specific programme for sustainable development, yet also retains the focus on sustainable development issues for ICT development in its Information Society Technologies programme (IST).

There is a clear need to maintain the momentum generated by the ACTS trials and its GAD working group to ensure that the links between the IST and the sustainable development agendas are not lost. If we accept that ICTs are really driving the re-organisation of worldwide social and economic systems, then we must not lose sight of the need for continued harmonisation between the traditional view of sustainable development and the accelerating progress of the IST framework for the information society.

Detailed ACTS trials information can be obtained from the ACTS INFOWIN[7], along with reports and detailed trials outputs offered for public access http://www.infowin.org. The briefings and guidance on using ACTS results from the ACTSLINE[8] are at http://www.actsline.org .The GAD chain guidelines[9] can be accessed via http://www.epri.org/gad.

References

1 ASIS Charter for a Sustainable Information Society (draft). CEC DG-XIII ACTS Programme, Project ASIS—'Alliance for a Sustainable Information Society', 1998. http://asis.jrc.es.

2 Radermacher, Professor F-J. cited in Second Annual Report of the Information Society Forum, Working Group 4, (1998).

3 Mitchell, H. The ACTS project 'European telework development', on-line support for new ways of working—'European Telework Online' at http://www.eto.org.uk, 1997.

4 Marsh, J. Sustainability in the Information Society. CEC DG-XIII ACTS Programme, Project AC079 EPRI-WATCH, report number 101, 1997.

5 Descamps, P.; and Wilson, F. Adapting CATV Broadcast Networks for Broadband Interactive Symmetrical Services. IBC '96, International Broadcast Convention 1996, Amsterdam. IEE Publications.

6 Wilson, F. Evaluation of City-TEN Trials. CEC DGXII TEN-IBC Programme, Project B3002 City-TEN, project deliverable D4, Dec. 1996.

7 ACTS INFOWIN—http:///.infowin.org

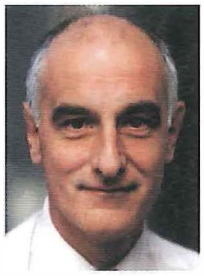
8 ACTSLINE dissemination actions —http://www.actsline.org

9 GAD Chain Guidelines—http://www.epri.org/gad

Biographies

Frank Wilson
Interaction Design Ltd

Frank Wilson is a human factors engineer and system designer. He worked with the Xerox Corporation as an interface designer for new products and then joined the London HCI Centre within the University of London as part of a UK initiative to promote technology transfer from research activities. He later joined the spin-off company Interaction Design Ltd which continues to promote research and innovation in human-technology integration, and which also supports EU collaborative projects where end-user issues and impacts on working life are important. He is currently chairman of the ACTS programme chain group on Generic Access (GA), and currently represents the Flexible Working and Sustainable Development focus in the ACTS project ACTSLINE.



Hill Stewart
Italtel

Dr Hill Stewart joined the UK Post Office (now BT plc) in 1972. Following research on optical communications technology and systems he moved to technology forecasting and strategic planning and managed BT research programmes on architecture and standards and network and service management. His final job with BT was producing the first release of an integrated architecture for the company's networks and systems. During his time with BT he was active in supporting the European Union activities in collaborative research and development and, since leaving BT, has worked as a consultant on European RTD. He is currently managing the ACTSLINE project for Italtel.

Robin Harris Foster and Stanny De Reyt

Re-inventing the Call Centre with Predictive and Adaptive Execution

Call centres have evolved from simple single-function centres to offer access, convenience, choice and courtesy to callers. Forecasting and staffing tools support planning, enterprise databases permit the business to craft specific caller treatments, and cross-trained agents using desktop applications can respond to a wider range of caller needs and business opportunities on a single call. One key element of the call centre, however, has changed only superficially—the question of 'What should each agent do next?'

The 'oldest waiting call' rule has answered that question for the last 20 years. Signs that this methodology is obsolete are seen in call centres where designs become more complex and results more difficult to achieve; where manual intervention moves agents from skill to skill chasing problems; where the most talented agents are overworked. This paper describes predictive and adaptive techniques† that answer the question 'What should an agent do next?' These techniques re-invent the call centre, creating a robust operation where performance is aligned with business intentions, without the manual, corrective intervention common in conventional centres.

What Should an Agent Do Next?

In any call centre, some routine must answer this question. First, what is the agent capable of doing? The agent's capabilities are referred to here as *skills*. Are the skills to be used in some hierarchical way, using some skills in preference over others? Are any of the calls waiting for the skills the agent holds queued at different priorities, where a call at a higher priority must be taken over a call at lower priority? Finally, how long has each call been waiting for service?

Robin Harris Foster:
Lucent Technologies, Room 1F-425
200 Laurel Avenue, Middletown,
NJ 07748, USA.
Tel: +1 732 957 5451
Email: robinhfoster@lucent.com
Stanny De Reyt:
Lucent Technologies Europe
Avenue Marcel Thirylaan 81
B-1200 Brussels, Belgium.
Tel: +32 2 7777 876
Email: sdereyt@lucent.com

This information can be combined to create three variations of the conventional oldest waiting call rule to assign a call to the available agent:

- the highest preference, highest priority, oldest waiting call is selected;
- the highest priority, oldest waiting call is selected; or
- the oldest waiting call is selected.

These rules alone may not achieve the results a centre seeks. Many call centres use call flow designs that change a call's queue priority or queue a call for one or more alternate skills in an effort to increase the chance of the call being served quickly.

The results of this type of control for a retail catalogue call centre in the United States are depicted in Figure 1. This diagram shows the daily average speed of answer (ASA) for sales and customer service calls. Although the centre provided an advantage to sales calls, customer service calls were too often handled sooner than sales calls.

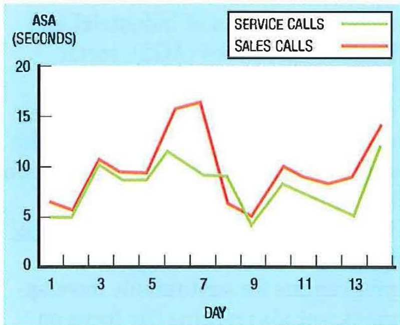


Figure 1—Call centre daily performance using conventional techniques

Re-inventing the Answer to What Should the Agent Do Next?

The use of the word *should* is deliberate in framing this question. Should implies some evaluation of alternatives or consequences in making a decision. What consequences are relevant in selecting a queued call?

New perspective on wait time

If two or more calls are waiting in the queue for the agent who has become available, the issue is one of call selection—deciding which call will be served. One element of consequence in this decision is the amount of time the calls *not* chosen will need to wait until other agents serve them.

A rare resource, if lost, presents a greater loss than a commonly found resource. If the available agent holds a skill that few agents handle, and

† The techniques described here were developed by Bell Laboratories, the research and development arm of Lucent Technologies. These techniques are embodied in software marketed by Lucent Technologies under the name CentreVu® Advocate and are the protected intellectual properties of Lucent Technologies.

another skill that many agents handle, the lost opportunity to use the rare skill must be considered. The understanding of how long a waiting call will continue to wait if passed up by the available agent is the foundation for building predictive and adaptive call centre operations.

A predictive evaluation of caller wait time

The amount of time a caller has waited and the amount of time a caller will remain waiting are both relevant in evaluating consequences. Bell Laboratories has developed a predictive algorithm, called *predicted wait time*: the sum of the caller's *current wait time* and the wait time until the next agent—not the one presently available—takes the call. The amount of time until the next agent is available is called *advance time*. Advance time is calculated for the calls the available agent might serve. The call with the greatest predicted wait time represents the neediest call among equally important calls.

Table 1 illustrates the evaluation of predicted wait time for queued calls. In this scenario an agent has become available and calls are shown queued for each of the agent's three skills—sales, customer service and inquiry. The calls arrived at various times and the current wait time for each call is known. The advance time calculation is the estimate of how long each call will continue to wait if not served by the available agent. The predicted wait time is the sum of the current wait and advance time for each call.

The call with the greatest predicted wait time in Table 1 is the customer service call. Notice that an oldest call waiting rule, which considers only the current wait time of the calls, would have selected the sales call. The sales call did indeed arrive three seconds before the customer service call, but a choice made according to predicted wait time avoids the longer wait the customer service caller would sustain.

Thinking about urgency

A consideration of only predicted wait time presumes that all calls are equally important. How long any call might wait for service, however, must be viewed in light of the urgency of the call. In a hospital emergency room some patients must be treated immediately. Others without critical injuries or illness will wait much longer. In a call centre, a 10-second wait may be much too long for one type of call but fine for another type.

Call centres have used both queue priorities and agent skill preferences to create attention for calls. Queue priorities and skill preference are both preemptive. The amount of attention a skill receives is dependent on the number of calls queued at higher priorities and on how many agents hold that skill at a preferred level.

A combination of queue priorities and skill preference levels seem practical in theory, yet when coupled with consideration of only the elapsed wait time (oldest call waiting) these designs yield fluctuating results. The centre depicted in Figure 1 (for example) used a straightforward design that provided advantage to sales calls.

A calibration of urgency

A calibrated and flexible method of evaluating concern for any caller's wait time would allow every type of call an agent handles to be considered while providing a meaningful distinction of calls in terms of relative urgency. A *service objective* or nominal gauge of good service can be defined for each skill and used as a comparison for the predicted wait time calculated for each waiting call.

Table 2 depicts the same call-selection scenario previously evaluated with the addition of service objectives or measures of urgency defined for each skill. A 15-second service objective is set for sales, a 20-second service objective for customer service, and a 25-second service objective for inquiry. These service objectives give distinct advantage, but not preemptive advantage, to sales over customer service and inquiry and to customer service over inquiry.

In this example, the call that will be served next is the one with the highest ratio of predicted wait time to service objective. The customer service call is selected through the effect of service objectives. A different business may value these calls differently and would choose different settings for the service objectives.

Creating alignment of service levels

The use of predicted wait time and service objectives in determining which call an agent should take aligns performance with business intent. Figure 2 shows the call centre in Figure 1 after adoption of these predictive and adaptive techniques. The sales calls have a consistently better ASA than customer service calls in alignment with the business' intentions. This centre used service objectives to bias call handling and obtained these characteristic 'heart beat' results. This alignment was accomplished, moreover, without real-time manual intervention on the part of call centre managers.

Identifying and Responding to Performance Problems

Call centres must perform well despite the random nature of call

Table 1: Using Predicted Wait Time in Call Selection

Skill	Current Wait Time	Advance Time	Predicted Wait Time (PWT)	Oldest Waiting Call	Call Selected
Sales	11 sec	9 sec	11 + 9 = 20 s	✔	
Customer Service	8 sec	24 sec	8 + 24 = 32 s		✔
Inquiry	3 sec	7 sec	3 + 7 = 10 s		

Table 2: Using Predicted Wait Time and Service Objectives in Call Selection

Skill	Service Objective (SO)	Current Wait	Advance Time	Predicted Wait Time (PWT)	Ratio PWT/SO	Call Selected
Sales	15 s	11 s	9 s	11 + 9 = 20 s	20/15 = 1·33	
Customer Service	20 s	8 s	24 s	8 + 24 = 32 s	32/20 = 1·6	✔
Inquiry	25 s	3 s	7 s	3 + 7 = 10 s	10/25 = 0·4	

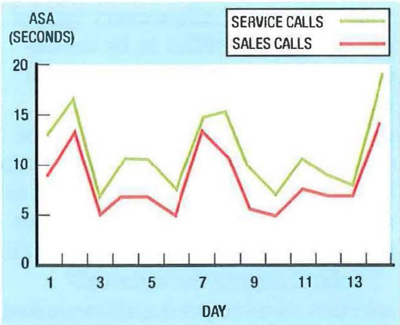


Figure 2—Call centre daily performance using predictive and adaptive techniques

arrival. Planning efforts, such as forecasting and scheduling, only prepare a centre for some baseline of operation. Deviations from the expected staffing, call volumes or mix or handle times for calls combine to create performance problems that conventional call-handling cannot overcome.

The manual intervention strategy

Call centre managers routinely try to adapt the real-time operation of the centre to maintain performance under unexpected conditions. These efforts follow a fundamental cycle. The manager must:

- watch the performance metrics of the centre,
- identify a problem condition,
- formulate a contingency solution,
- execute the solution,
- re-monitor the performance,
- determine when the problem has passed, and
- revert to the original method of operation.

This cycle takes precious time to execute. In addition, the metrics the manager observes are historical—a rolling ASA or the wait time of the queued calls. Caller wait times may already be too long before the manager can detect the problem. Often more than one problem must be solved simultaneously. Before a manager can return agents to their original skills the manager's attention might be drawn to the next problem.

A predictive and automated approach for adaptation

The technique developed by Bell Laboratories for managing performance in real-time allows potential problems to be identified and solutions applied before callers experience actual problems.

A method for predicting that a caller wait will become too long is to evaluate each caller's wait time as the call is placed into the queue and compare the prediction to one or more customer-defined *overload thresholds*. When a prediction exceeds the threshold, additional agents who could handle this type of call are activated or made eligible to serve that type of call. These *reserve agents* are part of the centre's contingency operation. When the prediction drops below the threshold, the pool of agents handling this type of call reverts to agents who typically handle the call under normal conditions.

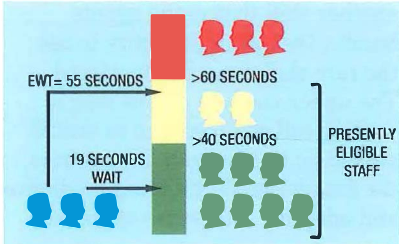
The prediction of a caller's wait time (used in contingency operations) is called *expected wait time*. The expected wait time calculation incorporates many dynamics of a call centre in its formulation. This predictive window into future call centre performance lets the centre's operation adapt to prevent calls from experiencing wait times above a desired threshold.

This concept is illustrated in Figure 3. A flexible system of thresholds allows a centre to align its automated contingency operations with its interests in eliminating long caller wait times. For example, an expected wait time calculation of 55 seconds for an arriving call is compared to the 40 second threshold set for the skill. Immediately, before this caller has experienced *any* of the predicted 55 seconds of wait, agents who hold this skill in reserve are eligible to handle this skill. The pace at which these calls are selected increases temporarily because the agent pool is larger. Multiple thresholds allow the reserve agents to become eligible in increasing numbers.

Perhaps only a few calls need to be served by this broader agent pool to reduce the potential for an unacceptably long wait time. When the expected wait time has dropped below the overload threshold, the contingency method of operations is halted and reserve agents are no longer eligible to handle this type of call. The thresholds can also be activated by the actual wait time of queued calls.

Reserve agent simulation

Table 3 shows the results of a very simple simulation using reserve agents. In this simulation, nine standard agents hold a single, standard skill where 110 calls arrive



Reserve agents become eligible to handle calls when an overload threshold is activated by a comparison of expected wait time (EWT) to preset thresholds or a comparison of callers' actual wait times to the thresholds. In this example, the last call's EWT triggers the eligibility of two reserve agents. Three additional reserve agents will not be eligible until a higher threshold is activated.

Figure 3—Activating reserve agents

per hour. The average handling time is four minutes. The performance results in terms of ASA, call completions and abandonments are noted. This simulation is compared to three scenarios using a 10th agent working first as a standard agent, then as a reserve agent with a threshold of 60 seconds, and finally as reserve agent with a threshold of 90 seconds. For simplicity in examining the contribution of the additional agent, the 10th agent is an exclusively reserve agent, meaning the agent has no other types of calls. In actual call centre implementations, however, reserve agents often handle other call centre calls.

Overall, any use of a 10th agent improves the throughput of the centre when measured as total completions per hour. However, the use of the 10th agent full time draws down the productivity of each agent by seven percentage points. When used as a reserve agent, however, the 10th agent's contribution has less impact on the standard agent's productivity.

A comparison of the additional calls completed to the amount of time the agent handled calls is one way to determine the beneficial leverage of the reserve agent. This is a measure of the leveraged contribution of the intermittent use of the reserve agent. In Table 3, the 10th agent, when working as a standard agent, is able to boost the completion rate by 1·8 calls/hour and is occupied 72% of the time. The reserve agent works with much greater leveraged contribution. Call completions rise by 1·4 per hour but with only 25% occupancy for the reserve agent with a threshold of 60 seconds and by 1·1 per hour with only 16% occupancy for the reserve agent. Considerable time remains for the reserve agent to perform other non-telephone duties.

Table 3: Reserve Agent Simulation Analysis (110 calls/hour, 4 minute average handling time)

Performance Measurement	Simulation Scenario: Baseline 9 Standard Agents	10 Standard Agents	9 Standard Agents 1 Reserve Agent 60 sec threshold	9 Standard Agents 1 Reserve Agent 90 sec threshold
Completions/Hour [A]	106·7	108·5	108·1	107·8
Completions /Hour over Baseline	Not applicable	1·8	1·4	1·1
Calls/Hour–Standard Agent [B]	11·9	10·9	11·6	11·7
Calls/Hour–Reserve Agent [C]	Not applicable	Not applicable	3·7	2·48
% Abandoned Calls	3·2%	1·5%	1·9%	2·2%
Average Speed of Answer	27·3 s	13·6 s	17·0 s	19·8 s
Maximum Delay (typical)	254 to 386 s	208 to 334 s	213 to 334 s	215 to 334 s
Maximum Queue Length (typical)	12 to 15 calls	10 to 12 calls	10 to 12 calls	11 to 12 calls
Occupancy–Standard Agents	79%	72%	77%	78%
Occupancy–Reserve Agent [D]	Not applicable	Not applicable	25%	16%
10th agent's leveraged contribution [E]	Not applicable	1·8 /0·72 = 2·5 calls/hour occupied	1·4/0·25 = 5·6 calls/hour occupied	1·1/0·16 = 6·9 calls/hour occupied

[A] Completions were greatest when the 10th agent held the skill as standard, but only marginally lower when the 10th agent worked less than 25% of the time as a reserve agent.
[B] Standard agents handle 1 less call/hour when the 10th agent is added as standard, but only 0·3 fewer calls per hour when the 10th agent is a reserve agent.
[C] The reserve agent handles more calls/hour at the lower, 60 second threshold but this work level does not appreciably lower the maximum delay over the contribution at the higher, 90 second threshold.
[D] The reserve agent's occupancy drops with the higher, 90 second threshold. In this simulation, the 10th agent is assumed to be at his/her desk, available for a call, but able to do other deskwork in the meantime.
[E] The leveraged contribution of the 10th agent is calculated by comparing the greater number of calls the centre can complete per hour divided by the occupancy of the 10th agent. This calculation magnifies the part time contribution of the low usage of the reserve agent.

Predictive and Adaptive Execution Each Time an Agent Becomes Available

If there are calls waiting for the agent to handle when the agent becomes available the various techniques described above are put into use sequentially to activate an appropriate contingency operation or to maintain normal operations.

If the agent has any reserve skills presently in an overload condition or any standard skills presently in an overload condition, business rules determine whether this call:

- is to be handled without consideration of any other calls waiting for standard skills;
- can be considered along with the other calls waiting for standard skills; and
- can be served only if no other calls are waiting for the agent.

If no reserve or standard skills are in an overload condition, a call is selected by examining the predicted wait times and service objectives for each waiting call:

- the predicted wait time is calculated for each call that the agent might handle;
- each waiting call's predicted wait time is compared to the service objective set for the skill; and
- the call with the greatest need, the highest ratio of predicted wait time to service objective, is selected and handled by the agent.

Distributing Work Fairly

Even in the busiest of call centres, there are many callers that do not have to wait for an agent. In some of these cases of agent surplus there are two or more qualified agents available. When a choice can be made, how the calls are distributed can create equitable workload for the agents, eliminating the 'hot seats' found in call centres.

When the *most idle agent* is selected from a group of available agents, the choice of agent is sensitive only to the amount of time since each agent completed the previous call. Selecting the agent who has had the least amount of call work—the *least occupied agent*—reduces the occupancy of busier agents and raises the occupancy of the less busy agents. The difference in these methods is illustrated in Table 4.

Often the agents who are the busiest are the agents with multiple skills. Their workload can be much higher than an agent who holds only one skill. The more skills an agent has, the more likely it is that at any time the agent becomes available there will be at least one call waiting for that agent to handle.

If a choice between two agents is possible and the least occupied agent is selected, it is more likely the case that the agent chosen has fewer skills, perhaps only a single skill. Allowing a very busy, multi-skilled agent to remain idle has two important benefits. First, the multi-skilled agent has a longer period of rest, resulting in a lower occupancy and perhaps less stress over the course of the day. Second, during that extended period of rest, another call may arrive needing one of the skills that this available, multi-skilled agent holds. This coincidence allows a call that would have otherwise waited to be answered immediately.

Table 4: Most Idle Agent Versus Least Occupied Agent as a Method of Agent Selection

Agent	Idle Time since last call	Work Time on ACD calls	Time since staffed-in	Occupancy (work time/ staffed time)	Most Idle Agent	Least Occupied Agent
Agent 1	30 s	50 min	60 min	83%		✔
Agent 2	35 s	90 min	100 min	90%	✔	

Benefits of Predictive and Adaptive Call Centre Execution

The benefits of using predictive and adaptive techniques in call centres accrue to callers, to the profitability of the centre, to agent fairness and agent efficiencies, and to overall ease of management.

First, callers will experience fewer long wait times because predicted wait time allows the detection of long waiting consequences for individual callers. Service objectives bias call selection to provide appropriate advantage to different calls, bringing performance into alignment with business intentions. If sales calls should be answered sooner than service calls or top tier callers should be answered sooner than mid-tier, the service objectives allow this distinction to be made.

Sensitivity to the consequences of not choosing a call eliminates excessive delays, often causing maximum delays to drop. With fewer long wait times, abandonments often drop and more calls are completed. In a revenue-centric call centre each additional call handled raises revenue and reduces the caller's likelihood of calling a competitor. Each additional call completion eliminates a later retry and additional network costs, or perhaps a lost customer.

With fewer call abandonments, more calls are completed and agent productivity as measured in time spent on calls increases. The rules-based activation of reserve skills may reduce the number of back up calls agents take, potentially reducing call handle times. For example, agents may need to switch between transaction tools less often or the agents' increased focus on standard skills helps them proceed quicker or sell more on each call.

In actual call centre implementations, the automated mechanism for moving from routine to contingency operations has proven to eliminate real-time manual intervention. The time that managers had devoted to monitoring performance, planning and executing solutions and returning agents to standard operation can now be applied to more strategic issues of planning, coaching, and quality management.

Conclusion

The oldest call waiting rule, while practical in early single-function call centres, limits today's call centre in achieving performance levels aligned with the business' intentions. As the role of the call centre becomes more strategic to the concerns of the business, the call centre must incorporate the business' intentions more directly into its operations.

The predictive and adaptive techniques developed by Lucent Technologies and outlined in this paper reinvent the call centre at the core of its operations, with new considerations made each time a caller and agent are brought together. This re-invention permits the call centre to execute in alignment with its intentions in a consistent fashion and to eliminate corrective manual intervention.

Biographies

Robin Harris Foster
Lucent Technologies

Robin Harris Foster is the Manager of Research for Lucent Technologies, Call Centre Advocacy, and a former member of the technical staff of Bell Laboratories, the research and development arm of Lucent Technologies. Robin holds several patents in the field of call centres and multi-media call centres, including predictive and adaptive techniques offered as CentreVu® Advocate. She is active both in partnering with Bell Labs to advance the state-of-the-art in call centre technologies and in translating those advances into business-useful call centre applications for Lucent's customers worldwide.

Stanny De Reyt
Lucent Technologies

Stanny De Reyt has over 10 years of diversified experience in the telecommunications arena. He played a key role in starting the PABX business for Lucent Technologies (formerly AT&T) in Belgium and Luxembourg. He is known for his ability to apply innovative technology to meet customers' needs. Together with a large Belgian call centre customer he introduced to the Bell Labs scientists at Telecom Geneva in 1995 the concept of the *customer contact centre*. Two years later this call centre was able to provide to its customers the first Internet call centre applications outside the United States. His expertise spans a wide range of call centre technologies including automated call distribution, voice response, computer telephony integration, and outbound telemarketing systems. He has developed in-depth knowledge of the components necessary to implement and manage a successful call centre, especially in the insurance and financial industries and in the global system mobile operator business sector in Europe and the Middle East.

Elina Nikaki

Internet, TV and Telephony: Are They Finally Converging?

This paper presents some of the forms in which telephony, computer, Internet and television technologies—and markets—'integrate'. It analyses the technologies deployed, the products offered, and the major industry activities in each area: cable telephony, Internet telephony, Internet television, computer telephony integration, and Internet—Telephony—TV integration.

Cable telephony involves the offering of telephone services, and high-speed data, over the hybrid-fibre-coaxial networks that multiple service operators (MSOs) own. Cable operators have embraced the idea by offering cable modem solutions, and major telcos are considering it an opportunity to expand their network footprint.

Placing and receiving telephone calls over the Internet, 'Internet telephony', becomes increasingly popular. Through the various implementations, PC-to-PC, PC-to-telephone, and telephone-to-telephone, IP telephony is expected to take over significant percentages of the communications market.

'Internet television' products offer the benefits of interactive, on-demand, multimedia videos to complement existing television programming on the PC or even TV screen.

Call centres, the main products of computer telephony integration, are part of our everyday life and a significant component of customer services of many businesses.

Finally, products that enable television programming, telephony features, and Internet access over our TV sets promise to make 'web-tone' as popular as dial-tone.

All areas of convergence seem to gain increased industry interest, and market attention. The costs of implementation and the resolution of any technological constraints, along with the industry activities and overall market acceptance, will determine the forms of telephony integration that will prevail.

Introduction

With the recent announcement of USWest's '@TV' trials later this year, the issue is again in the foreground: are Internet, TV and telephony finally converging?

The answer to this question probably captures the vision of our future communications: a single screen that connects us with the rest of the world so we can talk and see our beloved ones, send our documents to our colleagues, while watching movies and researching the web at the same time.

Elina Nikaki:
Deloitte Consulting,
Two World Financial Center,
New York, NY 10281-1414
Tel: +1 212 436 2923
Fax: +1 212 655 8176
E-mail: enikaki@dttus.com

This paper attempts to present some of the various forms in which computer technologies, Internet global networking, television programming, and telephony advanced services and features intertwine to produce a wide range of products for the present and the future.

Alliances and mergers among the major leaders of the computer, telecommunications and entertainment industries promise to redefine our present notion of 'communications' and guide us into the unimaginable interconnection possibilities of the Millennium.

At the end of this century, and the beginning of the new one, the blurry boundaries between our PC, conventional telephone sets and TV screens disappear for the sake of 'convergence'. Diverse naming schemes are used to describe the numerous integration forms: *cable telephony*, *computer telephony integration*, *Internet telephony*, and *Internet television* are only few among them.

Any attempt to define accurately these 'convergence' terms and classify strictly the various existing products that fall under them would be worthless. Especially, since most of these terms are often used loosely by the industry, and their definition areas are overlapping.

The following paragraphs will try to cast light on the most recent industry activities in the main areas of convergence, in an attempt to evaluate the current market pulses, and estimate the industry's future plans. More specifically:

- In the 'Cable Telephony' section, the transmission of telephone signals over cable (HFC) networks is described.
- In the 'Internet Television' section, television serves both in its conventional form to receive broadcasted cable entertainment signals and as an access device to the Internet world.
- In the 'Internet Telephony' section, the delivery of telephony over the former data networks, the expansion of Internet protocol (IP) telephony, and the use of PCs to place and deliver telephone calls are discussed.
- In the 'Computer Telephony Integration' section, computer data are entered through the telephone handset, and telephony features are appearing on a computer screen.
- Finally, in the 'Internet, TV and Telephony Convergence' section,

our former TV set may equally be our telephone, the home entertainment theatre and a user-friendly web browser.

Convergence 'happens' either in the telecommunication networks, or at the receiving-end devices. And in all of these convergence forms we watch the 'converging' of the formerly distinct industry segments.

Cable Telephony

Cable telephony enables users to receive telephone service over the cable network, instead of the traditional twisted-pair copper telephone network.

The communication networks owned by cable companies, called *hybrid fibre coaxial* (HFC), are composed of fibre optics that run up to the head-end stations and coaxial for the last mile to the customer premises. HFC networks offer increased bandwidth to the user, and, with slight network alterations, they can provide cable, telephony and high-speed access to the subscribers' homes.

Pure cable HFC network implementations offer 450 MHz up to 750 MHz or more capacity. Upgrades to the network increase the upstream bandwidth, so that the network can operate 'two-way'. In the upstream, 330 MHz–450 MHz is used for the cable, and the 5–42 MHz band is used to carry telephony and data signals, while in the downstream traffic 50–750 MHz channels are allocated for telephony and data. Cable modems at the head-end stations and the customer premises communicate through these channels allocated for data and telephony.

Cable telephony may be either circuit-switched or, more recently, IP cable telephony.

Circuit-switched cable telephony implementation requires the multiple service operators (MSOs) to build another logical network over the same physical cable infrastructure (high-speed data architecture). The disadvantages of such an implementation are clear: operational, economical and technological (spectrum) inefficiencies; cable operators are required to allocate separate upstream and downstream bandwidths to each service, and maintain and operate two technologically distinct operating support systems and platforms.

In IP cable telephony, telephone audio signals are digitised, formatted into packets and transmitted over data networks, received and decoded at the other end in the receiver's telephone handset. MSOs can carry such telephone signal packets over their existing networks without major changes once the networks have been upgraded to support two-way data traffic and the appropriate modems are installed at the subscriber's premises.

Upgrading the cable network to two-ways may be very costly, estimated around \$200–\$250 per home. Therefore only some of the MSOs have already upgraded their existing network infrastructure†. Others are offering telephony over one-way cable systems: fast downstream transmission through cable modems, and use of traditional telephone modems for the upstream communication over the public switched telephone network (PSTN). MSOs that are following this strategy are trying to gain a small share of the Internet services market before they raise the capital to upgrade their networks.

With cable IP telephony, cable operators will be able to offer multiple services over one platform. The services will include voice, data, enhancements for cable TV services, and value-added features, such as integrated voice mail and e-mail messaging, and real-time provisioning of additional telephone lines without having to rewire a home. Technological challenges to offer toll-quality telephone service seem to be the only obstacle.

The cable market offers opportunities for software vendors, computer components manufacturers, telephone equipment suppliers, cable operators, telephone carriers, and Internet service providers. Mergers, acquisitions and strategic agreements have drastically changed the competitive landscape, which is expected to change even more.

Cable operators, instead of offering cable telephony and data services directly, have responded to the demand for high-speed Internet services by selling cable modems. North America MSOs serve 550 000 cable modem subscribers*, outpacing telephone companies in the race to offer broadband Internet access. Cable modem services have been available to more than 20 million homes, or 20% of all cable homes in North America. More than 85% of these subscribers are receiving service with two-way cable modems and the remaining with telco-return products. As a group, the North American cable operators are currently adding more than 2000 cable modem subscribers per day.

The potential revenues are expected to be enormous. Internet telephony may cause telcos \$8 billion loss by 2002, and reduce by 6% the US telephone traffic that the incumbent local exchange carriers (ILECs) are currently handling‡.

Some of the recent activities from the market leaders include:

- *AT&T's* 'flirting' with cable telephony. With the acquisition of TCI, North America's largest cable operator AT&T is planning to deploy HFC networks to enter the local telephony markets. The acquisition of TCI has not been AT&T's only move into the cable market. AT&T has been trying to sign up other MSOs for cable telephony joint ventures to increase its US broadband service footprint.
- *@Home's* high-speed Internet service offerings provide an alternative to PC-based Internet access platforms. @Home connects its subscribers to its broadband IP backbone network with cable modems. The company is also offering comprehensive networking and systems integration services to MSOs that are entering the Internet business.
- *Road Runner*, Time Warner's high-speed data venture with Media One, Microsoft, and Compaq, already claims more than 180 000 cable modem subscribers.

Internet Television

Internet TV is perceived as a single product that combines Internet access and television programming. In reality, this product may have technologically two forms: intercast technology, and streaming video.

---

† In the US market: Time Warner, Media One, Comcast, Cox Communications and Rogers Cable Systems are among the ones that have upgraded their HFC networks to two-ways.

* Kinetic Strategies estimates, Byrnes, Cheryl. Cable Telephony Market Trends, Faulkner Information Services, April 1999.

‡ Byrnes, Cheryl. Cable Telephony Market Trends, Faulkner Information Services, April 1999.

Intercast technology is the result of the combination of the programming of television, the global connectivity of the Internet, and the interactivity of the PC. The intercast standard has been developed and promoted by the Intercast Industry Group (IIG), which is a consortium of computer hardware, and software manufacturers, on-line service providers, and television programming networks, including industry leaders such as: AOL, CNN, Intel, CNN, Compaq, Netscape, and Time Warner Communications.

How does it work? Content providers, such as television stations, enhance their existing programming and provide interactive content. This interactive content is 'pushed' via the cable TV wires, along with the TV signal to a PC that is upgraded with an intercast receiver card and software. Intercast content is cached in the computer's hard drive and users can see the intercast programming on their PCs at any time. To receive an intercast broadcast a PC should be connected to a television antenna, a cable TV connection, or a TV-satellite dish. To access the hyperlinks featured in the intercast broadcast the PC should be connected to the Internet via a traditional modem.

Intercast applications are currently limited to complement existing product offerings. An intercast broadcast, for instance, can include extended interviews and videoclips to supplement a nightly news programme, or interactive video games added to a children's television show. Intercast technology redefines in its own way interactive multimedia.

The chief producers of intercasting content are the members of IIG including CNN, and CNBC. Unfortunately, intercasting is only available in some areas.

Streaming video is a live or pre-recorded video broadcast that is transmitted over a network or the Internet and presented in real time on the computer screen. Streaming video also can be stored on a hard drive or some other storage device to be replayed at a later time. A PC with a modem and a standard Internet connection is adequate to receive and view streaming video, once the user downloads one of the many free streaming video players available on the Internet. Such videos are found at news- and entertainment-related web sites.

Producing streaming videos, although not cheap or simple, can be done by anyone. One should get the production software package, which may cost $400 to $1400, and obtain (prices ranging around $1000 and up) or contract (prices start from as low as $100) a streaming video server. Nevertheless, streaming video cannot match the quality of television because of the current limitations of compression techniques and bandwidth.

Another interesting venture in the Internet TV applications area is *DirecPC*, the high-bandwidth, satellite-based access medium. DirecPC transmits television broadcasts from major networks, such as CNN, and ESPN to the user's computer system, but it does not combine Internet data with television programming, as intercasting does.

By combining the television programming, the global connectivity of the Internet, and the interactivity of the PC, Internet TV technology develops a new series of services for the home and the office. Such services involve digital TV entertainment, data-enhanced TV programming, and video-enhanced multimedia magazines and can be used for entertainment, business, or educational purposes.

The potential applications range from the obvious—users can read the biography of an actor while watching the movie in which he is the star—to the fantastic—users can watch a featured movie whenever they have the time, instead of when it is broadcast. Additionally, Internet TV enables non-profit organisations and special interest groups to produce their own news programmes to counter the mass media. Equally, entrepreneurs may easily start their own Internet television stations broadcasting programmes to and from anywhere in the world.

Nowadays, Internet TV is a novelty that offers only a few practical benefits for the average individual or business. There are only few application available, which appear to be redundant and unimpressive and very narrowly focused at best. For instance, most of the features of CNBC's intercasts are available on the Internet or are of interest to a very limited number of people.

Due to the generally poor quality of streaming video and the narrow market for producing intercast programming, very few businesses and individuals are producing programmes for Internet TV. Those who are, tend to include high-income businesses and individuals who want to be on the cutting edge of technology, entrepreneurs who see Internet TV as a potentially lucrative investment, and television stations that have resources to explore the technology.

In a few instances, however, Internet TV has been and is being used in an exceptional manner. For example, when the Mars Pathfinder Mission video-taped the surface of Mars, the videos were displayed through streaming video feeds on the Internet. Such a case shows the possibilities of Internet TV.

It seems like the only thing that is currently limiting the capabilities of Internet TV is technology: a lack of bandwidth and poor video quality leave it lagging behind the television and computer markets. As soon as a universal high-bandwidth solution prevails such limitations will no longer exist.

Internet Telephony

Telephone calls via the Internet have already become commonplace. They have been very popular mainly because of their low cost compared to the long-distance or international call prices offered by telcos. Most PCs on the market have currently sufficient multimedia hardware that they can be effectively used for Internet telephony, and Internet telephony software is available for free or at very low costs. Additionally, the sound quality has significantly improved, and the existing customer base of Internet telephony products is significant, especially in groups such as university communities.

There is a wide range of phoneware products, that allow PC-to-PC communication, though the most recent industry activities in this area involve PC-to-conventional telephones Internet communication.

For instance, IDT's Net2Phone Direct service allows users to place international calls at much lower prices than the ones offered by conventional long-distance telephone companies†. IDT deploys its main telephone switches to handle calls placed by PCs and routed to conven-

† The cost of such an international telephone call, from other countries to the US, will be about $0.10 as opposed to $2 per minute rate offered by a traditional telephone company. Miller, Stewart. Internet Telephony Implementation Guide, Faulkner Information Services, 1998.

tional telephone sets. Software installed on the PC simply initiates the telephone call. In this implementation there is no need to establish an Internet connection between two PCs that have the same phoneware software in order to set up an 'Internet telephone line'. Such implementations of the past have been very cumbersome at times. IDT is currently improving its Net2PhoneDirect telephone-to-telephone service, and is installing servers in the countries that have been selected to receive such calls. Quality is still not at toll-level, and the product is primarily expected to be used by residential consumers, than businesses.

In the arena of PC-to-PC Internet telephony, Microsoft and Netscape are offering Internet telephony as part of the software platforms.

Finally, in this area where computer software companies have traditionally shown interest, traditional telephone companies are also starting to participate. AT&T WorldNet, for example, is planning to offer an Internet-based long-distance telephone service.

Computer Telephony Integration

Although, CTI seems to have significant growth, its applications area seems to have been narrowed down mainly to call centre implementations or integrated in other equipment such as PBXs instead of being a separate technology.

Its growth has mainly been supported by demands at the work environment, of very large businesses with increased customer support functions. CTI provides the way to apply the power of a computer to the functions of the telephone.

Major communication leaders, such as Lucent, and Nortel are continuously employing their resources and growing their business in this area. CTI functionalities are expected to be integrated in even more devices in the future.

The CTI industry is only beginning to take advantage of the integration of voice and data communications within the LAN environment. The benefits of the integrated voice and data communication networks involve the ability to treat voice messages and electronic messages in unified way and to edit a document with voice annotations.

Internet, TV, and Telephony Convergence—All in One?

During the second quarter of 1999, USWest announced an innovative 'converging' product. Initially named *AtTV* and later christened *Web Vision*, this product is about an original integration of telephone, television and Internet services.

This service will allow users to send and receive e-mail, place and answer telephone calls, surf channels and the Web—or surf the Web and TV channels at the same time—on their televisions. One will be able to check e-mail messages between one's favourite sit-coms, or view caller identifications right on the TV screen before answering the telephone during a crucial plot twist.

The product is still at the technical trial stages, and is expected to reach the market during the fourth quarter of 1999, initially in selected metropolitan areas, and later in the year the product will be launched in regions of several states.

The technology deployed comprises a television set-top box equipped with a speakerphone and proprietary software to receive and make telephone calls and access Internet-based features. Some of these features include programming guides, electronic commerce, news and electronic mail. In addition to access over conventional connections, support for high-speed digital subscriber line (xDSL, VDSL and ADSL) technologies will be provided to offer data-transmission speeds up to 200 times faster than the conventional dial-up connections.

Although the product aims to offer the benefits of Internet access to the households that do not own PCs, through the user friendliness of one of our most familiar household devices (TV), anyone who seeks for the convenience of an 'all in one' communications medium will be delighted by this product offering. Or, better put in the words of Eric Bozich† 'the goal is to make Web tone as common as dial tone'.

Conclusion

It seems that in addition to vendor and carrier activities, the costs of implementation and resulting market place acceptance will play key roles in defining the future of telecommunications convergence.

The potential of such applications is limited only by the imagination. They have the power to make computers and the Internet part of every home and to bring every business or product into every home, by combining the forces of entertainment and computer industry

Until now only minor steps have been realised towards this direction, but there is so much more to be seen in the future.

Biography

Elina Nikaki

Elina Nikaki was born in Athens, in 1972. She completed her undergraduate studies in Computer and Electrical Engineering at the National Technical University of Athens, and her Masters in Information Networking at Carnegie Mellon, Pittsburgh, Pennsylvania. In her professional life, she has had various internships in the telecommunications industry. Among them, she participated in the design of the geographic information system to manage the network maps of the Greek Telecommunications Organisation (OTE). She has also worked for USWest in Boulder, Colorado evaluating alternative technologies to integrate cable and telephony networks. Since September 1998, she has worked as a telecommunications consultant with Deloitte Consulting, in New York. She has been involved in diverse projects involving strategy planning, technology benchmarking, and systems integration in major telcos. Her most recent project entails AT&T's entrance into the local market in Texas.

† US West Vice President of Internet and Applications.

Christos Solomonides and Mark Searle

An Intelligent Network-Based E-Commerce Protocol

This paper describes a protocol for electronic payments using existing intelligent network (IN) infrastructure. This is of particular benefit to network operators who already have a large investment in IN infrastructure. Through this protocol they can extend their existing billing systems to support electronic commerce (e-commerce). Users can be charged for micro-payments on their telephone bill. This way they may enjoy the benefits of having an easy, relatively accepted and recognised billing system.

Introduction

Most European and United States operators have invested heavily in intelligent network (IN) platforms to support key bearer control services such as freephone and premium rate. Such systems have complex structures for interfacing with existing billing systems that are resilient and trusted by the end users. In contrast there is still much mistrust over the security of the Internet.

Currently, many electronic payment systems require the use of a credit card. Although reported instances of fraud are rare there is still reluctance by individuals to divulge their credit card details. This framework is limiting both for end users and businesses.

The purpose of this paper is to describe a protocol for electronic payments using existing intelligent network infrastructure. This allows users to charge for micro-payments on their telephone bills. The paper firstly presents the protocol designed and simulated using Java. It then goes on to talk about the relationship between the proposed system and existing electronic payment systems. The final part of the paper presents a number of other applications where IN systems can be used in support of the Internet.

Christos Solomonides:
c.solomonides@ee.ucl.ac.uk
Mark Searle:
m.searle@ee.ucl.ac.uk
Both authors are at:
University College London,
Torrington Place,
London WC1E 7JE.

The Intelligent Network-Based E-Commerce System

The protocol allows users to perform on-line commerce, including micro-transactions, without the need for credit cards. Currently most electronic payment systems require the use of a credit card. However a lot of studies have shown that users are uncomfortable with submitting their credit card information on-line. This has been one of the most limiting factors regarding the widespread use of on-line shopping.

A restriction is imposed by the high transaction costs of credit cards. These arise because of the overheads involved in processing credit card transactions. An impracticality arises from the requirement that, in most situations, users must enter their credit card number, expiry date, cardholder name as well as their address. This is tedious, time consuming and is a deterrent for such transactions. Credit cards are therefore uneconomical and impractical for use in micro-transaction services. However such micro-transactions are common in IN services like freephone where the cost per transaction is very low to support such services.

Operators who have implemented IN infrastructures will have essentially invested in a set of interfaces to complex management systems such as billing, customer services systems etc. It is in establishing such interfaces that much of the cost of IN deployment cost is expended and such systems represent important and reusable assets.

The application described in this paper allows end users to perform on-line commerce without the limiting requirement that they must be credit-card holders. The system reuses the telephone billing mechanism, which has for the most part been accepted by, and is trusted by, end users as a means of charging end users. Three parties are involved in the system: a network operator, Internet shops and end users. The scenario is outlined as follows: a network operator is responsible for managing an Internet shopping mall. Internet shops wishing to be part of this mall must sign a contract with the network operator and reach an agreement regarding the transfer of funds from the network operator to the Internet shop. The network operator also provides end users with a small application that is used when they wish to use the system.

The framework provides some degree of protection to end users because they feel they have the protection of the intermediary party—that of the network operator. A shop can use a combination of credit card billing for larger payments or where a user is happy to divulge credit card details, and use the micro-payment system for other transactions.

The Protocol: High-Level Architecture View

The protocol is divided into two phases: a registration phase and a transaction-processing phase. During the registration phase, the identity of the end user is established and authenticated so that payments can be billed to the home telephone bill. Having passed the registration phase, end users can

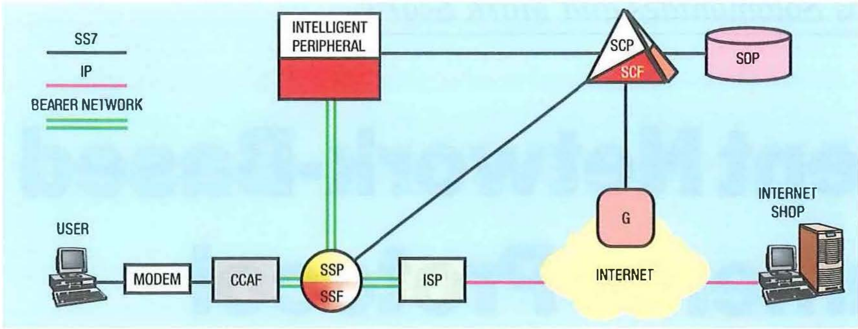


Figure 1—High-level architecture view

perform on-line transactions. The transaction phase can also be subdivided into two phases, an information interchange stage and a billing stage. Figure 1 depicts the high-level architecture view of the system. The Call Control Agent Function (CCAF) may be viewed as a terminal through which a user interacts with the network.

A key point regarding the protocol is the link between the service switching point (SSP) and the service control point (SCP). This is a Signalling System No. 7 (SS7) link, which is relatively resilient and suitable for allowing the secure identification of the user. The link between the intelligent peripheral and the SSP is used as a means of collecting information from the user. Also the final link between the gateway (G) and the SCP, which is again an SS7 link, is used for the billing phase.

An important element depicted in Figure 1 is the element between IP networks and the IN infrastructure (marked G). There are a number of possible ways for implementing such a gateway, two of which are described in reference 1. The authors take the view that the gateway should be considered as a standardised IN entity with SSP capabilities.

Registration phase

The registration phase of the protocol makes use of the existing SS7 network. During the registration phase, the user calls the Internet service provider (ISP) as depicted in Figure 2, flow 1. The number assigned to this service constitutes a service access number (SAN). The service switching function (SSF) will recognise the number as a SAN (universal access number) and will query the service control function (SCF) for call-handling support, flow 2. The SCF will then instruct the specialised resource function (SRF) to collect the personal identification number (PIN) of the user using the *PromptAndCollectUserInformation* operation[2], flows 3 and 4. The *ReceivedInformationArg* returned by the operation will then contain the PIN number entered by the user, flow 5. The SCF must now check the PIN number of the user stored in the service data point (SDF). For this, the SCF will use the 'query' operation to check for a match in the SDP, flow 6. If a correct match is made, the user is authenticated and the SCF will provide the user with a connection number, flows 7 and 8. This number is used in the final stage of the protocol for the billing.

At this stage a user has a normal IP connection but the software on the user's machine holds a connection number, which can be used to perform the on-line transactions. When the connection is terminated, the software on the user's machine will invalidate the connection number.

A user is free to browse web pages and make selections for purchase by filling a shopping basket. The transaction phase is triggered when a user has finished making selections and wishes to place an order with the Internet shop. The user may initiate this by clicking on an icon to place the order. The transaction phase is described in the next section.

Transaction phase

The transaction phase of the protocol is depicted in Figure 3. Three servers have been set up, representing the gateway, an end user and an Internet shop (IS). The simulation was implemented in Java. For the simulation a servlet simulator was used (using a server), as the web server available did not support servlets. The shaded area in Figure 3 indicates the fact that the *servletSimulator* and the Internet shop server (ISServer) are running on the same PC. In a commercial implementation of the system, the

Figure 2—User calls Internet service provider

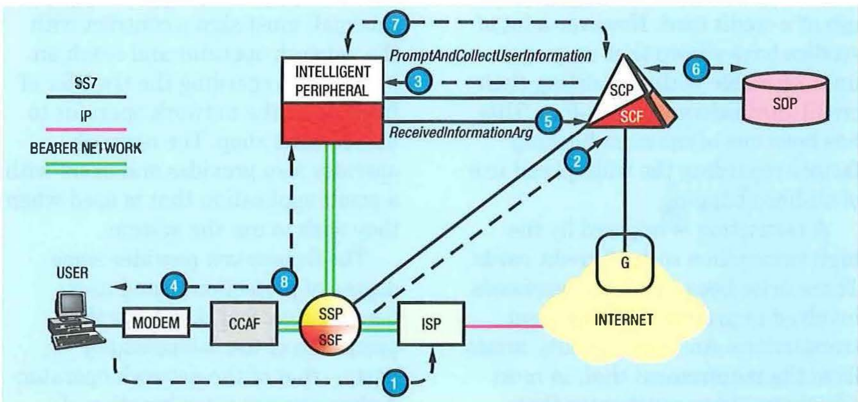
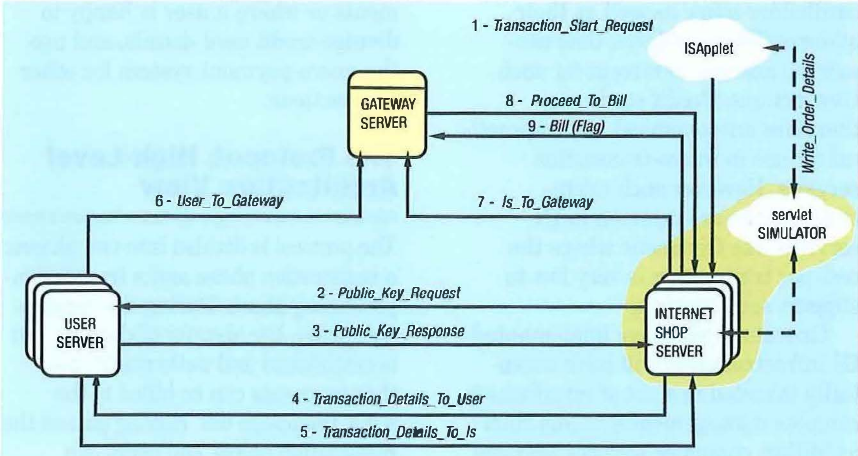


Figure 3—Protocol implementation



UserServer (UServer) is supplied to the user as part of the provided software package. The network operator is responsible for implementing the gateway server (GServer). Internet shops are required to implement their own servers, which would operate on a clean interface provided by the network operator to them.

The transaction phase is triggered when a user has finished shopping and has selected to proceed with placing the order. This causes the Internet shop (IS) Applet (ISApplet) to send a *Transaction_Start_Request* to the ISServer. The ISApplet also submits the details of the order to the ISServer using the *write_order_details* packet. This flow is not required if a servlet is running on the ISServer.

Packet 1 causes the ISServer to send a *Public_Key_Request* to the UserServer (UServer) to which the UServer replies by sending the public key. The request sent by the ISServer also contains the IP address of the IS. This is required in order for the UServer to know to which IS the reply must be sent.

The next stage is for the ISServer and the UServer to exchange information. This is achieved by the sending of packets 4 and 5. These packets contain information regarding the transaction cost, the connection number assigned by the SCP to the user during the registration phase and the unique transaction numbers generated by the IS and the user.

Having exchanged information, both parties must now submit the information to the gateway. These packets are encrypted using the public key of the network operator, which is known by all users and Internet shops. Once the GServer has received a matching pair of packets, it checks the information contained within the packets and makes sure that all the transaction details match. Another check performed by the GServer is the authentication of the ISServer as an authorised Internet shop through agreement with the network operator. This is achieved using the IP of the Internet shop, which is contained in packet 6, as well as interrogating the originating address of packet 7.

Following the correct matching of this information, the GServer will send a final *Proceed_To_Bill* message to the ISServer. The need for this packet is for the ISServer to perform a concluding check that the transaction number contained within the packet is valid and allows the ISServer to check stock levels by interfacing to existing stock control systems etc. The ISServer will then reply to the GServer by returning a true/false flag to the GServer. A false flag will cause the cancellation of the transaction. A true flag will cause the protocol to move to the final stage of billing. This is discussed in the next section.

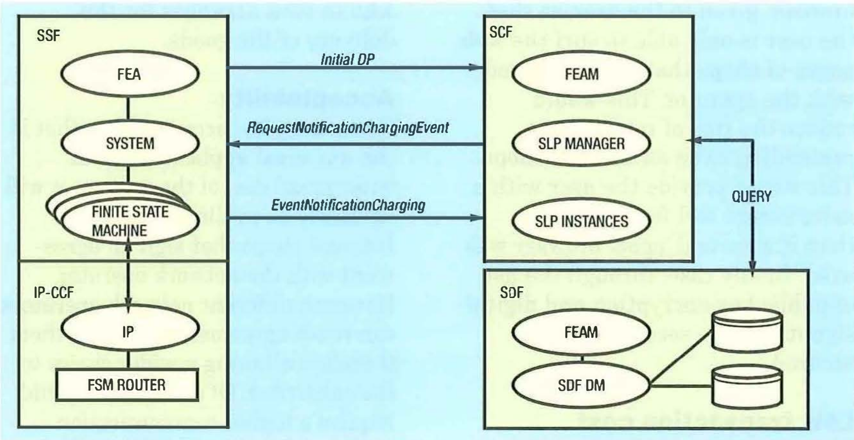


Figure 4 – Billing stage of protocol

Billing stage

Figure 4 depicts the information flows and operations sent from the Gateway to the SCF in order to bill the customer. Firstly there is an *InitialDP* operation from the gateway (SSF) to the SCF. The *InitialDPArg* contains the serviceKey argument, which will be used by the SCF to trigger the service logic. The SCF will then reply by sending a *RequestNotificationChargingEvent* (after it has sent a query to the SDP). The final stage involves the gateway sending an *EventNotificationCharging* operation to the SCF to perform the billing.

Characteristics of the Protocol as an Electronic Payment System

The proposed protocol does not require an operator to make changes to the intelligent network application protocol (INAP) or SS7. Standards CS-1 information flows are used. The network operator needs to create a billing service logic script in the SCP and associated announcements and digit collection procedures in the intelligent peripheral. This should be a relatively simple procedure with a standards service creation environment.

One new functional element is required for the system, the gateway. This device allows IP-based applications to trigger SSP-like transactions. However, once implemented, the gateway can be used to support a number of new services. Much of the gateway functionality can be reused in other applications. The system also requires the end user to install billing software, which is provided by the service provider or network operator.

System security is a major issue. Password information and secure keys are transmitted over a relatively secure link and is not transported over the Internet. Clearly, the first phase of the protocol is in isolation from the Internet. However, unlike most systems, which utilise a network that is in isolation, it does not require the setting up of a complete network infrastructure. It makes use of existing SS7. The use of the SS7 network for high-value data is a concern to network operators. Currently SS7 networks are not particularly the targets of malicious intent but sending billing information can lay it open to future attack. The second phase of the protocol, which makes use of the Internet, is relatively insecure for intrusion. Hence, the use of public keys transmitted over the secure link is essential.

Currently, there is the restriction that the user can only make purchases via their own home telephone line. Hence the PIN number entered by the use identifies them with the home line. It is acknowledged that this is restrictive. Universal personal telecommunications are a set of standards for allowing user registration on any terminal; such mechanisms could be used to support the billing application in future.

Another option open to the network operator or service provider is to configure the web browser, given to the user, so that the user is only able to surf the web pages of shops that are registered with the operator. This would reduce the risk of rogue shops pretending to be authorised shops. This would provide the user with a more secure feel for the system than if a normal 'open' browser was used. In any case, through the use of public key encryption and digital signatures the second phase is secured.

Low transaction cost

The system has a low transaction cost. This is because the on-line clearance that is required is obtained by the gateway after it requests an authorisation from the SCP. The SCP only needs to make a database enquiry in the SDP to check the credit limit of the customer before it authorises the transaction. The protocol could also be enhanced to operate as a form of a prepaid service, where the subscriber buys credit from the operator in advance and therefore further limits the need for on-line clearance.

Traceability of payments

Concerning the traceability of payments, the protocol could be adopted from conditional to user-controlled traceability. This means the user can choose whether or not his/her identity is revealed to the Internet shop. From the user's perspective to the network operator, the first phase of the protocol is classified as being unconditionally traceable (that is, the transaction generates a record that identifies the buyer, seller and the amount). This is due to the fact that the PIN number is associated with the physical connection. However, from the Internet shop's perspective the system could be classified as offering a user-controlled traceability (that is, the user can control the level of traceability). If the user is buying a service, then the system is totally user-controlled. However, if the user is buying goods then the Internet Shop needs to know the delivery address of the user. Currently the protocol offers unconditional traceability in this respect (that is, payer and payee are always identified). This can be changed and instead of the user providing the information to the Internet shop, the information could be sent to the gateway, submitted to the network operator, who in turn arranges for the delivery of the goods.

Acceptability

Concerning the acceptability (that is, the universal applicability and widespread use) of the system, it will obviously be available only to Internet shops that sign an agreement with the network operator. However, different network operators can reach agreements between them therefore allowing a wider choice to the subscriber. Of course this would require a higher communication overhead between the involved network operators, and thus slightly increasing the transaction cost of the system. At present the system does not offer transferability of funds (that is, the transfer of funds between the parties involved) and there should be no need to do so.

Comparisons with Commercially Available Systems

This system, unlike some currently available systems does not require the user to open an 'account' with each of the Internet shops. In does require the user to subscribe to the e-commerce system with the network operator/ service provider. The billing account is already in place and in a sense the user is pre-vetted as an existing telephone customer. Unlike other systems it does require for the providing organisation (network operator) to form alliances and reach an agreement with each of the shops. However, since the shops themselves are not responsible for managing the per-transaction billing their overheads are very low. They also have the advantage of having a large customer base being given over to them.

Also an advantage of this system in comparison to existing ones is that it does not use credit cards and hence completely avoids the existing controversy present among the public of whether submitting credit card information over the Internet is safe or not. It could easily be marketed as a system that does not require credit cards because such systems are not secure.

IN Infrastructure and IP Services

This section presents a brief overview of two further potential applications that are being considered where an existing IN system can be used in support of the Internet. The first of these provides support for mobility through the domain name system (DNS). The SCP can treat the DNS as a service data function. Hence, since CS2 provides enhanced support for user interaction and service profile customisation the user could register a binding between the IP address and URL through the IN. Number translation services are relevant to the Internet. The DNS currently services as a resolution protocol allowing the translation of fully qualified domain names to IP addresses. In this sense the DNS is an IN platform. The difference is that the end terminal rather than the switch (SSP) interrogate the DNS. A structure such as the DNS has the potential of being an interworking function for the control of services in the Internet. In such a scenario the DNS could be used to provide personal and terminal mobility services.

The second application treats the SCP as a general certification authority. The SCP can provide access for all manner of secure information such as passwords that could be used in Internet applications.

Finally, the role of application programming interfaces (APIs) such as Parlay[3] should be acknowledged. The Parlay Organisation is a working group formed by a number of leading telecommunication companies as well as software developing organisations. Its aim is to produce an open, technology-independent and extensible API specification for access to telecommunications control equipment. Parlay is currently defining interfaces to enable third-party service providers to provide customised access to telecommunications equipment such as SCPs. Parlay interfaces are object oriented and are described using Unified Modelling Language and Interface Definition Language. The potential of Parlay and other APIs is to allow other IP-based applications to make use of existing management and control interfaces in a way that is potentially more powerful that the INAP-based solutions described in this paper.

Conclusions

This paper has presented a system that utilises existing IN infrastructure to support an e-commerce protocol. The clear advantage of this system is the incorporation of existing billing mechanisms, which are accepted by end users. In this way, the controversy surrounding the security of credit card use on the Internet is sidelined.

This protocol nicely fits with the current trend of mergers and acquisitions between traditional PSTN network operators and Internet service providers. It can easily be used to complement the network operator's product portfolio by also providing e-commerce solutions in a manner acceptable by the end users.

References

1 SOLOMONIDES, C.; and SEARLE, M. Relevance of Existing Intelligent Network Infrastructure to the Internet, Lecture Notes in Computer Science 1597, Springer.
2 ITU-T Recommendation Q.121x Series on Intelligent Networks for CS-1, Geneva 1995.
3 Parlay API Specification—Issue 1.01. The Parlay Organisation, 14 Jan. 1999.

Biographies

Christos Solomonides
University College London

Christos Solomonides is working toward his Ph.D. in telecommunications from the Department of Electronic and Electrical Engineering, University College London (UCL). His dissertation concerns the issues surrounding the area of service creation in open networks with multi-vendor, multi-network APIs. He received his B.Eng in Software Engineering from the University of Manchester Institute of Science and Technology (UMIST) and his M.Res. in Telecommunications from UCL.

Mark Searle
University College London

Mark Searle was awarded the degrees of B.Sc. and Ph.D. at the University of Manchester Institute of Science and Technology in 1982 and 1989 respectively. He is currently involved in research and development in the area of telecommunications. He is currently engaged in a number of topics connected with Signalling System No.7, intelligent networks and the relevance of the control plane to the Internet. He is also currently the Director for the Telecommunications for Industry Programme based at UCL.

Manuela Broccoletti and Giuseppe Zilioli

Electronic Commerce is the New Way to do Business

Technology Driven or Regulatory Driven

Global networks could make it as easy to do business with a company on the other side of the world as with one on the next street. But the communication medium alone is far from sufficient. Contractual and financial issues, ownership of goods distributed electronically, privacy and security, still have to be resolved at a global level. After the Global Ministerial-level Ottawa Conference on e-commerce, several specific national and international initiatives are taking place around the world: the Monti Directive of the European Union and the Observatory on Electronic commerce of the Italian Minister of Industry are only two examples of initiatives in which Telecom Italia is involved.

New Service Providers

There is no universally accepted definition of e-commerce so far, but it is clear that it involves commercial transactions, over telecommunications networks, by using electronic means. It includes indirect e-commerce (electronic ordering of tangible goods), as well as direct e-commerce (on-line delivery of intangibles). The transactional nature of the exchange is the key feature, involving agreement to deliver goods, to perform services or to transact intellectual property rights. However, e-commerce should be distinguished from basic telecommunication services and from the simple use of the Internet and of electronic databases to obtain information without the establishment of a commercial transaction.

Indeed, it is not a completely new phenomenon, as companies have exchanged business data for many years using different communication networks. As the Internet is now growing rapidly, e-commerce is no longer limited to big companies, but open to everyone around the world. Driven by this Internet revolution, e-commerce is expanding and undergoing radical changes and presents enormous potential opportunities for consumers and for businesses, particularly small and medium enterprises (SMEs). It will have a considerable impact on Europe's competitiveness in global markets and will offer possibilities for job creation, which is at the top of the European Union agenda.

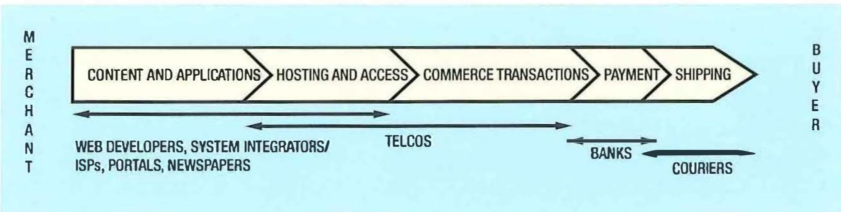
Electronic commerce differs from traditional means. It changes the communications models, which ultimately affect the flow of information, induced by the technology. The reduced cost of communicating, transmitting and processing information demonstrates the advantages over traditional commerce and also allows direct or relationship marketing via Internet at a lower cost than other direct-marketing channels. Improvements in information and communications technologies (ICT) reduce information-processing costs, which deeply affect the organisational structure of the business. Also the roles of the players involved will look different because they all take advantage of a global reach with faster communication at lower costs.

In this scenario telcos will dominate the market for basic access to the Internet in both the business and consumer segments with a progressive consolidation and with few Internet service providers (ISPs). However, telcos will probably have to change becoming, in effect, telco/ISPs.

New entrants will become powerful competitors to the traditional telcos in nearly all-European countries before 2002 (source: Yankee Group), but they will probably not dominate the higher value-added segments of the market. Due to the diffusion of the extranets (an extension of the corporate Intranet to suppliers, partners, and key customers) and e-commerce solutions they still will play a relevant role in this market together with software companies, systems integrators, vertical industry sector suppliers and specialised data network providers.

Figure 1 shows the new value chain which identifies the new

Figure 1 – The new value chain with the new service providers



Manuela Broccoletti, Telecom Italia, DST. Research & Development Viale Europa 190, 00144 Roma.
Tel. +39 06 3687 5702
Email: manuela.broccoletti@telecomitalia.it
Giuseppe Zilioli, Telecom Italia Rue de la Loi 15, 1040 Bruxelles
Tel. + 32 2 282 0640
Email: zilioli@telecomitalia.be

service providers and their roles. Each of these (systems integrators, portals, newspapers, banks and couriers) has unique value propositions, target markets, and existing customer relationships.

Telecommunications operators

There are many opportunities currently available to international carriers (IXCs) and regional telecommunication carriers participating in the e-business scenario. The IXCs are jumping into the market by leveraging relationships with existing clients, primarily large and medium enterprises. They can use their knowledge in developing and implementing integrated network solutions to solve traditional telecommunications network problems.

The other licensed operators (OLOs) and domestic operating companies are also participating in the e-business arena. These companies can use the strong brand awareness in their respective regions to quickly establish an e-business footprint. In addition, many of these use their enhanced Yellow Pages directories to promote their e-business capabilities.

The challenge facing the telecommunications operators is achieving credibility in the e-business market and one way to accomplish this is to enter into partnerships with reputable e-commerce service providers to benefit from their expertise.

Web developers/Internet services providers

There are clear opportunities for ISPs to thrive in today's e-business market who tend to target high-end, the large corporations who are seeking assistance in the implementation of their Internet business solutions. The local and regional ISPs usually target smaller regional organisations.

However, many ISPs will need to broaden their current capabilities to address the additional skills, knowledge, and expertise necessary to successfully design, develop, and implement e-business solutions for their clients. In addition, the ISPs will need to expand their customer service capabilities to support the new e-business requests.

To establish credibility, it will be critical for the ISPs to successfully develop solutions for few high-visibility clients. This will provide experience in this new area—and success stories would showcase their capabilities. Eventually ISPs will provide a full range of e-business services and meet the high-end customers' needs and the needs of smaller organisations. The ISPs may also have to establish partnerships with other companies including web design organisations and system integrators.

Systems integrators

Systems integrators are well positioned to help large enterprise organisations implement Internet solutions, which can affect their businesses positively. The primary strength of these companies is the integration of corporate legacy systems with the Internet to create complex e-business solutions. Systems integrators have extensive experience in developing business process re-engineering, data warehousing, decision-support systems, and systems integration solutions, which they can leverage to support their clients. Since a vast majority of companies have not yet implemented complex Internet business solutions, the demand for systems integration services is currently limited. As more companies recognise the benefits of using e-business solutions and implement complex functionalities, the demand for systems integration capabilities will expand significantly.

Portals

Portals, including America Online (AOL), Yahoo!, Excite, and Lycos, can leverage interactions with their Internet audience and increase their reach by providing web hosting and Internet business solution services. Local portals like Microsoft's Sidewalk and AOL's Digital Cities can support the e-business needs of smaller local organisations. Entering the electronic solutions arena enables these portals to expand their revenue streams beyond their primary advertising revenue base.

Newspapers

In general, newspapers have been slow in taking advantage of their position in the e-business industry. However, they are beginning to leverage their on-line presence by expanding their service offerings into the e-business scenario. The *New York Times*, for example, sponsors the *New York Today* service, which is a micro-site targeting users in the New York region. The service includes a *Daily Calendar*, where businesses can post upcoming promotional activities and events, and receive personalised announcements for events and activities that are important to them.

Businesses are also given the opportunity to include their banner advertisements in relevant sections of *New York Today*, which are linked directly to corporate home pages.

Banks

The entry of banks and financial institutions into the e-business market may appear to be somewhat of a surprise. Some financial organisations, such as Fleet Bank, provide merchant account services to their customers. Fleet Bank's Storefront@fleet product provides (primarily) small business users with a variety of e-commerce related services, including the ability to conduct secures transactions at a low cost.

Most banks, however, have not yet taken advantage of their position in the e-business market.

Couriers

Couriers like FedEx, SDA (an Italian leader in national and international delivery of small parcels and packages since 1994) have the opportunity to close the value chain. They realise the integration of the sale services, offered by the virtual shops, through the delivery of goods directly to customers' homes.

Electronic Commerce: Technology Driven

Broadcast networks have been used in the past specifically for content and the public switched telephone network (PSTN) mainly for voice. E-commerce is creating the impetus to transform infrastructures so that they can handle a variety of services and applications. The growth in demand for network resources created by competitive e-commerce services is expected to be important. High-speed network infrastructures require platforms from three main groups: the public switched telephone digital networks (PSTDN), the broadcasting infrastructures, and the mobile ones (both cellular and satellite networks). These networks, despite differences in architecture, bandwidth, performance and reliability, offer different means of access, which would allow users and service suppliers to choose how they access and use e-commerce. For everyday transactions the three infrastructure groups mentioned above can be viewed as substitutes.

Telecom Italia Group and other operators support all of them, but the bottlenecks are not exclusively technological.

Investments in infrastructure are required, but not at the cost of increasing the price of the services, which should follow the current downward trend to enable competition with the United States. Industry and government should work together and competing infrastructures should be allowed. The United States has already devised an action plan, and allocated $100 million, to develop new infrastructures to cope with the number of users and the technical requirements of the future, offering a faster and more reliable service. The EU needs to launch a similar programme as it cannot be left to the market because it requires long-term and high-risk investments. It needs to rest on partnership between researchers, governments and industry, as well as with other countries, especially the United States.

Interoperability has to be ensured to allow transactions between any operators, and standards should not jeopardise the development of infrastructures or build barriers to the entrance of new providers. The standardisation process, leading to increased interoperability, should be market driven and standards technology neutral so as not to favour one specific solution.

At the moment, there are several possibilities for the creation of a secure payment system. Some are based on traditional cheques and credit cards, others on more innovative and technologically advanced systems which concentrate on small payments and ways to remain anonimous (Digicash, Cybercash, Mondex, etc.). Several payment systems are at an experimental phase, developed by academic institutions and do not have extensive or significant test marketing presence. Most of the payments currently in use are based on traditional credit cards and more security-oriented clients tell only their credit card number to sellers via telephone or fax.

It is expected that the development and use of smart cards and e-money will make payments on Internet easy and secure, building the necessary confidence among users. At the moment it is not yet clear who will be allowed to issue electronic money, the banks only or other institutions, and how it will be controlled. EU Directives on distance selling of financial services and on electronic payments should clarify this matter.

Terminals are the initial interface for users (residential and business) to access e-commerce through the different infrastructure platforms. High-speed local loops and sophisticated applications are of little value to users if they have no means to access them or if the technical capacity of their terminals is much less than required by applications, and if terminal prices are high. Ease of use for terminals is an important factor. The importance of the personal computer (PC) as an e-commerce terminal is known but, despite the installed base being relatively high in the United States and slightly less in Europe, residential customers can find the price and complexity of PCs a barrier, even though price/performance ratios have recently shown improvements. The potential of televisions and mobile telephones as e-commerce interfaces for residential customers have to be considered given their worldwide penetration (56·9% and 11% respectively in the Organisation for Economic Development (OECD) area in 1996). But television is a receive-only terminal so that interactive applications need terminal adjustments, while at the moment mobile communication has a potential, despite the present limitations on the speed of data services. E-mail applications and home-banking transactions require limited bandwidth and may become very popular among mobile users. Further, the GSM smart cards system offers a very good platform for e-commerce because it is easy to use and supports a large number of standardised services, such as security.

Electronic Commerce: Market and Regulatory Driven

Serveral questions need to be clarified in the on-line environment. It is not clear whether a commercial web site should be treated like advertising, direct marketing or simply a shop window. On the basis of work already undertaken in this area (the Commission Green Paper on Commercial Communications and the EU Parliament's resolution on it) these questions should be addressed within the framework of the forthcoming Commission Communication on this issue.

The liberalisation of telecommunications is crucial to the further development of e-commerce because it will increase competition and ensure lower prices. This is important, as price levels now are higher in Europe than in the US, our main competitor in e-commerce today.

The conclusion of electronic contracts in a networked environment brings about a number of important questions to be addressed in order to facilitate e-commerce cross-border transactions. For instance, the determination of where and when an electronic contract is concluded and which country's law is applicable could be addressed differently by the Member States. These questions, which go beyond the legal recognition of digital signatures, should be clarified in order to facilitate electronic contracting within the EU (Monti's Directive Proposal).

The issue of liability is often raised in the copyright context and there is no doubt about the horizontal nature of this problem. The liability of Internet service and access providers may arise in multiple areas: copyright, trademarks, misleading advertising, protection of personal data, product liability, obscene content, hate speech, etc.

It is felt that no new taxes should be imposed on e-commerce, and that a *bit tax* would negatively affect e-commerce. Taxation issues together with tax neutrality should be pursued. EU is developing ways of adapting VAT to the particular characteristics of e-commerce, in order to ensure equal taxation of all services whose consumption takes place in the Community. Regarding direct taxation, issues like compliance, the characterisation of income (royalties or not), residency, source rules and the concept of *permanent establishment* are being considered.

The Ottawa Conference

The Organisation for Economic Co-operation and Development (OECD) and the Government of Canada held a Ministerial Conference on 'E-commerce' in Ottawa, 7–9 October 1998, to reflect on the need to create a global framework to harness the economic potential of e-commerce and to ensure its continued growth in a socially responsible manner. Building on the results of a previous OECD Conference held in Turku, Finland, in November 1997 and on the European Ministerial Conference in Bonn, Germany, in July 1997, governments, the private sector, and social interests met in Ottawa to advance their common commitment to create a certain, secure and predictable international environment for the conduct of e-commerce.

The Ottawa Conference took specific steps to establish the cornerstones for the future growth of transnational e-commerce. Government representatives and business, labour and non-government organisations' leaders considered the economic and social potential of e-commerce, the barriers that currently limit this potential, and solutions that may be employed. They also identified the following elements towards a shared vision for global e-commerce:

- *Building trusts for users and consumers* Ensuring that the frameworks and safeguards which provide confidence in the physical marketplace are adjusted to instil equivalent confidence in the digital marketplace.
- *Establishing ground rules* Providing protection into this new platform which is as effective as that provided by legal and commercial frameworks in the physical world.
- *Enhancing the information infrastructure* Creating effective competition in telecommunications markets can ensure a long-term trend towards lower costs increased quality and expanded access to information infrastructures and services. The growth of global e-commerce relies on universal and affordable access to the information infrastructure.
- *Maximising the benefits* Understanding the social and economic impacts of e-commerce is critical for the transition to a digital economy and towards a global information society. All players must contribute to understand and address the impacts on growth, productivity, and employment, and needs of business, including small- and medium-sized enterprises, in both developing and developed countries.

The Conference identified the challenges implied by maximising the benefits of global e-commerce, clarified roles, responsibilities and commitments in terms of an action plan and outlined for the first time who is doing what to solve these problems.

The Ottawa Conference represented an important milestone in establishing a plan of action for the groups participating in the Conference, but its ultimate success will be measured by the degree to which all parties will meet their respective commitments to future action (see www.ottawaoecdconference.com).

The Commission Initiative on Electronic Commerce

The aim of the Commission Communication 'An European Initiative in Electronic Commerce' is to encourage the growth of e-commerce in Europe. Building upon the Commission's work to date, it provides a policy framework for future Community action, and aims at establishing a common European position to achieve global consensus through international negotiations.

After Ottawa it is recognised that the coordination at the EU level is essential in order to avoid fragmentation of the internal market and to establish an appropriate European regulatory framework and a common and strong negotiation position. At the moment the Commission, and the EU parliament, are discussing proposals and elaborating a road-map for the missing legal issues in the area of e-commerce (such as liability, definition of the place of establishment for service providers in the on-line environment, commercial communications, regulated professions, patents, trademarks and domain names) to clarify the regulatory framework for these issues and to safeguard the rights of the users of e-commerce.

Besides, the current trend is that there is a process of technological convergence going on in the telecommunications, audio-visual and information technology sectors. Through this process a number of media, previously attached to a specific hardware, are now available in a multiplicity of platforms. There is also convergence going on in another sense, as e-commerce has a great cohesion potential. Business location is not as important anymore, so rural or remote regions could compete on equal terms. Urban congestion problems could also diminish. However, the development of the infrastructure necessary for e-commerce in the EU varies greatly between countries and regions. To ensure access to e-commerce anywhere, support programmes and funds together with the inclusion of such issues in the framework of the structural funds are desirable.

Also, an open competition policy in the provision of facilities, products and services is a major goal. The state of the competition situation and the barriers to entry should be assessed. International cooperation concerning competition matters should be implemented.

The MICA Observatory

Following the Ottawa Conference, the Italian Ministry of Industry has recently instituted an Observatory's Experts Committee on Electronic Commerce. Telecom Italia and the main Italian companies are among the members of this initiative.

Several events are in progress in the major Italian towns to address industry and commerce associations and to make operational the joint effort between the Ministry and the industry to stimulate the market.

Specific initiatives have been organised for the public administration and SMEs (including workshops and promotional campaigns).

The role of the Observatory is to study, globally, the policy and the themes of the information society and will work with the main players of the market (Confindustria, Confcommercio, ICT Forum, etc.), other Italian national institutions (Presidenza del Consiglio dei Ministri, Ministeri del Commercio con l'estero, delle Comunicazioni, AIPA, Trade Union, etc.) and international ones (European Commission, OCSE, etc.).

Telecom Italia Offer

Telecom Italia has put forward a vast range of innovative on-line services for electronic trading known as *Village Commerce*. It is integrated within Telecom Italia's Village platform and offers a complete range of services in partnership with information technology suppliers who are interested in developing this business with Telecom Italia. This gives companies wishing to create a virtual shop-window on the Internet

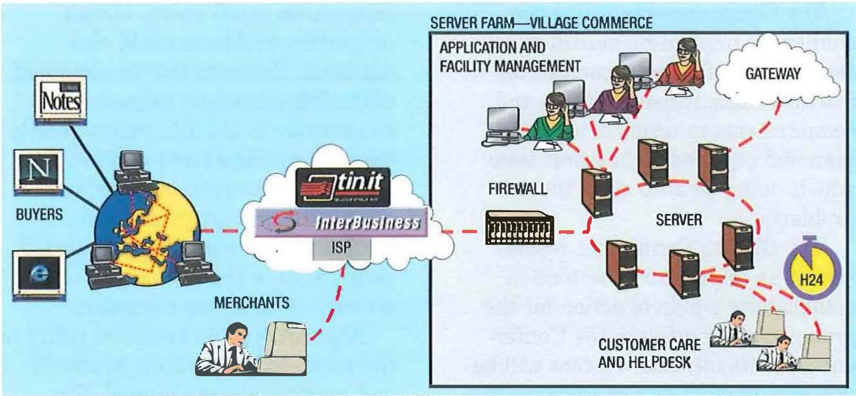


Figure 2 – The architecture of Village Commerce

the chance to produce their own site and catalogue for the presentation of goods to sell on-line to others companies or private individuals. It is integrated with a system for the management of ordered goods with accounting procedures for orders, the acquisition of customer data, payment, the setting-up of a customer database and total integration with the merchant's computer systems (Figure 2).

Village Service offers over 40 applications, for cross-industry and industry-specific solutions, developed, selected and certified by Telecom Italia business partners; such as:

- projects and business consulting;
- site building and management;
- e-commerce service activation;
- pre-sale support services;
- transaction and order management;
- fulfilment management;
- post-sale support services; and
- customer services.

Bibliography

1 The Yankee Group Report—Internet Marketing Strategies. Vol. 5, Feb. 99.
2 Intermediation and Electronic Commerce, DST/ICCP/IE(97)6.
3 Linee di politica per il Commercio Elettronico, 30 luglio '98, Ministero dell'Industria, del Commercio e dell'Artigianato, MICA.
4 Overview of the OECD work bearing on e-commerce, Jul. 97.
5 Global information infrastructure and global information society, May 97.
6 http://www.village.telecomitalia.it di Telecom Italia Spa.

Biographies

Manuela Broccoletti
Telecom Italia

Manuela Broccoletti gained her degree in Electronic Engineering at the Rome University 'La Sapienza' in 1988. She attended the Specialisation Course in Telecommunications at the SSGRR School in L'Aquila. She has worked in Telecom Italia since 1988. After six years in the Network Division, where she was involved in ISDN and ATM technological deployments, she is now working on the business services innovation in research and development. She attends international ETNO and Italian business/ministerial working groups on e-commerce as the Telecom Italia delegate.

Giuseppe Zilioli
Telecom Italia

Giuseppe Zilioli gained his degree in Physics at Milan University in 1966. He has spent over 20 years in international IT companies in the United States and Europe with technical and marketing responsibility. Since 1991 he has worked at the Telecom Italia's representative office for the European Union of in Brussels, where he is involved in the Research Framework Programme and in the business development initiatives financed by the European Union.

Paschalis Bouchoris

The E-Commerce Challenge: Strategic Implications For Enterprises

Electronic commerce is a serious strategic challenge to businesses. As the e-commerce paradigm prevails, enterprises, key among them European telecommunications operators, have to understand how they can enhance value and gain competitive advantages through e-commerce. This paper puts the current state of e-commerce development into some context and sizes up its short-term future potential. It discusses the recent phenomenon of very high market capitalisations of e-commerce ventures and analyses the prime competitive strategy implications arising from the introduction of e-commerce. In brief, it identifies three ways in which enterprises can benefit from e-commerce: cost reduction, differentiation and new market development. It analyses the impact of e-commerce throughout a company's value chain: inbound logistics, production, sales, marketing, outbound logistics, distribution, post-sales support, staffing etc. Finally, the paper focuses on the development of e-commerce in Europe and the reasons it is lagging behind development in the United States marketplace.

Paschalis Bouchoris, European Dynamics S.A., 24 Ebrou Street, Athens 115 27, Greece
Tel: +30 1 6987100
Fax: +30 1 6987105
E-mail: pbouchoris@hotmail.com

An Overview of the E-Commerce Market

Although the skyrocketing growth of electronic commerce from basically zero level of revenues in 1995 is an undeniable fact, there is little consensus among research firms regarding projected e-commerce revenues in the near to mid-term future. According to a recent report by OECD, electronic commerce worldwide is predicted to reach $330 billion in 2001–2 and $1 trillion in 2003–5, rising from a mere $26 billion in 1997. At these levels, in 2001–2 e-commerce will represent three times US catalogue sales, 24% of US credit card purchases and 18% of US direct marketing sales.

Analysing the e-commerce market entails a differentiation between the following segments:

Business-to-business

Business-to-business electronic commerce, a field which has existed for long in the form of electronic data interchange (EDI) over value-added networks and has now evolved adopting Internet protocols and encompassing intranets, extranets, etc. This is a relatively well-developed segment currently representing the bulk of all electronic commerce. Analysts estimate that business-to-business trade may already exceed $150 billion.

Business-to-consumer

Business-to-consumer e-commerce, which can be further segmented by main category of tangible and intangible products as follows:

- *tangible*: Computers/electronics, books, clothing, food/drink; and
- *intangible*: entertainment, software, travel, newspapers/ magazines, financial services, e-mail.

To date, the largest segments of e-commerce activity involve intangibles (most importantly entertainment, software, financial services).

In particular, financial services (stock trading, banking, insurance) have seen an explosive growth on the net over the last year. Nowadays, one out of seven share trades are taking place on the Internet with 6·3 million households, in the US alone, trading their shares over the Internet.

Internet discount brokerage houses were the first to exploit the advantage of the Internet and, as a result, have now excelled to a position to challenge established financial institutions by offering the full spectrum of financial services. Furthermore they are making plans to establish their own markets that could soon become competitors to the existing stock exchanges.

One of the tangible product categories generating a lot of sales on the net is computers but other products are seeing their Internet sales increasing rapidly too.

There are many factors that have proven to influence growth in business-to-consumer electronic commerce. A non-exhaustive list of the main drivers includes convenience, personalisation, choice, amusement and savings, while legal and psychological barriers are the most common inhibitors.

Competitive Strategies for E-Commerce

The emergence of e-commerce has serious implications on the competitive strategy of firms. E-commerce has been used as a strategic weapon in order to gain competitive advantage either through cost reductions or through differentiation. New, non-industry entrants have appeared in many sectors due to e-commerce while existing players have used it to diversify their product offerings, and enter new markets (usually addressing the global market). In this pursuit of a global market, localisation has remained important and as a result a combination of global services highly customised to the needs of local/niche markets have prevailed. As typical in a market which is at its introductory stage, product innovation has been the basis for initial entry and early success. However, as the sector has evolved and competition intensified other requirements for success have emerged: investment/financial resources, marketing, advertising and distribution strengths (as in the case of Netscape versus Microsoft). Thus, many actors have moved from introducing and exploiting their own technological standard to competing on issues such as acquiring customer information, exploiting first mover advantage to become niche/category leaders, or building strategic alliances.

In the paragraphs below, some of these competitive strategy issues are analysed in more detail.

From a company's standpoint the main strategy motives for doing business on-line seem to be:

Reduce costs

E-commerce technologies can be used throughout the value chain to reduce costs. More specifically e-commerce has proven to be a significant cost saver in the following activities:

Purchasing of production inputs

Companies using EDI over value-added networks (VANs) have been reported to save commonly 5–10% in procurement costs. This cost saving is even higher with Internet technologies, given their considerably lower investment and operating expense.

Inventory

The long-time 'zero inventory' target of production managers seems to be finally approachable with the use of e-commerce technologies. These technologies have enabled an unprecedented level of integration between marketing and production operations and abundant information on detailed customer needs and purchasing habits to be used for inventory management. Lower inventory means lower materials handling, warehousing and general administrative costs and more efficient utilisation of production capacity with subsequent reductions in the need for plant and equipment. For intangible content-based products (for example, newspapers and magazines) the inventory is completely eliminated as the paper-based editions are being replaced by 'bits-and-bytes' sent upon request.

Production

Again, the intangible products are the ones benefited more as e-commerce offers unparalleled economies of scale. Increases in output do not require proportionate increases in input. An e-commerce product can be profitable no matter how many units are produced—there is no extra unit cost while there is ample possibility for customisation addressing the individual consumer needs at no extra cost. In addition, manufacturers of tangible products can reduce their production costs by exploiting location advantages as they source production inputs from the lowest cost location worldwide in a seamless way. Through intranet/extranet applications they can link together resources in the model of a *virtual corporation* and spread throughout a possibly global organisation the costs of expertise available at even the most remote location, thus generating rare economies of scale.

Sales

This is an activity at which electronic commerce has already proven its efficiency and effectiveness. As the Internet users' base expands, enterprises can exploit their relatively much lower cost Web store to complement or even replace the real world alternatives (physical store, sales representatives visits, print catalogues, telephone sales, teleshopping). In the advantages of a Web-based point of sales one should add the little or no additional cost to add a new customer, the potential to pursue a global clientele at relatively lower costs than compared to the physical world alternatives, as well as the extra efficiency in order placement and execution.

Marketing/advertising

As with the sales activity, marketing has benefited a lot from the introduction of the Web due to the superior value-for-money it offers. Most importantly it is a unique means of collecting marketing information, executing highly-targeted promotions, or interacting with the customers in the direction of one-to-one marketing.

Product distribution

The potential for cutting distribution costs through improved coordination/organisation (particularly for intangible goods) is already being explioted by many organisations throughout the world. Furthermore, the potential of e-commerce to alter the cost structure of this activity is responsible for many ventures that have succeeded in transforming the industry structure by introducing a radical 'dis-intermediation' which has outplaced intermediaries such as traditional wholesalers or retailers.

Customer support/after-sales service

Companies such as Dell and Cisco have reported an increase in their customer support productivity by 200%–300% and savings of over $100 million in customer service costs. According to Forrester Research data, it generally costs $500–$700 to send a service representative into the field, $15–$20 to handle the request over the telephone, and about $7 to do the same over the Internet. As an indication of the cost-savings potential, Federal Express has reported that its on-line customer service system represented savings of about 20 000 new hires—almost 14% of its total staff. (Source: US Department of Commerce report.)

Personnel

The introduction of e-commerce technologies can reduce or eliminate organisational slack and improve efficiency in traditionally labour intensive activities such as sales, inventory management, customer service etc. E-commerce ventures require in general far fewer, but better-skilled staff. As an indication, Amazon.com the largest Internet bookstore has less than 700 employees while Barnes & Nobles, the largest physical US bookstore more than 27 000!

Furthermore, the introduction of e-commerce may alter a firm's cost structure in particular with regard to its linkages to other actors, most commonly wholesalers, retailers, subcontractors, raw material suppliers etc.

Differentiate

The Web offers numerous possibilities for differentiation. It enables product expansion either through new product features or through the combination of related product/ service offerings. It offers ample opportunities for new product development. Most retailers on-line have soon discovered the potential to expand their role as merchants to community organisers attracting an audience of existing and potential clients to their site which becomes more of a forum where trading operations are blended with content, after-sales services and opportunities to interact with producers and other product users. Service providers such as banks and brokerage houses have been able to add ancillary services in a way that could not be done in the physical world.

Perceived differentiation is also an important contribution of the Web, with customers seeking more convenience, personalisation and choice. However, with technology issues being mastered by more and more actors and competition intensifying, image differentiation requires increasingly more resources.

The ability to differentiate through the Web has been an opportunity for new industry entrants or smaller participants to gain competitive advantage.

This was the case for Amazon.com, and for auction sites that became leading e-commerce actors winning traditional players.

Enter new markets

Global by nature, e-commerce has enabled many commercial enterprises to overcome the boundaries of their limited geographic markets and expand worldwide. Small companies located in obscure areas of the world have managed to create a global brand and, usually through alliances with global distribution mechanisms, sell their wares worldwide. It should be noted that with technology not being an entry barrier anymore, advertising and brand development becomes the most significant obstacle. Another significant barrier is overcoming consumers' security considerations when involved in commercial transactions on the net and, thus, the need to build the necessary level of trust.

Other strategic implications

The great importance of strategic alliances has been profound in the e-commerce era.

Banks are allying with technology providers to develop facilities for secure transactions, media and retailers are joining forces to attract advertising revenues, groups of commercial sites are syndicating *banner space* to increase their bargaining power with advertisers and agencies, producers and distributors are collaborating to replace intermediaries and increase their margins etc.

The prime motives for strategic alliances in the e-commerce arena seem to be: economies of scale (in advertising, research and development, product development, etc.), access to resources and capabilities not available internally, improvement of margins through restructuring of the value chain.

Despite intense activity in both directions (forward and backward) of vertical integration, there is no common understanding yet as to the optimal degree of internalisation of transaction costs. At the same time that some actors opt for vertical integration others transform to virtual corporations where the primary function of the company is coordinating the activities of a network of suppliers while owning only one activity in the value chain (such as production, distribution, brand/advertising and sales).

A final note on the competitive strategy implications of the electronic commerce is the intense mergers and acquisitions (M&As) activity in the e-commerce sector.

According to Securities Data Corp. the biggest 1998 Net M&As were: AOL/Netscape ($4billion), Ticketmaster/CitySearch ($700 million), Disney/Infoseek ($465 million), AOL/ Mirabilis ($287 million), Verio/Hiway Technologies ($256 million), Microsoft/LinkExchange ($250 million), Cisco/Nextspeed ($233 million), Amazon/Junglee ($173 million) and Lycos/WhoWhere ($128 million).

There is a clear trend away from the *fragmentation* at the early stage of the market's development and towards an *oligopolistic structure.* This trend is indicated by the high concentration of revenues in the top 3–4 players in every product category and is reinforced by the increasing difficulty to compete for the consumers 'mindshare' in highly congested sectors.

E-Commerce Firm Valuations

The Internet stock fad of 1998–1999 is clearly the most extreme stock market frenzy in the history of global financial markets. The investors' enthusiasm about almost any company with a '.com' in its name is unprecedented even compared to other infamous periods of overheated stock markets such as the tulip bulb craze in 17th century Holland, the 19th century railways stock surge, the 1980s software mania etc. A series of phenomenally successful initial public offerings (IPOs) of Internet-related companies, the stock price explosion of listed e-commerce companies, and a rapidly growing number of individual shareholders using the net and electronic brokerages to trade their own shares are aspects of this phenomenon.

Statistics are revealing: In 1998, Amazon's share price rose 638% reaching a market capitalisation of $18·9 billion, more than six times that of Barnes & Noble, the world's biggest bookseller. Amazon is not expecting profits before 2004 and it is not the exception in a financial world that is not scared by the losses of e-commerce/Internet related ventures. Other high-profile stock market successes of the e-commerce sector are: Yahoo who saw its share rising 452% to a market capitalisation of $11·7 billion in 1998, AOL 294% (market cap. $40.9 billion), eBay 252% (market cap. $7·6 billion), Excite 227% (market cap. $2·5 billion) and Lycos 153% (market cap. $2·3 billion) *(source: Yahoo Finance).* Reflecting the success of its on-line trading site, the market capitalisation of Charles Schwab, the discount brokerage, had exceeded that of the financial conglomerate of Merill Lynch by the end of 1998!

Highly reputable US venture capitals and investment banks have secured hundreds of millions of dollars venture financing for e-commerce companies with non-existent profits and sometimes even revenues! As reported by the magazine *The Industry Standard* (Jan 1999), StarMedia, a Spanish-language portal for the Latin American market, has raised $80 million, Buy.com (an Internet department store) $60 million, Priceline.com (an auction site) $55 million, Realtor.com, (a real estate site) $50 million and Talk City (a network of Net chat sites) $34 million.

Even the most faithful believers in the potential of e-commerce market cannot explain current valuations at these levels. In the absence of a track record of revenues, visible future profits, hard assets, or even (sometimes) experienced management teams, what drives these unprecedented valuations remains an enigma—according to conventional analysis.

There are many people who tend to believe that a new set of net-economy tools should replace mainstream economic analysis when assessing investments in the e-commerce sector.

E-Commerce in Europe and Implications for European Telecommunications Operators

Europe is clearly lagging behind in e-commerce activity. Sales estimates suggest that US is credited with about 80% of the total e-commerce activity with a potential decline to 65%–75% by 2002. With the possible exception of Scandinavia, the adoption of e-commerce practices is still sluggish in Europe.

IDC forecasts that Western European revenues from e-commerce will rise from 900 million ECU in 1997 to 26 billion ECUs in 2001. According to another estimate by Datamonitor, on-line shopping at European Web sites will rise from ECU 95 billion in 1997 to ECU 4·3 billion by 2002.

The key factors still limiting the growth of e-commerce in Europe seem to be the following:

- bandwidth comes at relatively higher cost and lower quality/ capacity;
- liberalisation of the telecommunications sector is lagging behind;
- European consumers are not used to buy through mail order/ catalogue/ distance schemes;
- penetration of PCs/modems in households and enterprises is lower;
- societal awareness about the benefits of e-commerce is lower;
- European national markets are still relatively fragmented;
- European citizens are not comfortable entering cross-border commercial transactions by distance; and
- linguistic diversity becomes a barrier in developing content which can attract communities of users and stimulate e-commerce activities.

However, despite a slow start, Europe is catching up fast. According to a survey presented in the 1999 report of the European Information Technology Observatory (EITO) most European businesses believe that electronic commerce will become the norm for sales, post-sales, purchasing and marketing functions before the end of 2001. This fast growth is also demonstrated by research data which indicate that at the end of 1998 29% of European business were using Internet-based e-commerce applications—up from only 6% in 1996—while, by the end of 1999 an estimated 47% of European firms will be using e-commerce.

The state of e-commerce in Europe and, more importantly the mid-term outlook of its growth present very significant implications to European telecommunications operators. This impact can be analysed as follows:

Telecommunications companies as enablers of e-commerce

Telecommunications operators as infrastructure providers (access, backbone etc.) are key enablers of e-commerce.

The continuing expansion of the Internet and the diffusion of e-commerce practices suggest a serious need for an upgrade of the European telecommunications infrastructure including both access and bandwidth resources. With more than 45 million hosts and more than 30 000 Internet service providers operating worldwide, costs and capacity bottlenecks in European local loop, switching/routing and backbone infrastructures may inhibit the growth of electronic commerce unless a lot of additional investment takes place in the next 2–3 years.

When estimating the viability of their investments in this area and taking into account the above analysis, telecommunications operators should take for granted the insatiable demand for bandwidth-intensive services and commit resources to the expansion of their basic telecommunications infrastructure in Europe.

Telecommunication operators as e-commerce industry participants

European telecommunications operators are also actors themselves in the e-commerce sector. In most European countries, Internet service providers owned or controlled by incumbent PTTs are holding leading positions in their domestic Internet markets. These ISPs should build and utilise strategic alliances with other actors as described previously to generate more value as pivotal players in the e-commerce market. Taking advantage of the financial markets' endorsement of e-commerce ventures, ISPs may reach significant market capitalisations enhancing shareholder value for controlling telecommunications operators.

Telecommunications operators as adopters of e-commerce facilities

Telecommunications operators should learn from the competitive strategies for e-commerce described above to re-engineer their own operations, saving costs throughout their value chain. In an increasingly commoditised market in which telecommunications operators fight against churn, e-commerce offers them opportunities to differentiate themselves, deliver more and enhanced services to their customers and excel to one-stop centres for their clients. Accessing detailed, customised billing information over the Web, paying telecommunication bills electronically, offering customised services to businesses and niche audiences are indicative examples in this direction.

Biography

Paschalis Bouchoris
European Dynamics S.A.

Paschalis Bouchoris is an Assistant Managing Director at European Dynamics S.A., a fast-growing European systems integrator headquartered in Athens, Greece. The company's affiliate, Telecom Dynamics, is Greece's leading private telecommunications operator developing a state-of-the-art alternative telecommunications network throughout Southern Europe. He is an electrical and computer engineer and an MBA graduate from Rotterdam School of Management, the Netherlands. He has also worked as a management consultant and industry analyst. His main areas of professional expertise are: business strategy, investments and international business development.

Welcome to the New Telecoms World

'The New Telecoms World'. This is the world where Ericsson operates. An exciting working environment where everyone is on top of the latest developments. Dynamic and with unrivalled growth prospects. There are more than 100,000 people in Ericsson worldwide working every day on the integration of the latest developments in telecommunications and information technology. Ericsson has particularly high expectations for mobile applications in the near term.

Ericsson is a global player with decades of experience in the ICT world and mobile communications. Ericsson is one of the top three international mobile telephone suppliers. And it is no accident that Ericsson has a leading position in the most magnetic market of the coming decade.

'A world of new opportunities'

People are the difference

Maintaining and strengthening this leading position depend largely on the people working for Ericsson. In spite of the hi-tech character of the working environment, the employees are the differentiating factor at Ericsson. Enthusiasm, passion and professionalism are needed for supplying innovative products and services. So too is a modern business environment where individual initiative is continuously encouraged and promoted. Independence is therefore an essential trait of our people. This is also why Ericsson staff direct the development of their own careers and skills themselves to a large extent.

Ericsson's activities in the Netherlands

Ericsson has four locations in the Netherlands. The largest facility is in Rijen, Brabant, where approximately 2500 employees meet the commercial needs of the telecoms and other companies located in the Netherlands. Ericsson Rijen also plays a leading part in the worldwide R&D activities, consultancy and service support to national and international customers. There are 250 staff in Enschede working on large, globally significant R&D projects that are vital for the future. The same is true of the location in Emmen, where crucial R&D projects are also being worked on, including Bluetooth – a joint venture with other major international companies including IBM, Toshiba, Intel and Nokia. A special unit in Amsterdam takes care of the international marketing and sales of particular product groups.

New colleagues

Ericsson in the Netherlands has grown very rapidly in recent years. Future prospects are positive. There will therefore be an ongoing need for new colleagues in the next few years. People with a background in information technology, telecommunications or related fields are particularly welcome.

Interested?
Call the Search & Selection Department or take a look on the Internet.

Ericsson Nederland
Search & Selection
P.O.Box 8, 5120 AA Rijen,
Netherlands
Telephone 0031 (0)161 24 24 39
Fax 0031 (0)161 24 25 92
http://career.ericsson.nl

ERICSSON

BRITISH TELECOMMUNICATIONS ENGINEERING

ISSN 0262-401X

Published in April, July, October and January by The Institution of British Telecommunications Engineers, Post Point 2D05, The Angel Centre, 403 St. John St., London, EC1V 4PL.

Formerly *The Post Office Electrical Engineers' Journal* Vols. 1–74: April 1908–January 1982.)

British Telecommunications Engineering and the *Structured Information Programme* are printed in Great Britain by Headley Brothers Ltd.

Subscriptions and Back Numbers
Annual subscriptions (including postage)—companies, universities, libraries and other bodies £60·00 UK, £65·00 overseas; private individuals £36·00 UK, £41·00 overseas.

Single copies and back numbers—companies etc. £15·00 including postage UK, £16·25 including postage overseas; private individuals £9·00 including postage UK, £10·25 including postage overseas.

Overseas customers paying by cheque should use a sterling cheque drawn on a London bank, a sterling Eurocheque or a sterling travellers cheque payable in London.

Orders can be paid by credit card by calling +44 (0)20 7843 7626.

Orders by post should be addressed to *British Telecommunications Engineering Journal* (Sales), PP: 2D05, The Angel Centre, 403 St. John St., London EC1V 4PL.

Cheques should be made payable to '*IBTE*' and should be crossed '& Co'.

An index of recent back issues is on http://www.ibte.org/

IBTE Members requiring back issues should contact the IBTE Office on +44 (0)20 7843 7626.

Advertisements
All enquiries relating to advertisement space reservations should be addressed to The Advertisement Manager, *British Telecommunications Engineering Journal*, PP: 2D05, The Angel Centre, 403 St. John St., London EC1V 4PL. (Telephone: +44 (0)20 7843 7626. Fax: +44 (0)20 7843 7888.)

Communications
All communications should be addressed to the Editorial Office, *British Telecommunications Engineering Journal*, PP: 2D05, The Angel Centre, 403 St. John St., London EC1V 4PL. (Telephone: +44 (0)20 7843 7622. Fax: +44 (0)20 7843 7888.)

THE INSTITUTION OF BRITISH TELECOMMUNICATIONS ENGINEERS

IBTE Local Contacts

The following is a list of local contacts to whom enquiries about IBTE membership and services can be directed.

Zone / Centre	*Contact*	*Telephone Number*
London	Carl Alderman	(020) 8666 5956
Martlesham Heath	Bryan Allman	(01473) 605696
Midlands	Dave Plumb	(024) 7655 6034
North East	Geoff Jenkinson	(01226) 248141
North West	David White (in Manchester)	(0161) 600 2468
Northern Home Counties	Keith Setchell	(01234) 277937
Northern Ireland	David White (in Belfast)	(028) 9021 5354
Paris	IBTE Office	+44 20 7843 7622
Scotland	Cameron Scott	(0131) 668 5804
South West	Tim Jeanes	(0117) 920 5481
Southern Home Counties	John Dymott	(01202) 554001
Wales	Mike Thomas	(01492) 538103

IBTE HelpLine: 0800 028 0209

IBTE On-Line (Internet): http://www.ibte.org/

(BT intranet): http://networks.intra.bt.com/ibte_nat/

BRITISH TELECOMMUNICATIONS ENGINEERING

Journal Subscription Order Form

PLEASE COMPLETE THE FORM IN BLOCK CAPITALS AND RETURN IT, WITH THE APPROPRIATE SUBSCRIPTION PAYMENT, TO THE FOLLOWING ADDRESS:

The IBTE Administration Manager
Post Point 2D05, The Angel Centre,
403 St. John St.,
London EC1V 4PL
United Kingdom

Annual subscription rates (four issues, including postage and packaging) are as follows:

Companies, universities, libraries and other bodies: **£60·00 UK; £65·00 Overseas**
Private individuals (to be paid for by personal cheque): **£36·00 UK; £41·00 Overseas**

Please supply four quarterly issues of *British Telecommunications Engineering*. I enclose a cheque/postal order to cover the yearly subscription as given above.

(Cheques and postal orders, payable to 'IBTE', should be enclosed with the order. Overseas customers should pay a sterling cheque drawn on a London bank, a sterling Eurocheque or a sterling travellers cheque payable in London. Customers may wish to use a local subscription agent. Customers wishing to pay by credit/debit card should call +44 (0)20 7843 7626.)

Title (Mr/Mrs/Ms etc) ____ Initials ______ Surname ____________________

Mailing Address ____

____

____

Postcode (Zip Code) ______________ Country ______________

Signature ____

Date ____

Please state the issue you wish your subscription to commence: ______________
(April, July, October or January)

VAT Registration Number (zero rate): 243 5917 54 *Photocopies of this form are acceptable.*

VOLUME 18 PART 2 SPECIAL AUGUST 1999 ISSUE

Contents

The Journal of The Institution of British Telecommunications Engineers