Ericsson 1/2001

REVIEW

THE TELECOMMUNICATIONS TECHNOLOGY JOURNAL

Wireless-Trench

More power, smaller sources

AXE 810—The evolution continues

Open architecture in the core of AXE

Cable modems—Broadband highway to the home

The R380s smartphone

ERICSSON

Ericsson
REVIEW

THE TELECOMMUNICATIONS TECHNOLOGY JOURNAL

The purpose of Ericsson Review is to report on the research, development and production achievements made in telecommunications technology at Ericsson. The journal is published in English and Spanish and distributed quarterly to readers in more than 130 countries.

Address:
Telefonaktiebolaget LM Ericsson
SE-126 25 Stockholm, Sweden

Phone: +46 8 719 00 00

Fax: +46 8 681 27 10

Internet:
http://www.ericsson.com/review

New subscriptions:
http://www.ericsson.com/review

Address changes and distribution:
Fax: +46 8 681 27 10
birgitta.blom@lme.ericsson.se

Publisher: Jan Uddenfeldt

Editor: Eric Peterson

Production manager: Eva Karlstein
eva.karlstein@lme.ericsson.se

Layout: Paues Media, Stockholm

Illustrations: Claes-Göran Andersson, Hara

Printer: Ljungföretagen, Örebro

ISSN: 0014-0171

Volume: 78, 2001

Cover: The new Wireless Trench™ technology makes it possible to operate a power amplifier without any bonding wires to ground (bringing a new meaning to the term *wireless*!), which translates into considerable savings (less pad area, simplified bonding, smaller and cheaper package) without degradation of performance. What is more, it opens the way for the use of silicon in wireless integrated power amplifiers for mobile terminals.

Contents

Wireless-Trench technology for portable wireless applications

Until recently, the use of silicon for the production of wireless integrated power amplifiers has not been viable. Now, however, the revolutionary new Wireless-Trench technology makes this possible. Ericsson is currently developing a family of related power amplifier products for GSM 900 and 1800/1900MHz as well as GSM dual/triple-band, PDC, other cordless systems, and Bluetooth.

Page 6

AXE 810—The evolution continues

Thanks to an architecture that was developed to support change, the AXE exchange continues to evolve. Recent adaptations of hardware and software prepare AXE for the next generation of networks. The new system architecture will serve as the basis for migration toward a server-gateway architecture for third-generation mobile and multiservice networks. **Page 10**

Open architecture in the core of AXE

Having been based on industry-standard microprocessors, the APZ 212 40 central processor is a milestone in Ericsson hardware implementation. The latest in a line of central processors, the APZ 212 40 is the first of a new generation, and the platform for future multiprocessor solutions. **Page 24**

Cable modems—Broadband highway to the home

Cable operators are eyeing a huge business opportunity in bringing high-speed Internet access into customer homes. The market for cable modems is expected to increase dramatically over the next few years. This increase will be the result of rising user demands for increased speed and access to services.

Page 32

The R380s—The first smartphone from the Ericsson–Symbian partnership

The R380s, which is the first smartphone on the market to incorporate Symbian's EPOC32 operating system, opens up the future of devices by matching hardware and software to give a full range of desirable functions in a user-friendly, efficient, and portable package. The R380s points the way to a community of devices whose common architecture and open platform allow user requirements to be met by specialized developers. **Page 44**

Contributors

In this issue

Steve Bridges

David LeCorney

Ted Johansson

Mats Jonback

Steve Bridges joined Ericsson in 1997, after having worked many years in the aerospace, environmental controls and personal computer industries. He was the object leader on the R380 development project and served as the Symbian liaison officer before becoming applications support manager for SBU Communicator, in 1999. He is currently working on the new generation of open smartphone and communicator devices and with the Ericsson Developer Zone. Steve holds a B.S. in electronic engineering from Bolton Institute of Technology.
Steve.bridges@eml.ericsson.se

David LeCorney joined Ericsson in 1977 and has worked as a designer, tester, installer, subsystem manager, methodology coordinator, operation and maintenance coordinator, and project manager. He is currently a product manager for Core Switching Products in the AXE Switching Platform. He has worked with AXE for a total of 13 years, in both the control system (APZ) and the traffic-handling system (APT). He has also worked for six years with the AKE system, a predecessor to AXE, and five years with the AXE-N system, which was the forerunner of the current TSP Server Platform system. He holds an M.A. in mathematics from the University of Cambridge.
David.lecorney@uab.ericsson.se

Magnus Enderin joined the Methods, Tools and Training department at Ericsson Utvecklings AB in 1997, and in 1998, he became a product manager for the adjunct processor platform. He is a certified upper secondary school engineer in telecommunication, and holds an M.S. in industrial work environment, production systems from Luleå University and a Licentiate of Engineering in industrial ergonomics from the Royal Institute of Technology in Stockholm.
Magnus.enderin@uab.ericsson.se

Juan E. Figueroa is currently director of the project office for the Cable Access Product Unit within the Home Communications operation of Ericsson's Multiservice Network division. Since joining Ericsson in 1999, he has served as product manager for the PipeRider and for new technologies efforts in the Home Communications group in Lynchburg, Virginia. Juan also previously worked for Bell Labs and Harris Semiconductor (now Intersil). He has a Ph.D. in physics from the State University of New York, Binghamton.
juan.figueroa@ericsson.com

Bill Guzek is currently manager of systems design for the Cable Access Product Unit within the Home Communications operation of Ericsson's Multi-

Magnus Enderin

Juan E. Figueroa

Bill Guzek

Mikael Lindberg

Tomas Lundqvist

Stefan Schultz

service Network division. Since joining Ericsson in 1991, he has held several management positions within Ericsson's R&D units in Lynchburg, Virginia. Most recently, he has been involved in the development of cable modems and in evaluating wired and wireless home networking technologies. He holds an M.S. in computer engineering and an MBA in marketing/innovation management from Syracuse University.
bill.guzek@ericsson.com

Ted Johansson is an expert on RF transistor and process design at Ericsson Microelectronics in Kista, Sweden. He has spent several years developing high-power transistors for base station transmitters. He now divides his time between advanced semiconductor process development for wireless products and the development of power amplifiers in silicon for handhelds. He joined Ericsson in 1989 and holds a Ph.D. in electronic devices from Linköping Institute of Technology, Sweden.
Ted johansson@mic.ericsson.se

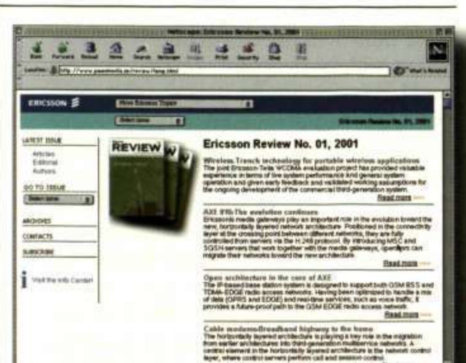
Mats Jonback joined Ericsson in 1982 to help develop the hardware architecture of the APZ 212 central processor. After several years of mainly working with AXE hardware design (which included a one-and-a-half year period at AMD in Texas), he became a project manager. He currently works as a product manager of AXE Central Processors at Ericsson Utvecklings AB.
Mats.jonback@uab.ericsson.se

Mikael Lindberg joined Ericsson as an ASIC designer in 1991. Over the years he has worked with ASIC and hardware system design, always with a focus on hardware development for the core network. He currently holds a line manager position at Ericsson Utvecklings AB, where he works with hardware development of the traffic-carrying parts of AXE. Mikael holds an M.S. in applied physics from the Royal Institute of Technology, Stockholm.
Mikael.lindberg@uab.ericsson.se

Tomas Lundqvist, who joined Ericsson as a system engineer in 1994, currently works with hardware system design in the AXE Systems Management department at Ericsson Utvecklings AB. He holds an M.S. in electrical engineering from the Chalmers University of Technology, Göteborg.
Tomas.lundqvist@uab.ericsson.se

Stefan Schultz, who joined Ericsson in 1987, currently works as a system manager at Ericsson Utvecklings AB. Prior to this, he has worked with system testing (APZ IO), project management, IO support (two years in Australia), in-service performance, and line management.
stefan.schultz@uab.ericsson.se

Ericsson Review is also published on the Web. Now with a new look and improved search capabilities. Visit us at http://www.ericsson.com/review.



Editorial

Eric Peterson

As you probably know, Japan will be one of the first markets to offer WCDMA on a very large scale, and massive efforts are underway there to set the stage for the roll-out this year. Last summer, I had the opportunity to visit Japan to get a first-hand look at the work being done "behind the scenes." I expected to be impressed (otherwise, I would not have scheduled a visit) but I was thoroughly "blown away" by the sheer magnitude of the undertaking as well as the top-drawer caliber of the people I met and the wonderfully contagious enthusiasm and dedication they exhibited! I was also deeply impressed by the maturity of the market—indeed, if any market is ready for third-generation technology and services, it is Japan!

I lived in Japan from January 1980 to November 1981. In those days, public phone booths were a common fixture outside of train stations and department stores. I remember watching with fascination as hoards of commuters would pour out of the train and stream from the station. Invariably, a queue would form by the phones. As a foreigner in the country, I enjoyed watching as callers concluded their conversations, bowing and bobbing as they mouthed their good-byes over the phone. Today, 20 years later, most of the public phone booths have vanished. Instead, the population relies on mobile phones. Virtually everyone I saw this past summer had one: businessmen, "office ladies," school children, even elderly white-haired ladies wearing traditional Japanese kimono (fortunately, some things haven't changed).

Although truly remarkable, I was not solely struck by the extent to which mobile phones have penetrated and taken root in Japanese society. After all, I'm used to seeing people carry and use mobile phones, since I live in Sweden, which is ranked very high in terms of mobile phones per capita. What really excited me was how, and the extent to which, the Japanese use their phones. As I've already stated, I saw people with mobile phones virtually everywhere I looked. But what I found particularly interesting is that I did not actually hear very many phones ringing or even see so many people conversing on their phones—imagine the cacophony that would arise if all the commuters' phones began ringing during the mad morning rush through Tokyo or Shinjuku station! Instead, I saw people everywhere "reading" or "surfing" (that is, they were using multimedia/data services). I observed people from all walks of life actively consulting their phones while they waited for or commuted on a train or bus, roamed the streets, shopped, and hurried to or from lunch. I even saw a motorcycle deliveryman pull out his phone and scan it quickly while he waited for the light to change.

Thanks to iMode/J-phone and the broad range of available services, the Japanese are already fully accustomed to the concept of the mobile Internet. In terms of technology, services and savvy, the market is very mature and ready to make the transition to full-fledged mobile multimedia. Elsewhere, most markets will be implementing packet data (GPRS, cdma2000) this year. In some measure, this will propel us to the point where Japan now stands, laying a secure foundation for our migration toward WCDMA/UMTS/EDGE. The end-user in me is very encouraged—this year, instead of just reading about it, a lot of ordinary folks outside of Japan will actually begin experiencing the mobile Internet first-hand. It is already a hit in Japan, and I predict that people elsewhere will also be thrilled by it!

Eric Peterson
Editor

Wireless-Trench technology for portable wireless applications

Ted Johansson

Until recently, the use of silicon for the production of wireless integrated power amplifiers has not been viable. Here, however, the author describes how a revolutionary new grounding method makes this technique possible.

Wireless-Trench technology was first applied to an integrated power output amplifier for the DECT system—a digital wireless standard for home and office handsets. Using this same technical approach, Ericsson is currently developing a family of related power amplifier products for GSM 900 and 1800/1900MHz as well as GSM dual/triple-band, PDC, other cordless systems, and Bluetooth.

BOX A, TERMS AND ABBREVIATIONS

DECT	Digital enhanced cordless telecommunications
f_T	Maximum transition frequency
GaAs	Gallium-arsenide
GSM	Global system for mobile communication
HBT	Heterojunction bipolar transistor
LDMOS	Laterally-diffused metal oxide semiconductor transistor
MESFET	Metal semiconductor field-effect transistor
NPN	N-type/P-type/N-type
PA	Power amplifier
PAE	Power-added efficiency
PDC	Personal digital communication
P_{in}	Input power
PNP	P-type/N-type/P-type
P_{out}	Output power
QSOP	Quad small outline package
RF-IC	Radio frequency integrated circuit
SEM	Scanning electron microscopy
Si	Silicon
SiGe	Silicon-germanium
V_{cc}	Supply voltage
VSWR	Voltage standing-wave ratio

An important rule of thumb often used by radio circuit designers says that a maximum transition frequency (f_T)—a general figure of merit for any high-frequency transistor—of at least ten times the operating frequency for the active device is needed to design a smoothly running radio frequency (RF) circuit for modern wireless applications. While this may be true for elementary radio building blocks, the design of the output power amplifier (PA) puts additional requirements on the semiconductor process, thus precluding the use of silicon processes for the power amplifier. For this reason, gallium-arsenide (GaAs) metal semiconductor field-effect transistors (MESFET) and other more advanced technologies have been widely in use in this area up to the present time.

As the wireless communication business continues to grow, there is a great demand for reducing the price of all system parts, as well as for using semiconductor manufacturing processes that can handle very high volumes during the short product cycles of many of the new "gadgets." Silicon bipolar integrated circuit technology (RF-IC) is one suitable candidate for non-digital parts. With today's production processes offering an f_T in the 25 GHz range and higher, silicon can even fulfill the technical requirements needed for power amplifiers. The use of RF-IC technology also facilitates the integration of more functionality in power amplifiers. In particular, power control and active linearization are two important features needed for evolving second-generation and soon-to-arrive third-generation systems.

Ericsson Microelectronics has long enjoyed a good reputation for its development of high-power bipolar and discrete laterally-diffused metal oxide semiconductor transistors (LDMOS) for cellular base stations. As a provider of low-voltage RF-IC and GaAs power amplifiers within Ericsson, and lately for use on the Bluetooth market, the company has wide access to the know-how and technology required to manufacture silicon power amplifiers for high-volume applications. Ericsson Microelectronics has gained this know-how through its own development and in-house production, as well as through its partnerships with other vendors. Today, this wealth of experience is being used to explore silicon RF-IC technology for power amplifiers.

Grounding: the essential element for power amplifier performance

One of the most important factors for achieving high performance in power amplifiers is being able to provide a good conducting path from the active devices to ground. The straightforward solution is to use many bonding wires from the chip's surface to the package. Preferably this includes a grounded lead frame, maybe even with a number of the pins permanently connected (fused) to outside ground connections on the circuit board, or a package in which the back side of the lead frame is exposed and soldered directly to the circuit board.

However, at these high frequencies, the bonding wire from the chip to the package severely limits the circuit's performance. To achieve low resistance and low impedance to ground, the number of ground bonding wires is increased; but since a power amplifier is usually quite a small circuit, the additional pads needed for bonding increases the size of the chip considerably, which can prove to be very costly. Packages with exposed lead frames represent no immediate solution to the bonding wire problem, since they are also considerably expensive.

Device technology

The semiconductor process used for Ericsson's silicon power amplifiers is a half-micron, 25 GHz f_T double-polysilicon bipolar process with additional features for improved wireless performance. This process allows operation up to 5 V and includes NPN and PNP transistors for use in analog and digital designs. For the integrated internal matching network, on-chip capacitors and inductors are used. Four layers of metalization (with a thick top layer for improved performance of the integrated inductors) are deployed. Advanced, deep-trench isolation is used to obtain small-sized, low-parasitic, high-performance transistors. Figure 1 shows a schematic view of a trench-isolated bipolar NPN transistor.

Wireless-Trench technology

As was pointed out above, a good ground connection is essential for power amplifiers. In addition to the bonding wires, the substrate (the back side of the die) is usually connected to ground (the package), which enables the use of contacts at the chip surface for ground connection. The contacts consist of metal on highly doped semiconductor material. To achieve low resistance, a high processing temperature and long processing time are required. A contact of this type might have to occupy a considerable area to achieve low contact resistance. Furthermore, the semiconductor structure with which it maintains contact consists of several medium-doped layers on the original substrate with more resistance adding up in the current path as a result. In many RF-IC processes, the substrate itself is low-doped, which precludes its use as a low-resistance path to ground.

In Ericsson's RF-IC process for power amplifiers, the substrate selected is as highly doped as possible, close to the solubility limit. Used together with deep trench isolation, the circuit achieves excellent protection against latch-up effects, which might occur in devices that switch large currents, such as power amplifiers.

To assure excellent grounding from the active device to the highly doped semiconductor substrate and the package, a new type of tungsten-metalized substrate contacts have been developed for the front side. These contacts are formed by etching additional deep trenches through the medium-doped upper silicon layers down to the highly doped substrate, filling the trenches with a barrier material, which is annealed for a short time to form a good metal-to-semiconductor contact, finally filling them with tungsten. The new process module is fully compatible with conventional RF-IC processing, and adds only a few more steps to the process flow.

Figure 2 shows a scanning electron microscopy photograph of a double NPN-transistor structure with its deep trench isolation to the left, and the new trench sub-

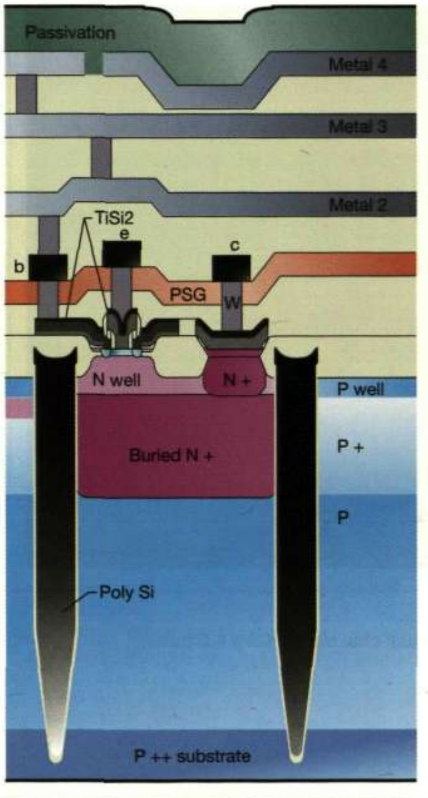


Figure 1
Cross-section of trench-isolated NPN transistor.

Figure 2
Scanning electron microscopy (SEM) photograph of double NPN transistor with new trench substrate contact to the right.

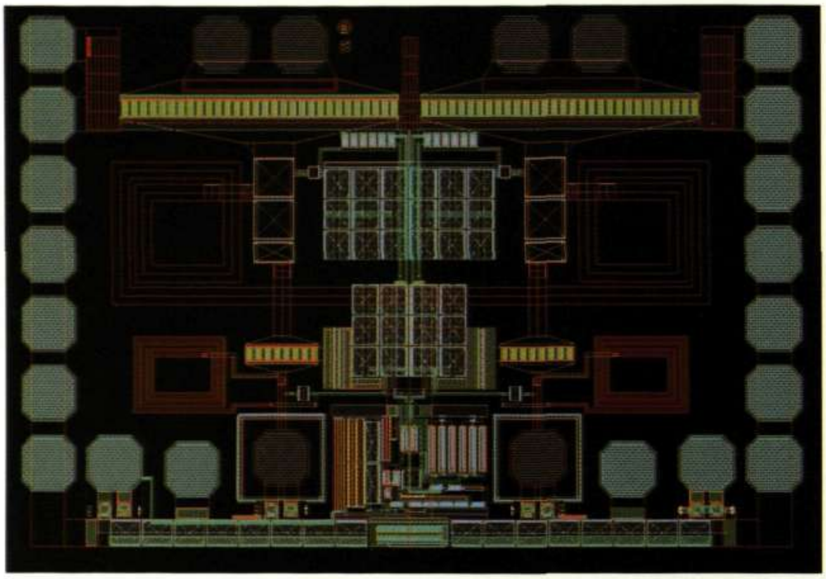
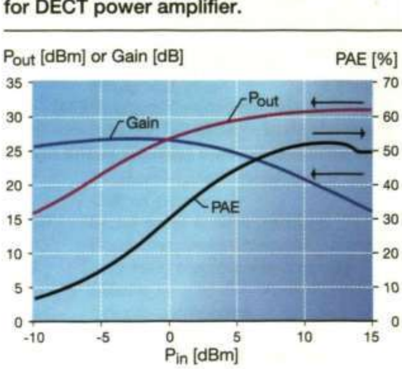


Figure 3
The PBL 403 09 DECT power amplifier chip. Actual chip size is 1.3 x 1.0 mm.

Figure 4
P_{out}, gain and PAE vs. P_{in} at V_{cc} = 3.6 V for DECT power amplifier.



strate contact to the right. Since almost all of the total resistance from the front to the back of the wafer is contained in the top layers—now penetrated by the metal-filled trench—a good ground reference plane and low resistance path to the package are obtained within the chip. Each contact occupies very little area (around one square micron), and can be used in large arrays anywhere on unoccupied circuit areas, say, between blocks or under metalization, to ensure excellent grounding for each part of the circuit. Furthermore, the impedance to ground is now mainly resistive, instead of inductive as is the case when many bonding wires are used. The substrate contacts can also be used to improve shielding and reduce crosstalk between different circuit blocks. In the future, this may prove to be crucial to the integration of radio or control circuitry into the power amplifier chip.

But the greatest potential of the new substrate contact was yet to be demonstrated. If the new metal-filled substrate contact, which provides ground connections from the substrate up to the circuit, is as low-resistive as expected, it should even be possible to use it as a current path from the circuit to ground, instead of the many ground bonding wires usually required in a good power amplifier design. To explore this hypothesis, an experiment was undertaken. A chip used for the DECT power amplifier PBL 403 09 (Figure 3), designed with the new substrate contacts from the beginning, was mounted in an open QSOP16 package. A total of 22 bonding wires were used in this design: fourteen for ground, and eight for supply voltage, input/output signals, and so on. The back-side contact from the chip to the package was of good quality and low-ohmic. The circuit was mounted on its evaluation board, and input/output matching was tuned for optimum performance. The ground bonding wires were then carefully removed, one by one, using a stereomicroscope. When the first wires were removed, the RF performance did not change. With all the ground wires removed, RF performance still remained virtually unchanged. This experiment shows that it is possible to operate a power amplifier without any bonding wires to ground (bringing new meaning to the term *wireless*!) using this new trench substrate contact.

By permanently eliminating the bonding wires to ground, considerable pad area is saved (20% of the chip area for DECT power amplifier), and bonding the chip is simplified (eight wires instead of 22 in the same example). It also becomes possible to select a smaller and cheaper package, thus saving costs and space without any degradation of performance.

This new, unique feature of Ericsson Microelectronics' family of power amplifiers for wireless applications has been dubbed Wireless-Trench technology. Several patents already cover the fabrication and use of the metalized trench contacts.

First application: silicon DECT power amplifier

Wireless-Trench technology was first applied to an integrated power output amplifier for the DECT system. DECT is a digital wireless standard for home and office handsets. Using maximum power of around 0.5 W (27 dBm), the standard communicates at 1.9 GHz, transmitting only 1/24 of the time. The power level and general requirements for the power amplifier are simpler to meet as compared with, say, GSM. A DECT power amplifier was therefore a good starting point for testing the potential of the new silicon RF-IC technology; thus it was chosen as the first product in the new family of silicon-integrated power amplifiers for low-voltage, wireless applications that will be introduced by Ericsson Microelectronics in the near future.

The PBL 403 09 integrated DECT power amplifier, which is housed in a fused QSOP16 package, delivers up to 30 dBm (1 W) of output power at 1900 MHz, with more than 50% power-added efficiency (PAE) The device can be operated up to a 100% duty cycle with minimum performance degradation. Input and output are of the differential type.

Figure 4 shows the input/output characteristics. The two-stage integrated amplifier has on-chip input and inter-stage matching (50 ohms at the input). It also contains biasing circuitry and on/off control signal, and operates using a single 3.6 V supply. It can withstand more than the 5 V supply voltage that occasionally occurs in systems during battery charging, besides withstanding open/short voltage standing-wave ratio (VSWR) conditions to over 5 V with no failures or degradation.

The PBL 403 09 power amplifier is specially designed to interface easily with the PBL 402 15 transceiver chip for DECT. The transceiver has a differential output that interfaces with the power amplifier (using only capacitors for DC-blocking purposes); it also delivers the on/off control signal with appropriate timing. Figure 5 shows how easily the power amplifier is interfaced with the transceiver to form the radio part of a DECT handset or base unit.

The PBL 403 09 silicon power amplifier matches the performance of comparable GaAs products—typically, GaAs products show efficiency in the 40% range, with up to 55% for a few high-performance designs using heterojunction-bipolar-transistor (HBT) technology—Si (35%) and SiGe (30% to 40%). The use of conventional silicon technology makes the use of negative bias or battery switches unnecessary, in contrast to most of the GaAs products that are currently available.

What's next?

Using the same technical approach as was used for the PBL 403 09, Ericsson is currently developing a family of related power amplifier products for GSM 900 and 1800/1900 MHz as well as GSM dual/triple-band, PDC, other cordless systems, and Bluetooth. Figure 6 shows the electrical data for a pre-production sample of the PBL 403 10, a single-ended power amplifier for 900 MHz GSM. At 3.4 V supply voltage, the amplifier delivers 35.5 dBm output power with more than 55% efficiency and 31 dB of small-signal gain. The power amplifier, which is a two-stage integrated amplifier designed for 2.7 to 5 V single-supply operation, has integrated input and inter-stage matching and analog input for the power ramp. In power-down mode, it consumes less than 10 µA of standby current. In addition, this single-ended amplifier uses Wireless-Trench technology, which means that it can be operated without ground wires, and without limits to its performance. Similar impressive results have been demonstrated with power amplifiers for other wireless systems.

Conclusion

The results obtained for power amplifiers designed with the new Wireless-Trench technology, as presented in this article, prove the capability of the new technology for low-cost, high-performance wireless power applications. Moreover, this marks the entrance of silicon into power amplifiers—an area that has been dominated by GaAs technologies.

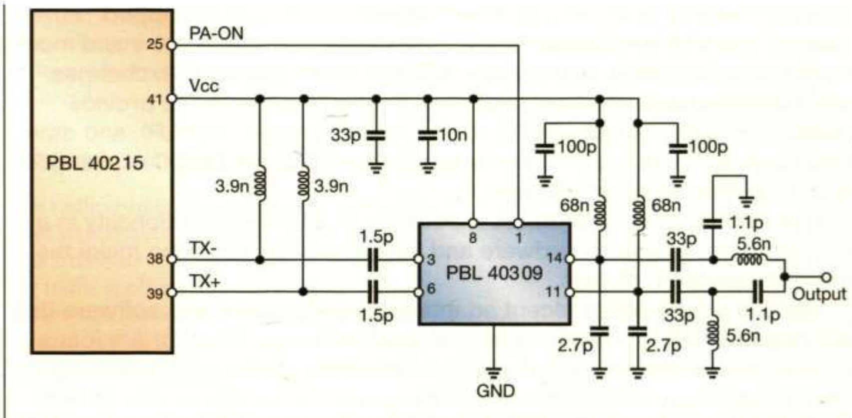
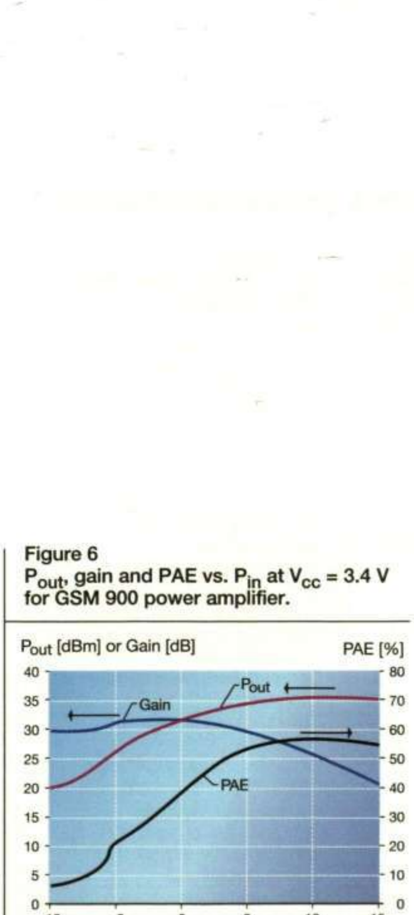


Figure 5
DECT radio with PBL 402 15 transceiver and PBL 403 09 power amplifier.

Figure 6
P_{out}, gain and PAE vs. P_{in} at V_{cc} = 3.4 V for GSM 900 power amplifier.



AXE 810—The evolution continues

Magnus Enderin, David LeCorney, Mikael Lindberg and Tomas Lundqvist

Thanks primarily to an architecture that was developed to support change, the AXE exchange continues to evolve. Its architecture and modularity have benefited customers—AXE has served as local exchanges and international exchanges, and even in mobile networks to provide mobile switching centers (MSC), home location registers (HLR), and other functions. This has resulted in a total number of about 13,000 exchanges and an all-time-high growth rate.

The modularity of AXE makes it possible to add new functionality in a cost-effective way, but hardware and software R&D must also make the most of new technologies.

This article describes recent adaptations of hardware and software that will prepare AXE for the next generation of networks. The authors focus on a system architecture that will serve as the basis for migration toward a server-gateway architecture for third-generation mobile networks and the next generation of multiservice networks. Adaptations will also enable improvements in existing networks where traffic is growing quickly.

Introduction

In the next few years, networks will evolve toward today's vision of an "all-IP" network. We can see two possible scenarios: one for traditional networks, and one for "next-generation" networks. For traditional networks, the scenario describes what will happen in multiservice networks and in second-generation mobile networks, such as GSM. The next-generation network scenario describes the development of mobile Internet and fast Internet access in fixed networks.

In traditional networks, evolution is driven by never-ending growth in the need for processing capacity; in mobile networks, by growth in the number of subscribers and usage levels per subscriber. The wireline network is also experiencing a sharp increase in traffic because of Internet "surfing."

In next-generation networks, traditional telephone and data networks will converge to become general networks designed for many different services and modes of access. The convergence of data and telecommunications makes it possible to combine the best of two worlds. Some requirements, such as the need for heightened performance, are fulfilled more easily when development is based on data communications products. Also, the variety of access modes—via second- and third-generation mobile networks, the multiservice networks, and broadband—will necessitate the coexistence of different transmission formats. Thus, gateways will be required at network interconnection points.[1]

BOX A, TERMS AND ABBREVIATIONS

AAL	ATM adaptation layer
ACS	Adjunct computer subsystem
ALI	ATM link interface
AM	Application module
AP	Adjunct processor
APC	AM protocol carrier
APIO	AXE central processor IO
APSI	Application program service interface
ASIC	Application-specific integrated circuit
ATM	Asynchronous transfer mode
BICC	Bearer-independent call control
BIST	Built-in self-test
BSC	Base station controller
C7	CCITT (now ITU-T) no. 7, a common-channel signaling system
CAS	Channel-associated signaling
CP	Central processor
cPCI	Compact peripheral component interconnect
CPP	Cello packet platform
DL	Digital link
DLEB	Digital link multiplexer board in the GEM
DSA	Dynamic size alteration
DTI	Data tranmission interworking
ECP	Echo canceller in pool
ET	Exchange terminal
ETSI	European Telecommunications Standards Institute
FTP	File transfer protocol
GCP	Gateway control protocol
GDDM	Generic datacom device magazine (subrack)
GDM	Generic device magazine (subrack)
GEM	Generic Ericsson magazine (subrack)
GS	Group switch
GSM	Global system for mobile communication
GSS	Group switch subsystem
HDLC	High-level data-link control
HSB	Hot standby
IO	Input-output
IOG	Input-output group
IP	Internet protocol
IPN	Interplatform network
ISDN	Integrated services digital network
ISP	Internet service provider
IWU	Interworking unit
MGW	Media gateway
MIPS	Million instructions per second
MSC	Mobile switching center
MSCS	Microsoft cluster server
MSP	Multiplex section protection
MTP	Message transfer part
MUP	Multiple position (timeslot)
NGS	Next-generation switch
OSS	Operations support system
PCM	Pulse code modulation
PDH	Plesiochronous digital hierarchy
PLMN	Public land mobile network
PSTN	Public switched telephone network
PVC	Permanent virtual circuit
RAID	Redundant array of independent disks
RAM	Random access memory
RISC	Reduced instruction set computer
RLSES	Resource layer service specification
RM	Resource module
RMP	Resource module platform
RP	Regional processor
RPC	Remote procedure call
RPP	Regional processor with PCI interface
SCB-RP	Support and connection board - RP
SDH	Synchronous digital hierarchy
SES	Service specification
SONET	Synchronous optical network
SS7	Signaling system no. 7
STM	Synchronous transfer mode
STOC	Signaling terminal open communication
TCP	Transmission control protocol
TDMA	Time-division multiple access
TRA	Transcoder
TRC	Transceiver controller
TSP	The server platform
TU 11, 12	Typical urban 11 (12) km/hr
UMTS	Universal mobile telecommunications system
VCI	Virtual channel identifier
VPI	Virtual path identifier
XDB	Switch distributed board
XSS	Existing source system
WCDMA	Wideband code-division multiple access

Typical of next-generation network architecture is the separation of connectivity and communication or control services. This new new architecture will appear first in GSM and WCDMA/UMTS core networks, where AXE exchanges will continue to serve as MSCs. For multiservice networks, telephony servers will become hybrid nodes which consist of an AXE exchange that uses an AXD 301 to handle packet-switched data.

Based on the traditional and next-generation scenarios, network evolution will demand increased processing capacity, greater switching capacity, and conversion to packet-switched technology. In addition, new ways of doing business will emerge as virtual operators pay for the right to use telecom equipment owned by other operators. Similarly, more operators are expected to enter the market, putting additional demands on charging and accounting functions.

To succeed in today's telecommunications market, operators must be able to provide their users with new functionality in networks and complete coverage by mobile systems at the same or a lower cost as time progresses. Operators put demands on the return on investment, cost of ownership, footprint (multiple nodes per site), plug-and-play functionality (short lead times in the field), and quality of software packages (quality of service).

This article describes the latest developments made in the AXE platform to meet these requirements, and what new products will be launched. For example, to shorten the software development cycle, a layered architecture was introduced in the early 1990s, resulting in application modules (AM). Current work on application modules focuses on the server-gateway split. AXE hardware is also undergoing a major architectural transformation to further reduce the footprint, power consumption, and number of board types.

Ericsson's goal is to cut the time to customer for AXE by targeting improvements in production, transportation, and installation. Far-reaching rationalization has been achieved through the introduction of the generic Ericsson magazine (GEM), an open, flexible, scalable, high-capacity magazine (subrack). A new distributed, non-blocking group switch (GS890) has also been introduced, as have new devices, such as the ATM link interface (ALI) and ET155-1, which enable AXE to serve as a gateway to an ATM network. Also, a new input-output (IO) system, called the APG40, has been developed using products from mainstream data communications vendors.

The application platform

The traffic-handling part of AXE has a two-layer architecture: the application platform, which consists of hardware and software; and the traffic applications, which solely consist of software. The application platform can be seen in terms of hardware and software, which present different but complementary views of the architecture.

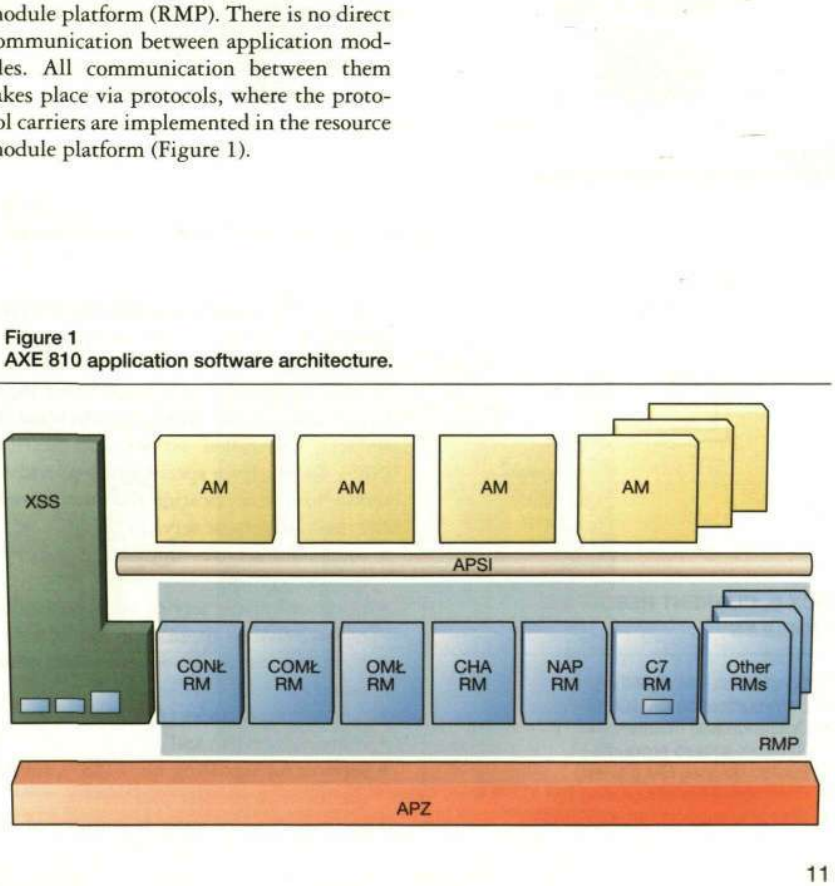
The software view

The AXE application software has continued to evolve since it was first layered and modularized in the early 1990s. Market demand for shorter lead-times was the main force driving the change to layers and modularization.

Application modules

Traffic applications are encapsulated in application modules that use the application-platform software—called the resource module platform (RMP). There is no direct communication between application modules. All communication between them takes place via protocols, where the protocol carriers are implemented in the resource module platform (Figure 1).

Figure 1
AXE 810 application software architecture.



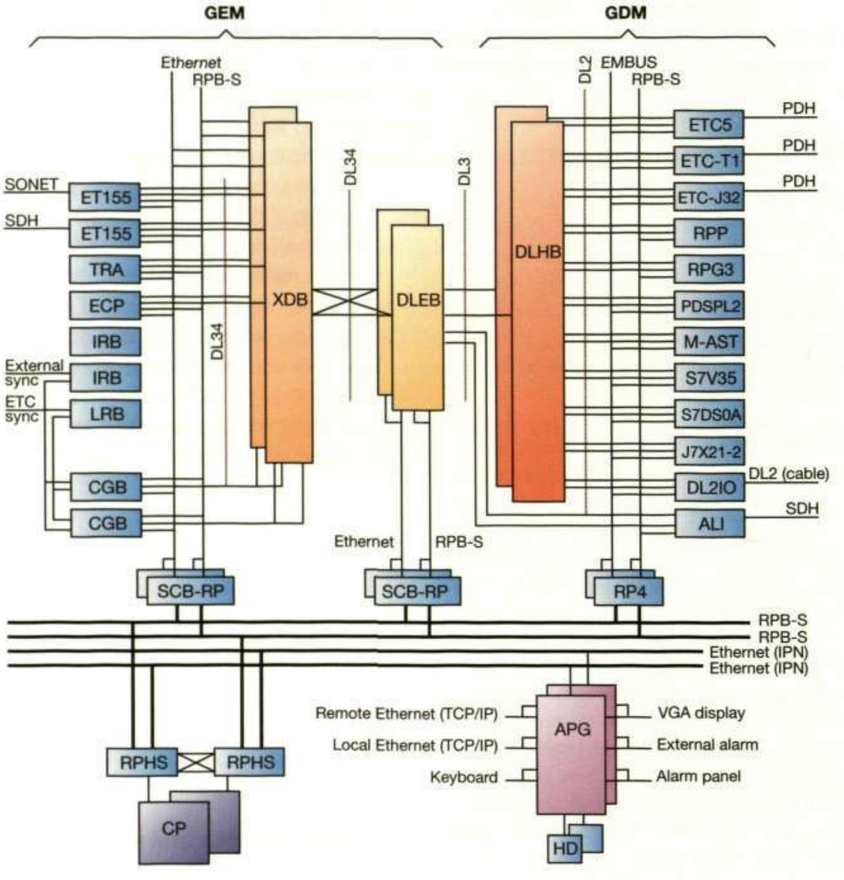


Figure 2
Hardware architectural overview.

The interface of the resource module platform with the application modules, called the application platform service interface (APSI), is a formalized client-server interface. Subsets of the APSI, or individual interfaces, are called service specifications (SES). Each service specification provides a service for the application modules, some of the most important services being:

- connection service—for setting up a physical connection;
- communication service—for communication between application modules;
- object manager—for operation and maintenance;
- AM protocol carrier (APC);
- charging service; and
- services for signaling no. 7 (SS7), such as the message transfer part service.

The total concept, which consists of the RMP, APSI, the AMs, and the inter-AM protocols, is known as the application modularity concept. It is largely thanks to this concept that the system lifetime of AXE consistently exceeds expectations—the AXE system is 30 years old, yet we see no end to its potential evolution.

Resource modules

When the RMP was first introduced, it was small in comparison to the large volume of existing software, called existing source system (XSS). Following several RMP development projects, the migration from the earlier architecture to the AM architecture is now virtually complete. The resource module platform, in turn, has become large and complex, leading to a refinement of the AM concept and the division of the RMP into several resource modules (RM). The interfaces provided by the resource modules are formalized as resource-layer service specifications (RLSES).

A priority in the development of RMs and AMs (known collectively as system modules) is to specify the interfaces (SES and RLSES) early in the development process. When the interfaces are "frozen", the separate system modules can be designed independently of one another, often at different geographical locations, as is common in the Ericsson organization.

The second major advantage for development of applications is that application-platform hardware is now associated with specific resource modules and controlled by them. Application modules simply request services via the APSI. Hardware is "owned" by the software in the respective resource modules.

Hardware

The hardware (Figure 2) revolves around the GS890 group switch, which has 512 K multiple positions, each with a bit rate of 64 kbit/s.

The biggest change in the hardware architecture is the addition of a new exchange interface (the DL34) to existing DL3 and DL2 interfaces. The DL34 interface supports from 128 to 2,688 timeslots to each device board, in steps of 128, via a backplane interface in the generic Ericsson magazine running at 222 Mbit/s. This interface made it possible to improve the devices connected to the group switch, further reducing input power, cabling, the number of board types, footprints, installation time, and other parameters. The high-speed DL34 in-

BOX B, CURRENT RESOURCE MODULES

- Connection RM (CONRM)
- Communication RM (COMRM)
- Operation and maintenance RM (OMRM)
- Network access products RM (NAPRM)
- Pooled devices RM (PDRM)
- Common channel signaling RM (C7RM)
- Charging support RM (CHARM)

terface has also facilitated a more efficient version of the group switch itself, also located in the GEM.

Another architectural change (Figure 2) is the interplatform network (IPN). In the first phase, the IPN will be used to speed up communication between the central processor (CP) and adjunct processors (AP). The interface is based on fast Ethernet. In a second phase, it will be upgraded to Gigabit speed.

All devices that support the DL34 interface can be mixed in the GEM more or less without limit. The maximum capacity of each GEM is 16 K multiple-position timeslots (MUP), which corresponds to eight STM-1 ET boards, for example. Physically, there is space for 22 device boards in one GEM.

If a switch is needed that is larger than 16 K MUPs, or when the number of devices exceeds 22, additional GEMs must be added. These can be configured without interruption while the system is processing traffic. An exchange based on the GEM is linearly expandable. The maximum switch size is 512 K MUPs at normal rate (64 kbit/s) or 128 K MUPs at a subordinate rate (n x 8 kbit/s).

By inserting a pair of digital link multiplexer boards (DLEB) in the GEM, we can convert the DL34 backplane interface into four DL3 cable interfaces. This allows all of today's devices in the generic device magazine (GDM), generic datacom device magazine (GDDM), and other formats to be used with the new switch core.

Software evolution

The original AXE concept has enabled ongoing evolution and modernization. Legacy software can be reused or modified, subsystems and system modules can be modernized or completely replaced, interfaces can be left unchanged or redesigned to meet new needs, and new architectures can be introduced. Yet AXE remains fundamentally AXE.

Several software development programs are under way in parallel. The APZ processor is being adapted to serve as a multiprocessor CP to multiply call-handling capacity. Cost of ownership is being reduced. Ericsson is improving its internal management of software to reduce time-to-market and time-to-customer. These activities are in addition to the four programs described below.

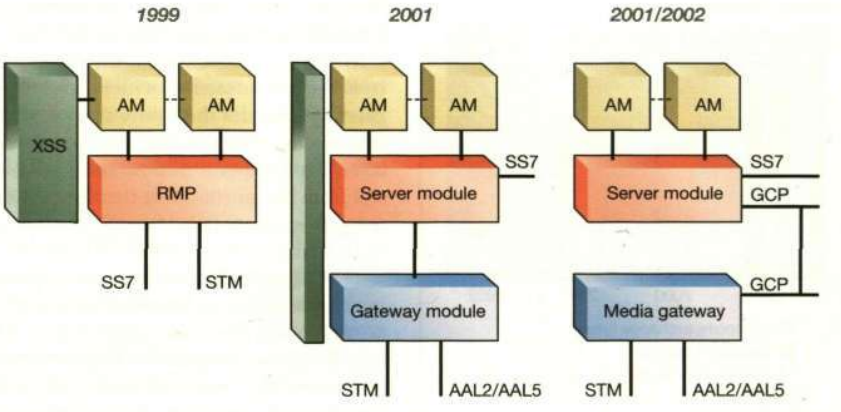
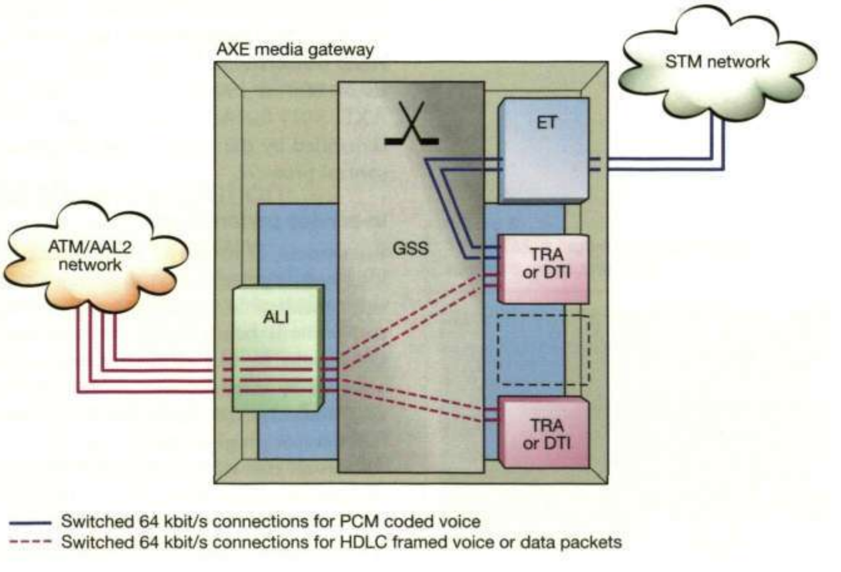


Figure 3
Software evolution toward the server-gateway split.

Software migration

The network architecture is changing. In particular, we see two new types of node: the server (or media-gateway controller) and the media gateway itself, resulting in what is known as the "server-gateway split." In order to meet the demands of the new architecture, the software in the application platform is being migrated from a traditional telecom environment—which is STM-based, mainly for voice traffic—to-

Figure 4
System architecture, ATM interworking function.



System downtime

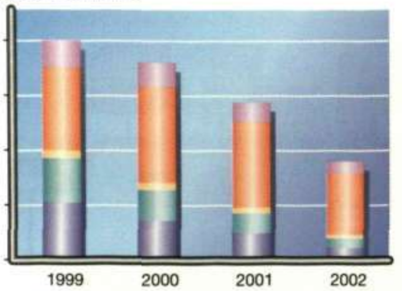


Figure 5
Reduction in system downtime.

ward a packet-based environment, initially for ATM, and subsequently for IP.

In this new architecture, the server controls the call or session, whereas the media gateway provides the bearer service, transmitting information over the connectivity layer. An advantage of separating call control from bearer control is that bearer technology can be upgraded from STM via ATM to IP without affecting the call control.

The original AM architecture has proved itself to be future-proof and is well suited to accommodate the server-gateway split. The split has been implemented in system modules (new and modified AMs and RMs) without any fundamental changes to the software architecture. For example, new RMs for bearer-access service and for TDM access are currently being developed for the WCDMA/UMTS program.

Two separate migration tracks are currently being followed. For the WCDMA/UMTS program, AXE is the natural choice of technology for the server, thanks to its high capacity and high level of functionality. AXE is also available as a combined server/media-gateway node, where the server part controls its own gateway part. By contrast, the Cello packet platform (CPP) is the preferred choice of technology for ATM-based media gateways. In such a network, the server nodes communicate using bearer-independent call control (BICC). The server controls the gateway using the gateway control protocol (GCP). ATM-based gateways communicate using Q.AAL2.

The second migration track is for the next-generation switch (NGS) program for the multiservice network. The server node is a hybrid node made up of co-located AXE and AXD 301 systems. The gateway node consists of an ATM switch (also an AXD 301) for AAL1 connections, and is controlled by the server using the gateway control protocol.

In-service performance

Robustness, or in-service performance, has long been a priority in AXE development, and considerable improvements have been and are still being made. No software is 100% reliable, but robust systems can be achieved by means of recovery mechanisms that minimize the effects of software malfunctions or program exceptions.

One of the most powerful robustness mechanisms is known as *forlopp*, which is a recovery mechanism at the transaction level. (Forlopp is an anglicization of the Swedish word *förlopp*, roughly the "sequence of events.") The forlopp mechanism is a low-level recovery mechanism by which only the transaction affected by the software malfunction is released. This minimizes the disturbance to the overall system. The forlopp mechanism, which has been refined in several stages and is now applied to all traffic handling in AXE and to many operation and maintenance functions, has significantly reduced the number of recoveries at the system level.

System restart is used as a recovery mechanism for certain faults and for system upgrades. The restart mechanism itself has been optimized in several ways, so that in the few cases that still require restart, its duration is as short as possible.

Activities for restarting software in the regional processor (RP) have been especially improved. For example, regional processors are restarted through minimal restart by default—that is, with the suppression of unnecessary actions. Complete restarts are performed on regional processors only when necessary. Regional processors are restarted asynchronously, which means that no regional processor has to wait for any other regional processor to become ready for a restart phase. These improvements have significantly reduced the time consumed when restarting application hardware.

If necessary, AXE can, via the IPN, be reloaded from the backup copy of the system on the APG40 file system. This approach is ten times faster than the design that applied before the APG40 and IPN were introduced. Also, the time it takes to make a backup copy of the system on APG40 is one-fifth of what it was before.

The restart duration for a system that is upgrading to new software has also been reduced. The most recent improvements involve RP software, which is now preloaded prior to an upgrade, instead of during the upgrade, thus saving restart time.

Major improvements have been made to the function-change mechanism used for system upgrades. Inside the central processor, the time needed for data transfer prior to a changeover to new software has been cut from hours to typically ten minutes. (The data transfer does not disturb traffic handling, but exchange data cannot be changed during data transfer.)

Another improvement applies to the retrofit of new CPs that replace old ones. The bandwidth of the connection between the processors has been expanded by using regional processor with industry-standard PCI interface (RPP) and Ethernet connections, thereby reducing data-transfer times ten fold. As a consequence of these many improvements, system downtime has been reduced significantly in recent years, a trend that is sure to continue (Figure 5).

Software licensing

The main force driving software licensing is customer preferences to pay for access to specific functions, features, and capacity. Thus, "software licensing" is really the "licensing of features and capacity." The second driving force is Ericsson itself, because it is in the company's interest to deliver standard nodes (those that contain standardized software and hardware).

With software licensing, Ericsson delivers a standard node for a particular market or market segment, such as an MSC for a cellular system using time-division multiple access (TDMA). This standard node consists of a complete software configuration with a standard hardware configuration that is deliberately over-dimensioned. Ericsson then limits the call capacity and functionality of the node by means of software keys. When a customer requests more call capacity, Ericsson personnel execute a password-protected command to increase the capacity in the node to the new level.

This method of increasing capacity, referred to as "traffic-based pricing" or "pay-as-you-grow," is much simpler than the traditional method, in which Ericsson personnel would deliver, install, and test new hardware on a live node. The commands can also be executed remotely from a maintenance center, making a visit to the site unnecessary. This arrangement benefits customers and Ericsson, especially in the mobile market where growth can be rapid and unpredictable.

Similarly, in the future, software will contain functions and features for which customers can purchase licenses on an *ad hoc* basis. The software for these functions and features will be unlocked using password-protected commands. This method of delivering software is much simpler than the traditional method, by which new software is delivered in the form of software upgrades.

The licensing of call capacity is already available on AXE nodes, and the introduction of a general mechanism to handle all software licensing of functionality and capacity is planned. Some of the characteristics of this general mechanism are as follows:

- Software licensing will be common to AXE 810, TSP, CPP, and other Ericsson platforms.
- Ericsson will maintain a central license register, from which new licenses can easily be issued.
- License keys will be distributed electronically, in encrypted form.

Improved handling

Several improvements to the operational handling of AXE have recently been made, such as parameter handling and hardware inventory. Other handling improvements are being planned, such as plug-and-play functionality. These improvements reduce operational costs and also often have a positive impact on the ISP, since they reduce the risk of human error.

One of the most important improvements to handling is called dynamic size alteration (DSA). The size of many data records in AXE is traffic-related, since the number of individuals, or instances, in the record varies from node to node, as well as over time, in accordance with the capacity demands put on the node. Thanks to dynamic size alteration, the number of individuals in the record is increased automatically, up to a preset limit, without any intervention from the maintenance technician. However, as a warning, AXE issues an alarm when, say, 75% of the reserved data capacity has been used up, so that the technician can raise the limit as necessary.

Dynamic size alteration also simplifies the handling of this kind of alteration, since the alarm (referred to above) specifies an action list of the data records whose size needs to be altered (increased). The technician merely enters a single command to increase the sizes of all these data records.

Hardware evolution

Generic Ericsson magazine

Another far-reaching improvement is the GEM, a high-capacity, flexible, and scalable magazine (subrack) that anticipates future developments (Figure 6). GEM-based nodes will be smaller, dissipate less power, and have greater maximum capacity. Their implementation will dramatically improve the cost of ownership and cut time-to-customer for AXE.

In previous versions of AXE, each function was located in a separate magazine, and

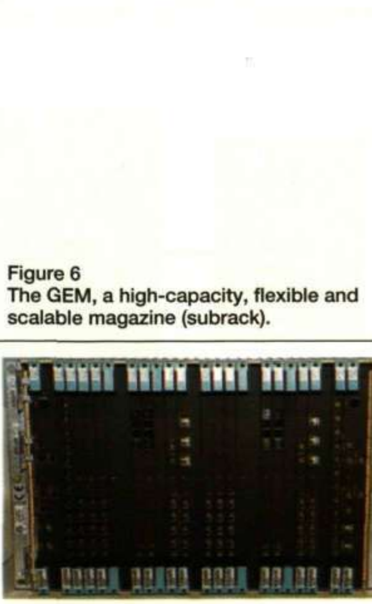


Figure 6
The GEM, a high-capacity, flexible and scalable magazine (subrack).

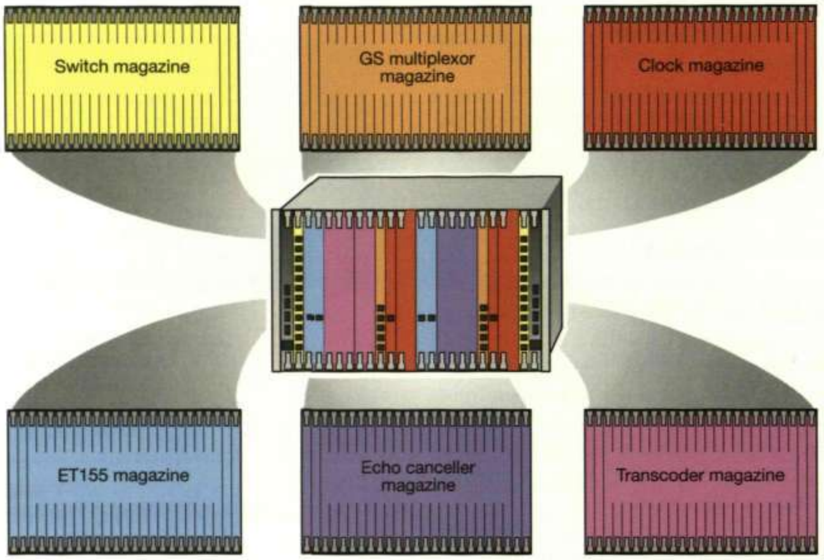


Figure 7
Many different functions can be mixed in the GEM.

the associated regional processors were on separate boards in the magazine. The GEM and the boards developed for the GEM represent fundamental change.

Several functions are mixed in a single magazine, and the RP function is integrated on each board. This makes the most of advances in technology to modify the architecture rather than merely shrink the hardware. One advantage of this change is that considerably fewer magazine and board types are needed in each node. A second advantage is that many internal AXE interfaces have been moved to the GEM backplane, reducing the need for cables. As shown in Figure 7, many different functions can be mixed in the GEM.

Boards developed for the GEM have high capacity, and because more functions are concentrated on each board, the total number of boards can be reduced. Mixing these high-capacity boards in a single magazine and removing many cables by moving the interfaces to the GEM backplane produces a more compact system.

The GEM concept can be used to build anything from an extremely small node to an extremely large one, using the same boards and magazine (Figure 8). The smallest node will consist of one GEM, and the largest, 32 GEMs. As capacity requirements grow, the node can be expanded smoothly and without interruption of traffic by adding one or more GEMs with a suitable combination of devices. A GEM can house

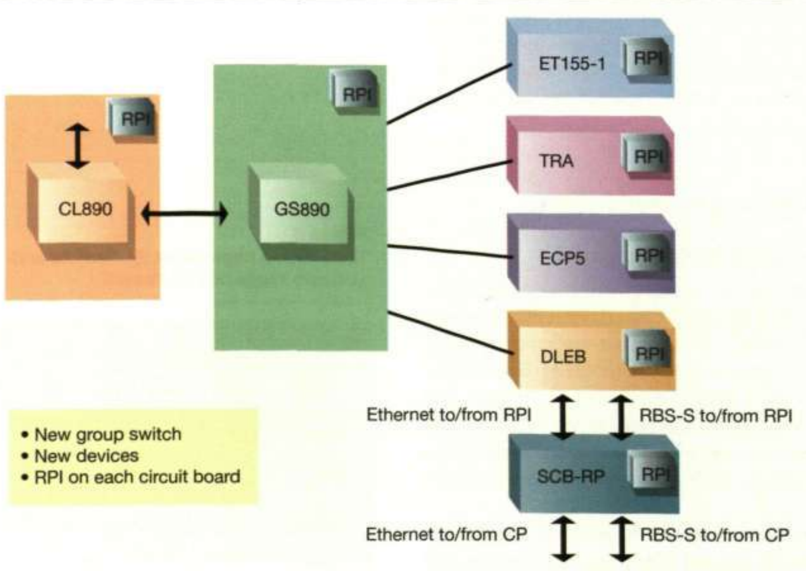
- two SCB-RPs, providing an Ethernet switch, control function, maintenance support, and power distribution;
- two switch boards (XDB), providing a 16 K plane duplicated group switch;
- up to 22 device boards with 15 mm spacing, such as ET155, ECP, or TRA;
- pairs of DLEBs, providing multiplexing functionality for DL3 cable interfaces, placed in device slots; and
- CL890 clock modules placed in device slots.

The GEM, which has been developed for use in the BYB 501 equipment practice, provides several infrastructure functions via the backplane for the boards housed in the magazine:

- duplicated power distribution;
- duplicated group switch connection;
- duplicated serial RP bus;
- maintenance bus for hardware inventory and fault indication;
- duplicated Ethernet, 10 or 100 Mbit/s; and
- extremely robust backplane connectors.

The option to use Ethernet as the control protocol makes the GEM an open magazine prepared for future developments. Another advantage of the GEM is that it is generic and can be used in non-AXE products from Ericsson.

Figure 8
The GEM concept can be used to build anything from an extremely small node to an extremely large one, using the same boards and magazine.



The GS890 group switch

The GS890 fully supports the GEM concept and its goal of dramatically improving cost of ownership and time-to-customer. The figures in Table 1 show an extraordinary reduction in power dissipation and number of boards and cables. This was achieved through a combination of recent advances in ASIC technology and high-speed interfaces with architectural modifications.

The maximum size of the GS890 has been increased considerably, making it possible to build larger nodes. For many network configurations, it would be better to build and maintain a small number of large nodes. The switch is also strictly non-blocking, so many limitations on the configuration of the node and the surrounding network have been removed.

The subrate switch function in AXE, which enables switching at 8 kbit/s, has also been modified. Primarily used in GSM networks, this function makes efficient use of pooled transcoders and transmission resources surrounding the base station controller (BSC). The subrate function was previously implemented as a pooled function, but in GS890 (Box C) it has been integrated into the switch core and increased to 128 K. The advantages are that much larger BSCs are feasible, less equipment is needed, and the traffic selection algorithms are easier and faster because the pooled switch has been removed.

GS890 group switch connection

All existing group switch interfaces are supported, making it possible to connect all current devices to the GS890. The DL2, which is a cable and backplane interface, carries 32 timeslots. Similarly, the DL3, which is a cable interface used by the ATM interworking unit (IWU) and GDM, carries 512 timeslots. The new DL34 backplane interface supports new high-capacity devices developed for the GEM.

TABLE 1, IMPROVEMENTS IN THE GROUP SWITCH

Characteristics	GS12 (Max 128 K)	GS890 (Max 512 K)	Reduction
Power dissipation in 128 K MUP group switch core	1200 W	250 W	81%
Internal cables in a 128 K MUP group switch core, and including clock distribution	1,024 (12-pair cables) and 48 (4-pair cables)	88 (4-pair cables)	92%
Number of boards in a 128 K MUP group switch core, excluding clock boards	320	16	95%
Equivalent figures for a 512 K group switch	--	1,000 W, 448 (4-pair cables) and 64 boards	--

GS890 hardware

The main functionality of the GS890 is implemented in two new ASICs: one core ASIC, which handles the switching; and one multiplexer ASIC, which concentrates traffic into the core. Both ASICs have been designed using 0.18 µm CMOS technology.

The core ASIC contains 4 Mbit of RAM, runs at up to 333 MHz, has 500 K gate logic and 20 high-speed interfaces running at 666 Mbit/s, allowing the ASIC to handle up to 13 Gigabits of data every second. In a fully expanded switch matrix, 64 of these ASICs will be needed per plane. Standard four-pair cable is used to interconnect the switch matrix elements located in different magazines. The cable can be up to 20 meters long and carries 1.3 Gigabits of bidirectional data per second, two 666 Mbit/s interfaces in each

BOX C, MAIN CHARACTERISTICS OF THE GS890

- Maximum size 512 K MUP, equivalent to 524,288 (64 kbit/s) ports
- Strictly non-blocking architecture, regardless of traffic type
- Fully distributed group switch, with switch matrix distributed among up to 32 GEMs
- Integrated subrate switching capability up to 128 K MUP, equivalent to 1,048,576 (8 kbit/s) ports
- Improved maintenance support by logic BIST and RAM BIST in all ASICs, reducing maintenance time and the time that the system runs on one plane
- Hardware support for fast dynamic fault isolation
- New and flexible high-speed-device interface
- Device protection support with no wasted capacity (used for ET155-1)
- New high-speed, internal group switch interface
- Rear cabling removed

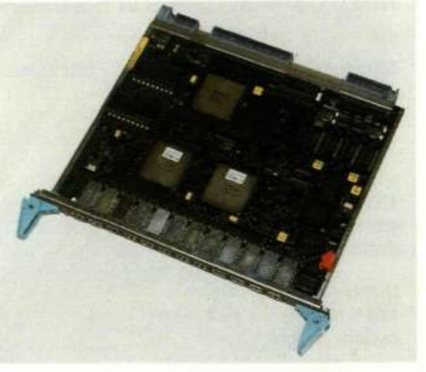


Figure 9
Photograph of the switch distributed board (XDB).

direction. The new hardware requires two new board types:

- the switch board; and
- the multiplexing board, which makes it possible to connect existing devices.

In addition, two more boards have been developed to support a smooth migration from the GS12 to the GS890. The migration consists of replacing one board in each GS magazine, thereby converting it into a multiplexing magazine. The concentrated traffic is then connected to the new GS890 core. These boards help safeguard investments already made in AXE. The GS890 (Figure 9) supports the following types of traffic:

- normal rate, 64 kbit/s;
- non-contiguous wideband, *n* x 64 kbit/s up to 2 Mbit/s;
- subrate, 8 kbit/s or 16 kbit/s;
- wideband on subrate, *n* x 8 kbit/s up to 256 kbit/s; and
- broadband connections up to 8 Mbit/s.

The CL890 synchronization equipment, also developed to support the GEM concept, consists of

- duplicated clock modules;
- up to two highly accurate reference clock modules; and
- up to two incoming clock reference boards for connecting additional clock references.

Devices

RPG3

A new version of the RPG2 with slightly improved CPU performance has been designed that only occupies one slot instead of two. The RPG3, which is still located in the GDM, is mainly used for different kinds of signaling application.

ET155-1

The ET155-1 is an SDH exchange terminal that will occupy one slot in the GEM and will be available for ETSI and ANSI/SONET standards. The ETSI variant will carry up to 63 TU12s and support both optical and electrical line interfaces. In ANSI/SONET mode, the board will be able to handle up to 84 TU11s.

The board fully supports channel-associated signaling (CAS) in both standards by extracting the signaling information from the incoming PDH frames and sending it through the switch to pooled signaling terminals. The ET155-1 can work either as a non-redundant single-board exchange terminal or in tandem, supporting MSP 1+1 and equipment protection.

One motherboard, which contains the main functionality, and two generic line interface modules, containing the optical respectively electrical line interface, have been developed. An ET155-1 board is produced by mounting the required module on the motherboard. The line interface modules are generic and will also be used in non-AXE products from Ericsson.

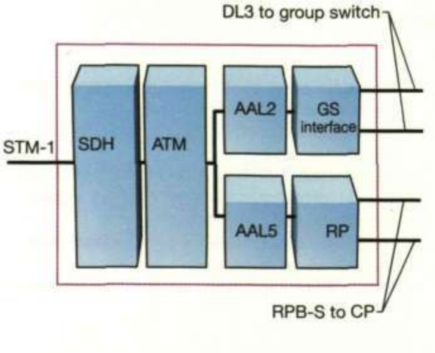


Figure 10
Block diagram of the ATM link interface (ALI).

Transcoders and echo cancellers

A big improvement has been made in the transcoder area. A new board designed for the GEM supports 192 channels for all the codecs used in GSM and TDMA systems. The same board is also used for echo cancellation, in which case it supports 128 channels.

A fully equipped GEM with transcoders supports 4,224 transcoder channels; when fully equipped with echo cancellers, the GEM handles 2,816 channels.

Future GEM devices

The devices described here were designed for the first release of the GEM. Future releases will include more devices ported from today's GDM and GDDM. Technology allows more traffic to be handled by each board in AXE, so the boards can efficiently take advantage of the configurable bandwidth of the DL34.

ATM interworking function

With the introduction of WCDMA/UMTS, AXE will be used as a combined server and media gateway and purely as a server. In the downlink direction (to the radio access network), AXE will be attached to the *Iu* interface, a new interface based on ATM. ATM will also be used in the core networks.

An ATM interworking unit has been implemented to add an ATM interface to new and legacy AXE nodes. The interworking unit enables existing AXE-based MSCs to be upgraded easily to handle signaling and the user plane in ATM networks. The interworking unit is connected directly to the group switch interface for the user plane, and the signaling information is transferred to the central processor via the RP bus (Figure 10).

The interworking unit provides a 155 Mbit/s ATM/SDH interface to AXE. New hardware and software for AXE enable it to perform as a gateway to an ATM network capable of handling voice, data, and signaling. The first release of the ATM interworking unit supports three types of payload in the ATM network:

- coded voice on AAL2;
- circuit-switched data on AAL2; and
- signaling on AAL5.

The device that implements the interworking unit is called the ATM link interface (ALI). This hardware unit is divided into three boards that are mounted together on a plug-in unit which can be located in a full-height GDM that provides RP bus, –48 V and maintenance bus via the backplane. The unit is then connected via DL3 cables to the group switch. The fiber for the STM-1 interface is connected to the external equipment.

Adjunct processor platform

Statistics from the telephony or radio parts of a network must be collected to optimize network efficiency. At the same time, virtual operators are willing to pay for the right to use telecom equipment owned by other operators, and network operators are demanding fast, almost instant billing (such as with prepaid cards). These requirements demand processes with close to real-time execution.

In terms of operation and maintenance, it is better to have one big node in a network instead of several small ones, although this demands greater switching capacity. To reduce operators' cost per subscriber, big switches are being planned. This has spurred the development of new high-end AXE switches. There are numerous methods of

BOX D, ATM VIA AXE

To transmit coded voice over ATM, a permanent virtual circuit (PVC) is set up between the ALI and the remote end. The PVC is specified by its VPI/VCI address. The first ALI release allows a total of eight PVCs to be set up in the user plane.

Within each PVC, AAL2 connections (based on the I.363.2 standard) can be set up and released using Q.2630.1 (Q.AAL2) signaling. Up to 248 AAL2 connections can be set up in each PVC.

The C7 signaling links for supporting Q.AAL2 and other user parts are configured as another type of PVC. For each ALI, up to 64 signaling PVCs can be configured to carry adaptation layer AAL5, based on the I.363.5 standard. Each of these AAL5 channels carries a C7 signaling link using SSCOP (Q.2110) as the link protocol.

Adaptation layers for both ETSI (SSCF-NNI, Q.2140) and Japanese C7 standards (JT-Q.2140) are provided.

The architecture for supporting circuit-switched data is similar to that for coded voice except that the payload packets are routed to a data transmission interworking (DTI) unit instead of a transcoder. The first release of ALI supports a total of 32 data channels.

In the future, the ATM interworking concept might be used to implement new functions such as 64 kbit/s voice on AAL1 or on AAL2 based on I.366.2. Another function in the works is "soft PVCs," which can be operated using B-ISUP signaling.

Figure 11
Photograph of a large APG40 cluster.

enhancing performance, since AXE processing power is distributed among several processors (CP, RP, and AP), and research and development can be pursued along different lines in parallel.

APG40 platform for near real-time applications

The most widely deployed AXE IO system, IOG20, was introduced in 1997. The IOG20 is a proprietary product for the CP file system, man-machine communication, boot server, and statistics functionality. In 1998, the first adjunct processor—the APG30—was released, primarily as a billing system. The APG30 is widely deployed in TDMA systems as a platform for billing but also for operation and maintenance of AXE as an IOG replacement. The APG30, which is based on a commercial computer from an external vendor, uses MIPS RISC microprocessors and NonStop-UX O/S.

The APG40, successor to both the IOG20 and APG30, is a platform for near real-time applications. Three needs have driven the development of the APG40 (Figure 11). One need is to offer a platform for new data-handling tasks and applications that are best performed at the network element level by a standard computer system. Another need is for improved performance, as a result of increased traffic per switch and new types of data extracted from each switch. The third need is to reduce time-to-market and time-to-customer.

The APG40 is a platform for the AXE central processor IO functions (APIO) that were inherited from the IOGs. It is also a platform for billing (for example, FOS) and statistic data collecting (for instance, STS), storage, processing and output from the AXE switch. For example, the APG40 can be a platform for collecting data related to in-service performance from the central processor. It can also format that data for distribution to an operations management

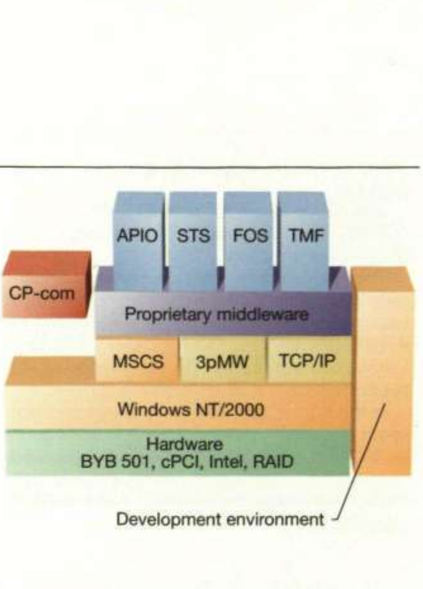


Figure 12
APG40 building blocks and development environment.

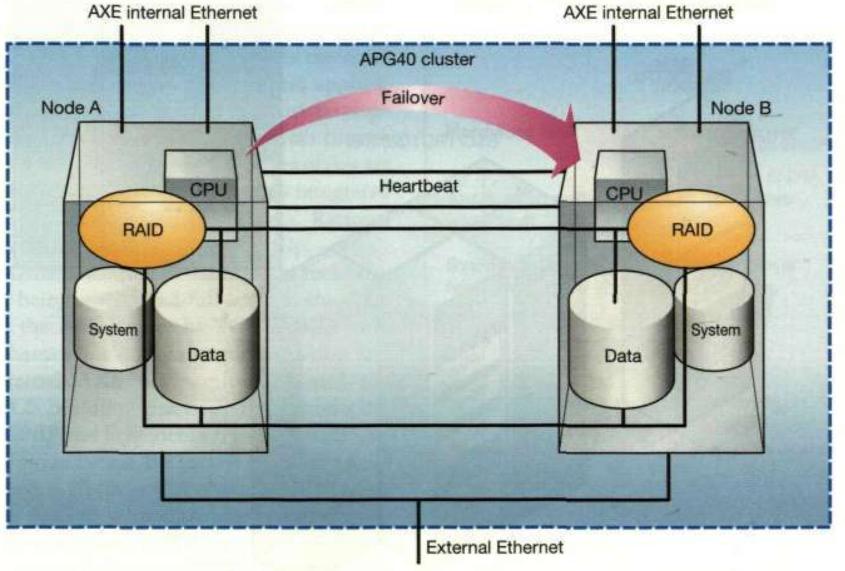


Figure 13
The RAID hardware system.

center, such as the operations support system (OSS), for preventive measures or for sending an alarm when a critical level is reached.

The applications that run on the APG40 do not communicate with one another and are not interdependent. The platform provides basic functionality, such as highly available processing capacity, safe storage, and data output mechanisms.

The adjunct computer subsystem (ACS) (proprietary middleware in Figure 12) provides the software platform for applications in the APG40. The ACS offers Windows NT, Microsoft cluster server (MSCS), and communication protocols, such as FTP, RPC, TCP/IP, and Telnet.

In-service performance

Operators require reliable input-output systems that store AXE system backup copy safely and perform reloads fast. A switch must always be available for operation and management from an operations support system. To ensure the in-service performance of the APG40 itself, and AXE as a whole, concepts introduced in the IOGs and APG30 have been brought to this new platform. In the APG40, mainstream technologies, such as MSCS and RAID hardware, form the basis for telecommunications-grade in-service performance.

The APG40 is a cluster server (Figure 13). In its first release, a cluster will consist of two nodes. Each node has its own memory and runs the operating system. All applications running on an APG40 cluster are active on one node only—the executive node; the other node is a standby node. Thus, if a major fault occurs, the standby node starts and takes over the application. A watchdog in the MSCS ensures that the adjunct processor software does not simply stop working. When an error is detected, either in hardware or software, the cluster server can restart the failed resources in the other node. This process is called *failover*.

The physical size of the APG40 is a half magazine for a small cluster and one full magazine for a large cluster. The small APG40 has one 18 GB disk per node, while the large APG40 can house up to three 18 GB disks per node. Later on, it will be possible to introduce disks with a capacity of 36, 72, or 144 GB.

Thanks to the RAID hardware system (Figure 13), the APG40 has a higher disk-integration speed than the IOG20. The RAID hardware system is used for safe storage. Only the executive node can write data to the data disk. To avoid loss of data due to a disk error, the executive node writes data to a Windows NT file system, and the RAID system writes to the data disk in both the executive node and the standby node. The executive node thus owns all data disks in both nodes. In the IOG20, on the other hand, each node owns its own set of data

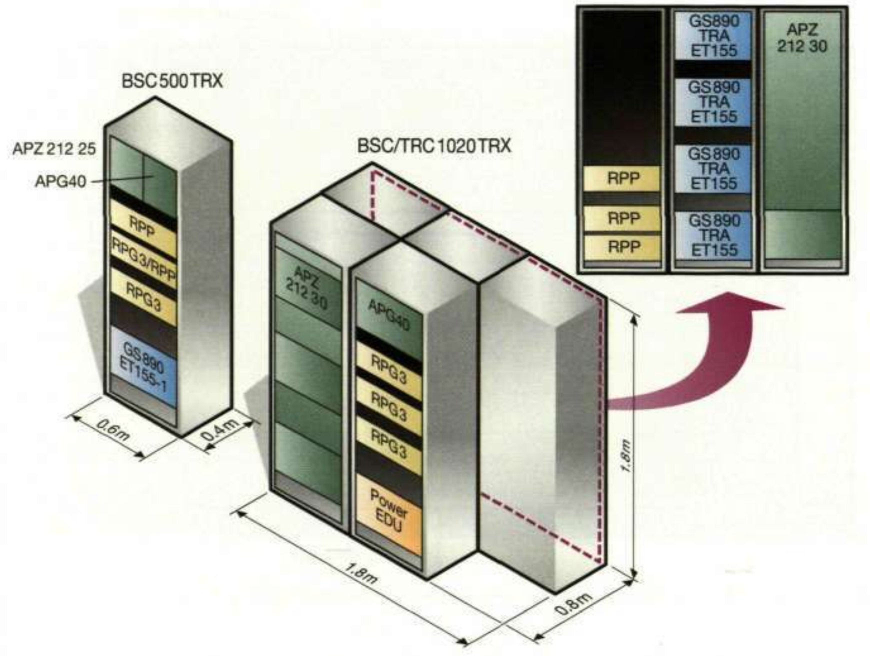


Figure 14
Node layout for BSC 500 TRX and BSC/TRC 1020 TRX.

disks, which must be re-integrated each time a failover occurs. Re-integration in the APG40 occurs when a node is replaced or the RAID system fails.

The AXE IO platforms listed in Table 2 are based on three different high-availability concepts. In the IOG20, the passive node is *hot-standby*, whereas in the APG40, the passive node is *warm-standby*. In this context, the main difference between hot- and warm-standby is that a larger part of the software is up and running in hot standby. In the APG30—with its fault-tolerant concept—identical processes are executed in parallel on separate processor units. If a hardware failure occurs, an immediate failover is initiated from the failing processor unit.

Behind the different availability concepts in Table 2 lie different assumptions about the most probable source of failure. If a hardware failure happens more often than a software failure, the fault-tolerant technology gives better in-service performance. A failover in the APG30 executes in an instant, because the same processes, including the applications, are running in parallel on both sides.

Nevertheless, because the risk of hardware failure is considered extremely low, the case for dimensioning has been based on software failure, and the high-availability concept used in the APG40 has been introduced as the preferred solution. Given the risk of hardware faults, application software faults and system software faults, the failover period has been held to a mininum. On the APG40, these three kinds of fault are well below the minimum CP buffer capacity of two minutes, which means that no data is lost.

Performance and functionality enhancements

The demands put on the adjunct processors are not the same as those put on the central processor. The central processor provides non-stop computing, so the adjunct processor concept is based on and provides high-availability computing. Consequently, new processor systems can be obtained from mainstream data communications vendors. The first generation APG40 is based on Intel Pentium III 500 MHz processors and Microsoft NT4 clustering technology. By adapting these mainstream products to spe-

TRADEMARKS

cific needs, Ericsson benefits from ongoing research and development conducted by others. This approach also simplifies the rapid introduction of new features and applications using Microsoft and Rational development tools integrated in an adjunct processor design environment. Examples of this are debugger and Rational test tools integrated in Microsoft Visual C++, Rational ClearCase, and HTML online help.

Other examples of mainstream technology being introduced full-scale in the AXE by the APG40 are the TCP/IP, FTP, and Ethernet (for external communication and internal AXE communication), and the cPCI building practice. By introducing TCP/IP and Ethernet in the IPN for CP-AP communication, the central processor reload speed is 10 times that obtained when using the APG40 with signaling terminal open communication (STOC).

The introduction of TCP/IP as the external interface to the OSS is a major step toward replacing the proprietary interfaces based on MTP and X.25. The external Ethernet board will enable a TCP/IP connection of up to 100 Mbit/s.

Telnet has been introduced for handling commands. Similarly, FTP and RPC are used for handling files. Support for a component object model-based (COM) OSS might be introduced in the future.

For local element management of the AXE 810, WinFIOL and Tools 6.0 (and later versions) have been adapted to the new APG40.

TABLE 2, THREE AXE IO SYSTEMS

	HSB (IOG20)	**FT (APG3x)**	**HA (APG4x)**
Hardware fault	Software failover 90 sec.	Hardware failover < 1 sec.	Software failover < 60 sec.
Application software fault	Software failover 90 sec.	Process restart 22 sec.	Process restart < 1 sec.
System software fault (major)	Software failover 90 sec.	Reboot 80 sec.	Software failover < 60 sec.

Results

Best-in-class hardware embodies compact layouts and easy extensions, exemplified in Figure 14 by one BSC and one combined BSC/TRC for the GSM system.

The BSC, which has a capacity of 500 transceivers, occupies only one cabinet. At the top, the CP (APZ 212 25) and APG40 share one shelf. Below this are three magazines with a mix of RPPs and RPG3s. The C7, transceiver handler and PCU applications run on this hardware. At the bottom of the cabinet, one GEM contains ET155, synchronization, DL3 interface boards (DLEB) and the group switch.

The combined BSC/TRC has a capacity of 1,020 transceivers and approximately 6,000 Erlang. The APZ 212 30 is needed for this larger node, occupying two cabinets. Another cabinet houses four GEMs, each of which contains a mix of transcoders, ET155, synchronization, DL3 interface boards and the group switch. The next cabinet contains one APG40, the RPG3s (for transceiver handling) and C7. The power equipment is located at the bottom. PCUs can be included in the last cabinet if the BSC/TRC is intended to handle GPRS traffic.

Conclusion

The AXE architecture allows us to continue to improve the different parts of the system as required by network evolution. This ensures that operators always have a network with the best possible cost-of-ownership. A limited number of high-capacity nodes with a limited footprint keep site costs down. State-of-the-art hardware with reduced cabling also ensures quick, high-quality installation procedures. Operation and maintenance is kept economical through low power consumption, an interface consistent with the OSS, and a high-capacity, user-friendly extension-module platform. The software architecture allows migration toward next-generation networks, whether fixed or wireless.

REFERENCES

1 Ekeroth, L. and Hedström, P-M.: GPRS support nodes. Ericsson Review Vol. 77 (2000):3, pp. 156-169.
2 Hägg, U. and Lundqvist, T.: AXE hardware evolution. Ericsson Review, Vol. 74 (1997): 2, pp. 52-61.
3 Hallenstål, M., Thune, U. and Öster, G. ENGINE server network. Ericsson Review Vol. 77 (2000):3, pp. 126-135.
4 Andersson, J-O. and Linder, P. ENGINE Access Ramp—The next-generation access architecture. Ericsson Review Vol. 77 (2000):3, pp. 146-147.

Open architecture in the core of AXE

Mats Jonback and Stefan Schultz

The APZ 212 40 is the first central processor of a new generation based on industry-standard microprocessors. It introduces a new *warm-standby/hot on-demand* system principle: the CP has two independent sides, each of which contains two processors—one that serves as an instruction processing unit and one that serves as a signal processing unit—which run as a two-way symmetrical multiprocessing computer. The APZ virtual machine handles ASA execution and is the middleware that guarantees telecommunications-grade availability.

The authors describe the APZ 212 40 hardware and software, the *warm-standby* concept for high availability, and differences and similarities between this and previous APZ processors.

Introduction

Being based on industry-standard microprocessors, the APZ 212 40 central processor (CP) is a milestone in Ericsson hardware implementation. The latest in a line of central processors, the APZ 212 40 is the first central processor of a new generation, and the platform for future multiprocessor solutions (Figure 1).

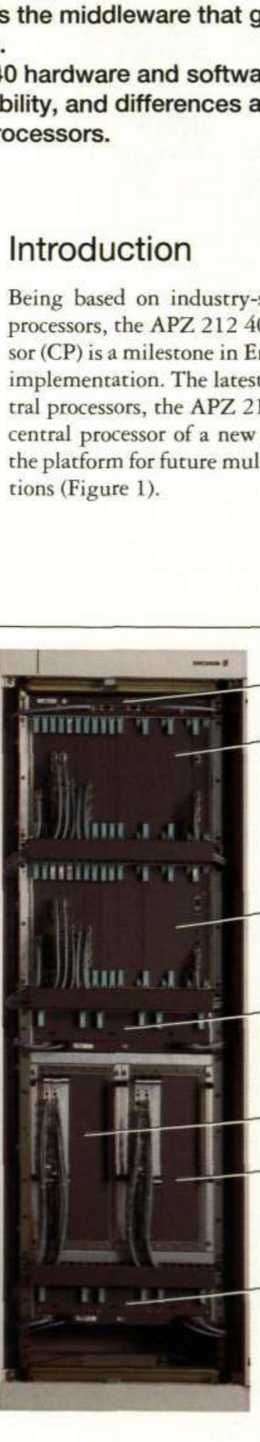


Figure 1
APZ 212 40 central processor cabinet. Note: CDU panel not shown.

The APZ 212 40 introduces a new system principle. In place of the parallel synchronous (lock-step) machine, we are introducing a *warm-standby/hot on-demand* principle. The CP has two independent sides, A and B, that are loosely coupled via a high-speed interconnect and a maintenance channel. Each side contains two processors: one that serves as an instruction processor unit (IPU), and one that serves as a signal processor unit (SPU), running as a two-way symmetrical multiprocessing (SMP) computer. The hardware and the operating system define a two-node cluster of two-way SMP computers.

The APZ virtual machine (VM), a new part of the system, handles the ASA execution and is the middleware required to achieve telecommunications-grade availability. It provides recovery and repair for hardware faults, software upgrades and the like while minimizing traffic disturbance. At the same time, Ericsson is introducing an industry-standard hardware equipment practice, cPCI, and a standard operating system. Parts of the CP core have been rewritten in standard C++ code using commercial development tools.

The APZ 212 40 will be deployed as a high-capacity central processor for all kinds of application. It offers processing capacity three times as great as the fastest APZ 212 30, and is fully backward-compatible—that is, it can run all existing AXE application software.

APZ 212 40 hardware

Hardware compatibility and structure

Processor technology is advancing rapidly, and Ericsson has decided to make the most of research and development in the mainstream of the computer industry by using industry-standard microprocessors. This reduces Ericsson's time-to-market, especially for upgrades, which can be made available as improvements are achieved in third-party development. At the same time, this makes it possible to concentrate in-house design efforts on developing services, improving robustness, and integrating solutions.

Each APZ 212 40 central processor magazine (CPUM, see Figure 3) contains five boards and a power module (Box A). The CPU board contains the industry-standard microprocessors. Connections are provided by

• the regional processor handler magazine

interface (RPHMI)—through cable connections;
- the base input-output (IO) board—through access ports and some general support functions; and
- the update bus board (UPBB)—through interfaces such as the new Ethernet bus and cable connections.

The new cPCI backplane is connected to the CPU backplane via a PCI interface module (PIM).

Ethernet and the interplatform network

To speed up central processor-to-APG40 communication, and thus reduce reload time, boot time, backup time, and the like, the APZ 212 40 uses direct communication

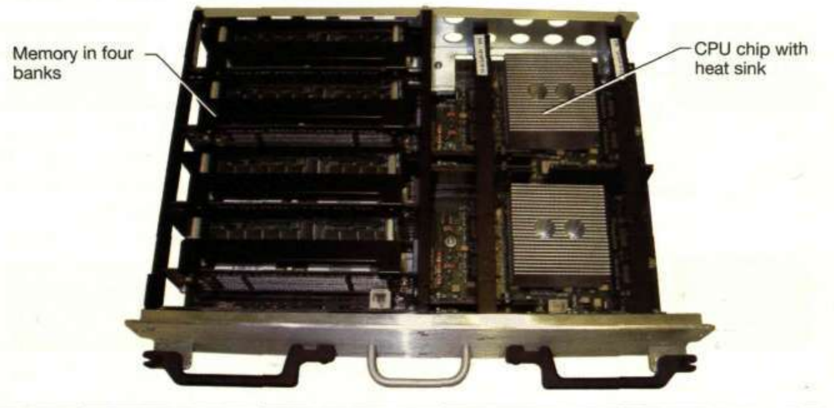


Figure 2
CPU magazine.

BOX A, APZ 212 40 BOARDS

CPU board
- 2 high-performance (GHz) microprocessors, each with 8 MB level 2 cache
- 8 GW (16 GB) SDRAM memory

RPHMI board
- Cable connections to the magazines (RPHB) for the A- and B-sides RPH-A and RPH-B
- Cable connection to the RPHMI on the other CP side for error information and for control of system states (WSB).

Base IO board
- Access port (connected to UPBB) for obtaining detailed low-level system-error information from the microprocessor system (CPU board)
- Some general support functions for the CPU board and power module

UPB board
- One 1 Gbit/s Ethernet optical fiber connection to the other CP side for updating
- Two 100 Mbit/s Ethernet cable connections to the adjunct processor over the interplatform network (IPN), connected via the IPNX Ethernet switch in RPHM
- One 10 Mbit/s Ethernet cable connection to the adjunct processor, a processor test bus (PTB) for access with the central processor test (CPT) command interface, connected via the IPNX Ethernet switch in RPHM
- Cable connection to the central display unit (CDU) panel at the top of the cabinet
- Cable connections to the CPU magazine fan units for monitoring and controlling the fans

The PCI interface module (PIM) card (not visible from the front) includes functionality for connecting the CPU backplane with the cPCI backplane.

Power module (in bottom left part of the CP subrack)
- Duplicated –48V external connection
- DC/DC converters for internal system voltages

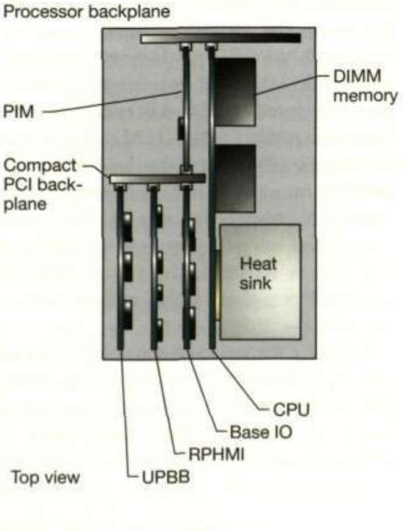


Figure 3
CPU subrack.

via Ethernet. Figure 4 shows the Ethernet switch used to bridge communication between the APG40, CP, AXD 301, and other equipment connected to the interplatform network (IPN). Communication between the central processor and APG40 uses cross-connected 100 Mbit/s Ethernet links. The two CP sides are connected by end-to-end 1 Gbit/s Ethernet link optical fiber.

The central processor test (CPT), which is used for maintenance and repair activities, is also connected to the Ethernet switch. All Ethernet connections to and from the central processor are connected to the UPBB board via the Ethernet switch. The use of Ethernet for CP-to-IO (APG40) communication also decreases the load on the regional processor handler bus (RPHB).

Cross-connected regional processor handler bus

As in previous models of the APZ, an RPHB runs between the signal processor unit and the RPHM. However, unlike previous APZs, the RPHB is cross-connected in the APZ 212 40 for redundancy. This connection is between the RPHMI board in the central processor unit magazine and the regional processor input-output (RPIO) board in the RPHM. With the exception of the RPIO board, the two RPHMs are identical to those developed for the APZ 212 30. The RPIO board has been updated to support the cross-connection.

Of the two CP sides, one is always the executive side; the other is on standby. The executive CP side has control over one or both of the RPHMs. In a normal traffic situation, the executive side controls both RPHMs. Because one of the two RPHMs is always active (used in traffic), if an error occurs, the standby RPHM can take over immediately. The RPHMs can also be separated, and each can be assigned to its own CP side.

To avoid accidents, such as the inadvertent disconnection of the active RPHB, the state of the RPHMs is shown on the CP control and display unit (CDU) panel at the top of the cabinet.

APZ 212 40 software

Layered software structure

The structure of the APZ 212 40 can best be described as layered (Figure 5). From the bottom, we have the CPU hardware platform, with microprocessor, memory, firmware console, privileged architecture library (PAL) code, and other functions. On top of that, we find a suitable operating system that interfaces with the hardware and the application—in this case, software developed by Ericsson.

The APZ virtual machine is a PlexEngine that executes traffic. It is the interface between the existing application software and the hardware platform, similar to a microprogram. The virtual machine communicates with the CPU platform via two additional thin layers: an operating system application interface (OS API) and the hardware abstraction layer (HAL). These two layers are intended to make the APZ less dependent on a particular microprocessor architecture or operating system. The idea is that if the operating system or hardware platform is replaced, the virtual machine and the layers above it can remain more or less intact.

The APZ virtual machine accesses the resources and services provided by the underlying CPU hardware and operating system via HAL and OS API. In Figure 5, PLEX/ASA indicates where in this layered structure the APZ software units and the

BOX B, TERMS AND ABBREVIATIONS

AMU	Automatic maintenance unit	IP	Instruction processor
AOT	Ahead of time	IPN	Interplatform network
AP	Adjunct processor (external IO system used for man-machine communication, external boot media and charging output)	IPNX	IPN switch
		JAM	Jump address memory
		JIT	Just in time
		MAU	Maintenance unit
APG40	The latest generation of adjunct processor most suitable for the APZ 212 40	MW 16	Megaword (1,048,576 words) 16-bit word length
		OS	Operating system (commercial)
ASA	Programming language assembler, also machine language in earlier APZ CPs	OS API	OS application program interface
		PAL	Privileged architecture library
BOOTP	BOOT protocol	PCI	Peripheral component interconnect
BUMS	Backup in the main store	PIM	PCI interface module
CDU	CP control and display unit	PLEX	Proprietary programming language for exchanges
CP	Central processor (includes RPHM, CPUM and cabling)	PS	Program store
cPCI	Compact peripheral component interconnect	PTB	Processor test bus
		RP	Regional processor
CPIO	Central processor input-output	RPHB	RP handler bus
CPT	Central processor test	RPHM	RP handler magazine (subrack)
CPU	Central processor unit. In this machine, same as CPUM	RPHMI	RPHM interface board
		RPIO	Regional processor input-output
CPUM	CPU magazine (subrack), one CPU side	RS	Reference store
DIMM	Dual in-line memory module	SDRAM	Synchronous dynamic random access memory
DS	Data store		
FTP	File transfer protocol	SMP	Symmetrical multiprocessing
GW 16	Gigaword (1,073,741,824 words) 16-bit word length	SP	Signal processor
		UPBB	Update bus board
HAL	Hardware abstraction layer	VM	Virtual machine
IO	Input-output	WSB	Working state bus

rest of the application can be found. In the actual target machine, only ASA code is executed. PLEX-to-ASA compilation is done off target.

The APZ virtual machine employs the operating system application interface to communicate with the operating system and to use its functions when required. The OS API layer is intended to make the system less dependent on particular operating system platforms. If, for some reason, it becomes beneficial to change the platform, the main changes will be made in the OS API. The central processor design is based on a commercial CPU, making it feasible to use a commercial operating system that

- is available for the chosen microprocessor architecture;
- supports multiple processors;
- uses 64-bit processor architecture;
- supports dynamic-link libraries; and
- has memory protection between processes and threads.

Two ASA compilers

The APZ 212 40 CP compiles ASA binary code using two methods: just-in-time (JIT) and ahead-of-time (AOT). Thus, the APZ includes a JIT compiler and an AOT compiler to compile the ASA binary code into native CPU code. This is transparent to the user.

Just-in-time compiler

The APZ virtual machine calls the JIT compiler when there is no compiled code—that is, when neither AOT-compiled nor JIT-compiled code is available. The JIT program compiles and stores a "compilation unit," which is a small part of the ASA code—a sequence of instructions out of which there are no jumps (short compilation time is the main focus of the JIT compiler). The next time that piece of code is called, the compiled code will be executed directly. If anything happens that affects the ASA binary code, such as a correction being inserted or a trace on an instruction address using the test system, the generated code is invalidated and must be recompiled the next time it is called.

Compared to earlier versions of APZ, the JIT-compiled code has full register coherency. Consequently, the JIT compiler is the one to use during troubleshooting. It has full support for the test system and the current correction handling—the code is JIT-compiled until the AOT compilation is finished.

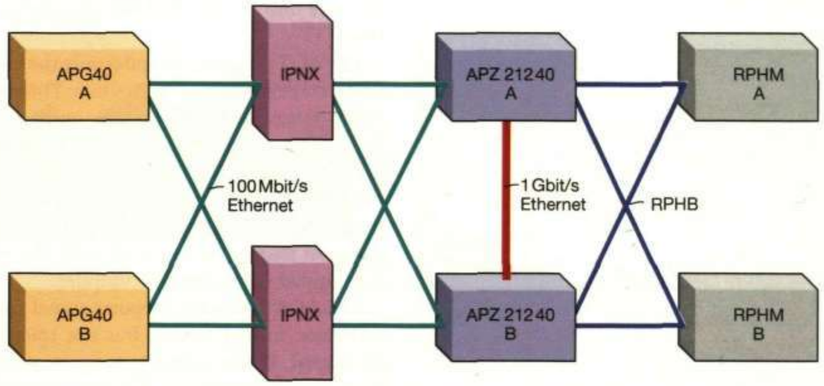


Figure 4
Ethernet links and cross-connected RPHM.

Ahead-of-time compiler

The AOT program compiles entire software units of ASA binary code in background execution, including all currently loaded ASA corrections. The AOT compilation is performed in a separate process on the SPU processor. Its purpose is to generate code that is as efficient as possible. The AOT compiler is allowed longer execution time to produce code. Thus, unlike the JIT compiler, it can optimize over an entire ASA

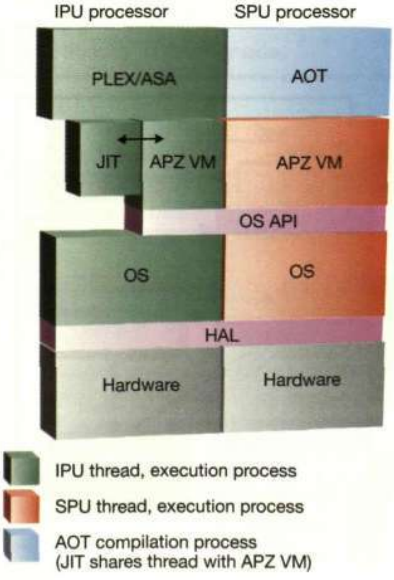


Figure 5
Central processor structure of the APZ 212 40.

software unit. Capacity is the main focus of the AOT.

The AOT program compiles only the most frequently executed software units. These are either selected by the system or decided manually—that is, they are preset. The number of blocks is estimated to be 10-15% of the total number, or around 150 blocks. This should be sufficient to cover at least 90% of the most frequently executed code, the rest is compiled using the JIT compiler.

The AOT compiler supports signal trace and some forlopp traces, but not trace on out-signal. If any other traces are requested by the test system, the JIT compiler will take over program execution. The *trace on every jump* option using the test system is no longer available.

The switch between AOT-compiled and JIT-compiled code is transparent to the user. For instance, when a trace is activated, the system switches from AOT-compiled to JIT-compiled code. When the trace is deactivated, the system automatically switches back to the AOT-compiled code.

Record-oriented data-store architecture

The APZ 212 40 introduces a record-oriented data store (DS) architecture. Record-oriented data allocation is based on the observation that after one piece of data has been read into memory, another variable in the same record (with the same pointer value) will likely be read soon after.

In contrast to existing central processors, modern microprocessors rely heavily on caching, so they spend much less time reading data if they have just read the preceding data. Thus, when the APZ 212 40 reads data, some data that follows immediately after is also read and put in the cache memory. The introduction of the record-oriented data-store architecture only affects a few PLEX or ASA software units in the APZ operating system and not the application.

APZ virtual machine

The APZ virtual machine is the compiled and linked module of C++ code that executes the application. The APZ virtual machine, which is compiled into native code off-line, is the first module loaded after the operating system when the system is powered on. This microprogram substitute takes care of the incoming signals from the regional processors (RP) and from the IPN network—it schedules, dispatches, and distributes the signals that result from the jobs. The APZ virtual machine also offers special services or support functions to the software layers above it (Figure 5).

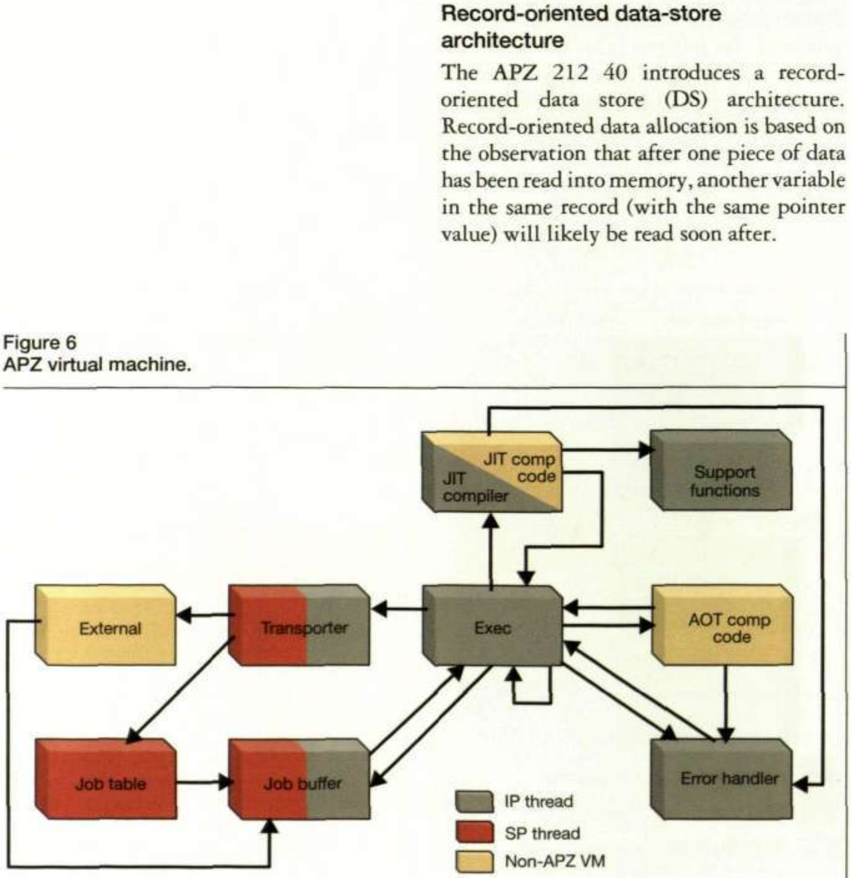
The APZ virtual machine executes in a process with two threads, referred to as the instruction processor (IP) thread and the signal processor (SP) thread. The instruction processor thread handles all job scheduling (Figure 6). The main loop of the instruction processor thread starts by polling the job buffers for incoming jobs. The jobs are then dispatched and executed. The main parts of the instruction processor thread are

- Exec—which is the virtual machine kernel, dispatcher, scheduler, and signal distributor;
- error-handling and recovery;
- the run-time log;
- the JIT compiler; and
- an interface to the AOT-compiled code.

The signal processor thread handles external communication and controls the job table. Basically, it runs in an infinite loop checking the external and IPU interfaces for new signals or messages to store in job buffers—for subsequent execution in the instruction processor thread. The main parts of the signal processor thread are

- job-table scanning;
- error-handling and recovery;
- communication support (transports and external signaling);
- RP signaling; and
- IPN signaling.

Figure 6
APZ virtual machine.



Warm-standby concept for high availability

The CPU sides do not run in parallel synchronous (lock-step) mode. Instead, a *warm-standby/hot on-demand* principle is applied. In normal traffic situations, one CP side executes and the other side is on standby (normal). The same system that is executing on the executive side is preloaded into memory on the standby side. The standby side is also regularly updated with transient data.

Likewise, the standby side is updated through the automatic backup function. When a side is scheduled to be switched or a function is scheduled to be replaced—for example, when testing, performing maintenance, or adding hardware—the standby side is "heated up"; that is, the entire memory is copied from the executive side to the standby side, so that the standby side becomes a mirror image of the executive side. A 1 Gbit/s Ethernet link (the updating bus) between the two CP sides provides the capability to heat up the standby side. The copy procedure can be divided into two phases (Figure 7): background copy (raw copy) and traffic-oriented copy (high-level copy). Memory is copied page by page using a fault-on-write mechanism. This means that a copied page is set to write protect. Then, when a write order to that page is received, the page is logged before it is written. The page is copied again later (Figure 8).

Measurements and estimates have been made for copying data between CP sides. With one 1 Gbit/s link, the capacity will drop marginally for a period less than 15 seconds. At the end of the copy interval, the system is frozen for around 0.5 seconds, allowing the CP states and the APZ virtual machine to be copied. During this interval, traffic from the RPs is buffered, so no data is lost.

If the standby side is put into operation because of a debilitating hardware fault in the executive side, a restart with reload is ordered. In the normal state, a system will already have been loaded, so what actually occurs is a nearly instantaneous reload followed by a large restart. The command log is later read from the adjunct processor without affecting downtime.

The advantage of the *warm-standby/hot on-demand* concept is in-service performance during software recovery, which fully compensates for any negative in-service performance from CPU hardware recovery. As with previous APZs, planned events do not result in any traffic disturbance.

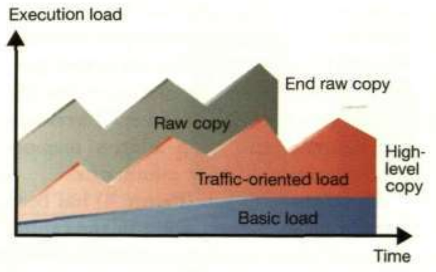


Figure 7
"Heating up."

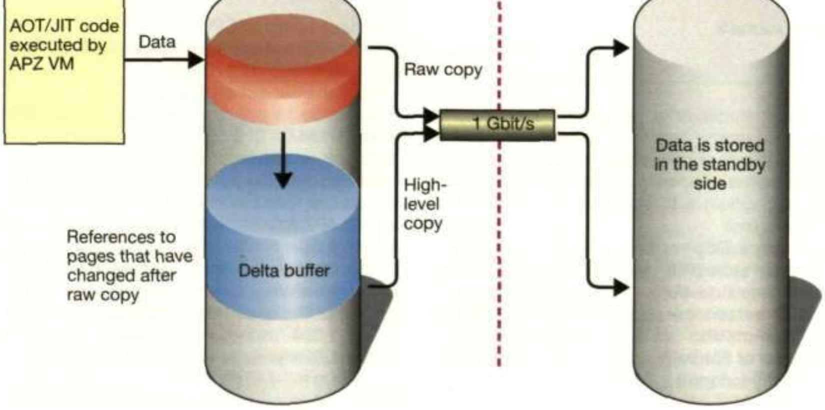
Previous APZ processors—similarities and differences

New concepts in APZ 212 40

The *warm-standby* concept, which is new to APZ design, obviates the need for maintenance hardware to detect matching errors between CP sides. Consequently, the APZ 212 40 has no separate maintenance unit (MAU).

As explained above, the APZ central processor is based on a commercial microprocessor that requires a substitute for the old

Figure 8
Data transfer to standby side.



microprogram. The substitute is found in the APZ virtual machine and the new compilers. Hence, the information that the system provides in fault situations and other specific states differs from that of earlier CP versions. For example, the jump address memory (JAM) contains different information.

An Ethernet switch (Figure 4) has been introduced to connect ports in the central processor to the adjunct processor and other systems, such as the AXD 301.

Hardware dissimilarities

The smallest replaceable hardware unit in the new central processor is the entire CPUM—that is, a complete CP side. For the RPHM, the smallest replaceable hardware unit is the same as in previous versions: individual boards. The CP working state logic is implemented on the RPHMI board; a special working state bus (WSB) cable connects the RPHMI boards in the two CPU sides.

The memory can be configured in steps of 1 GB. The normal way of upgrading or increasing physical memory is to update a configuration file on the adjunct processor, replace the CPUM, add physical memory, and then use regular size-alteration commands to inform the applications that more memory is available. As usual, this process is ended with a backup.

A new CP control-and-display unit panel indicates which side is executing and which side is on standby. It also indicates which RPHM is handling traffic at any moment. The CDU panel indicates normal system state when the executing side is switching traffic and the standby side is ready to take over control. The standby side is ready to take over when

- it has the same system generation preloaded in system memory as the executive side; and
- when it is receiving transient data.

A new state (normal, NRM on the CDU panel) has been defined for the standby side to indicate when these criteria are satisfied.

Software dissimilarities

The C and C++ parts of the system are treated separately from the rest of the system. These separate files on the adjunct processor are loaded using BOOTP and FTP. They are not part of the system backup.

To ensure short reload time, backup in main store (BUMS) must be active. When

BOX C, SYSTEM CHARACTERISTICS AND TECHNICAL SPECIFICATIONS

Equipment practice	BYB 501
Footprint	600 x 400 mm (23.62 x 15.75 inches)
Height	1800 mm (70.87 inches)
Cabinet subracks	2 CPU subracks (one each for CPU-A and CPU-B)
	2 RPH subracks (RPHM) (one each for sides A and B)
Electromagnetic compatibility (EMC) class:	Fulfills EMC Class B
Cooling type	Fans
CPU subrack	2 x 3 packages with 2 fans each
RPH subrack	1 package with 3 fans each
CPU subrack (one of the CP sides)	1 processor board including memory 1 interface board to RPHM 1 interface board to other CP side 1 base IO board
RPH subrack	1 interface board to CPUM Max. 16 RPH boards (parallel type) or Max. 8 RPH boards (serial type) 1 IPNX Ethernet switch board
Memory size	8 GW SDRAM
Power supply	–48 V
Power consumption	~1400 W

System limit	APZ 212 20	APZ 212 30	APZ 212 40
Data store (DS)	1.5 GW	4 GW	7.7 GW
Program store (PS)	64 MW	96 MW	256 MW
Reference store (RS)	2 MW	16 MW	16 MW
	32 bits	32 bits	32 bits
Number of RPs	1,024	1,024	1,024
Number of EMGs	1,024	1,024	1,024
Capacity increase from the APZ 212 20	1.0	3.5	10

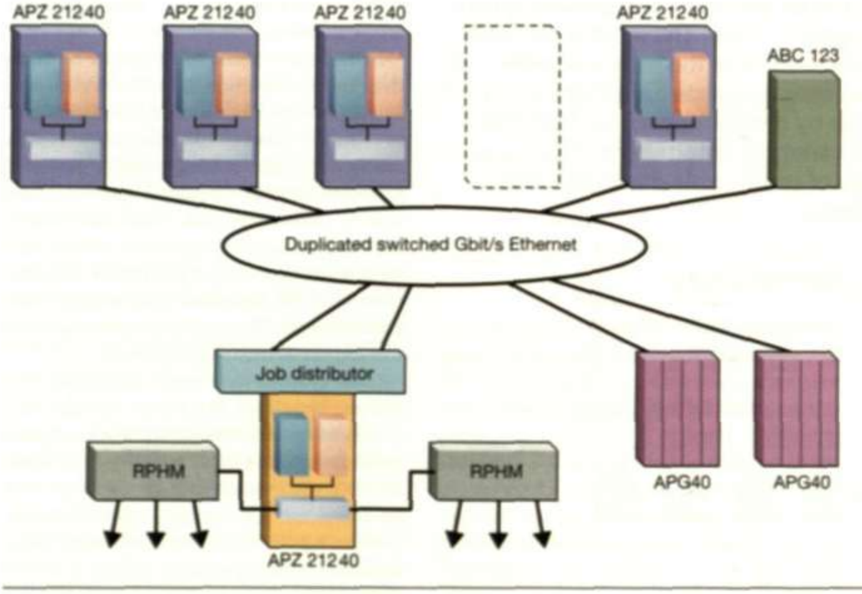


Figure 9
Example of a node in the future.

backup of the system data is first made, it is stored in the backup area of primary memory. As part of the backup function, the system is then transferred to external media—the APG40. The automatic backup function is routed via the standby side to the adjunct processor. In the standby side, an automatic reload is initiated to keep the standby side "warm."

When the APZ VM, JIT, HAL, OS API, AOT, or OS needs to be updated, the new version is loaded as a normal file onto the adjunct processor. A boot is then initiated in the CP standby side, and the new versions are loaded. The CP sides (roles) are then switched and the sequence is repeated.

Conclusion

The new APZ hardware architecture makes the most of increased processing speed through the use of industry-standard microprocessors and innovations such as internal Ethernet buses. The mainstream sourcing of certain processors allows AXE users to benefit from general advances in microprocessor technology.

The latest version of Ericsson's APZ central processor introduces several innovative concepts, but backward compatibility with previous versions of AXE hardware and software has been ensured, thereby safeguarding operator investments.

BOX D, KEEPING PACE WITH FUTURE DEVELOPMENTS

With its architecture and use of commercial microprocessors, the APZ 212 40 allows us to take advantage of rapid evolution in the computer industry. Development will keep pace with the mainstream computer industry and enable improvements in characteristics, such as capacity and robustness, with limited changes in software. The APZ central processor design is also open enough to simplify moves between different microprocessor suppliers and the operating systems they support.

The introduction of the interplatform network (IPN) via directly connected Ethernet makes the APZ 212 40 the obvious platform for future multiprocessor systems. In a distributed multiprocessor system, replicated processors can improve in-service performance and simplify handling when adding capacity to the node. Figure 9 shows an example of a multiprocessor system.

A powerful duplicated processor handles job distribution between a number of replicated processors (call handlers). The APG40, for IO communication, is connected to the Gigabit Ethernet used for communication. Other equipment can be attached to the network based on the needs of applications.

Cable modems—Broadband highway to the home

Juan Figueroa and Bill Guzek

Cable operators are eyeing a huge business opportunity in providing residential customers with high-speed Internet access. The market for cable modems is expected to increase dramatically over the next few years. This increase will be the result of rising user demands for access to the services and capabilities made possible by greater speed, such as high-quality voice, video on demand, and a variety of entertainment services.

The authors describe the technology of using cable plants to bring high-speed Internet access into the home.

Introduction

Surfing the World Wide Web typically offers residential Internet users a click-and-wait experience rather than an interactive extravaganza. Because connection speeds are typically limited to 53 kbit/s or less, frustrated residential online users are demanding higher-speed connections. However, in spite of the slow narrowband speeds currently available through existing dial-up telephone modem connections, residential Internet and online usage continues to grow rapidly.

Local service providers currently offer residential ISDN services that provide connection speeds up to 128 kbit/s. To offer downstream speeds in excess of 1.5 Mbit/s, service providers will have to look to alternative solutions, such as digital subscriber line (DSL) technologies, faster downstream data connections from direct broadcast satellite (DBS), fixed wireless access, and high-speed cable modems.

Broadband coaxial cable passes by more than 105 million homes in North America, and more than 75 million of these subscribe to cable TV. Coaxial cable connections offer nearly universal coverage and a powerful platform for providing high-speed data access to residences and small businesses. However, to support advanced communications services, one-way cable television systems must be upgraded into modern two-way networks. This is a technically complex and capital-intensive proposition.

Cable systems were originally designed to deliver broadcast television signals efficiently to subscribers' homes. To ensure that consumers obtain cable service via the same TV sets with which they receive over-the-air broadcast TV signals, cable operators recreate a portion of the over-the-air radio frequency (RF) spectrum within a sealed coaxial cable line or CATV network designed and used for cable TV distribution. The system must be upgraded with bidirectional amplifiers in the cable-distribution or CATV network before signals can flow in two directions. Most CATV networks are a hybrid of fiber and coaxial cables. Signals are passed through fiberoptic cables from the head-end center to locations near the subscriber. The signals are then transmitted in coaxial cables that run to the subscriber premises. Higher-frequency signals flow toward the subscriber and lower-frequency signals flow toward the broadcasting head-end.

BOX A, TERMS AND ABBREVIATIONS

ADC	Analog-to-digital converter	ISDN	Integrated service digital network
ARP	Address resolution protocol	ISP	Internet service provider
ATM	Asynchronous transfer mode	LAN	Local area network
BER	Bit error rate	LLC	Logical link control
BPI	Baseline privacy	MAC	Media access control
Cable network	Refers to the cable television plant that would typically be used for data-over-cable services.	MCNS	Multimedia Cable Network System Partners Ltd.
CATV	Cable TV	MSO	Multiservice operator
CM	Cable modem	NAT	Network address translation
CMCI	Cable modem-to-CPE interface	NSI	Network-side interface
CMTS	Cable modem termination system	OSI	Open systems interconnection
CPE	Customer premises equipment	PC	Personal computer
DBS	Direct broadcast satellite	PDU	Packet data unit
DHCP	Dynamic host configuration protocol	QAM	Quadrature amplitude modulation
DOCSIS	Data-over-cable service interface specification	QoS	Quality of service
DSL	Digital subscriber line	QPSK	Quadrature phase-shift keying
DSP	Digital signal processor	RF	Radio frequency
FEC	Forward error correction	RFC	Request for comments
Head-end	Central distribution point for a CATV system	RFI	RF interface
HFC	Hybrid fiber-coaxial	RISC	Reduced instruction set computing
ICMP	Internet control message protocol	SID	Service ID
IEEE	Institute of Electrical and Electronics Engineers	SNAP	Subnetwork access protocol
IF	Intermediate frequency	SNMP	Simple network management protocol
IP	Internet protocol	SU	Subscriber unit
		TFTP	Trivial file transfer protocol
		UDP	User datagram protocol
		USB	Universal serial bus
		VoIP	Voice over IP

Cable plant architecture

Several elements are required to bring high-speed data over cable (Figure 1). A cable plant consists of a hub or ring of hubs. In a typical cable plant, one or more of these hub facilities serve as the collection points. The head-end hub gathers television signals from various sources—primarily satellite TV transponders. The TV signals are picked up, decoded, and down-converted to selected channels. These channels are then combined into a local fiberoptic network for local distribution or placed on a higher-capacity optical network between the regional hubs. One or more of these hubs also serve as the main interface to the Internet and are tied to it via high-speed optical links.

Each hub has a head-end that uses smaller fiber bundles to distribute the television signal to smaller local distribution amplifiers. The fiber is then terminated and the signal is converted back into electrical signals that are sent over coaxial cable into the neighborhood (Figure 2). This scheme of using a mixed fiber and coaxial cable distribution scheme is called a hybrid fiber-coax (HFC) system. The amplifiers in this part of the network must be bidirectional.

In the past, the cable plant typically distributed signals from the head-end system to the customer premises using coaxial cable. Today, however, nearly all operators, including multiservice operators (MSO), have upgraded or are in the process of upgrading their plants to hybrid fiber-coax systems. These systems provide clean, high-quality signals to the neighborhood without replacing all of the cable to the customers' home.

To introduce data into the system, the IP infrastructure must be overlaid on existing systems. Connections to the Internet backbone via concentrators and routers are provided via high-speed optical connections, typically OC-8 and higher-speed connections. These connections are brought into the head-end hub and distributed via routers and optical links to various regional distribution hubs in the MSO network. The IP network is tied into a special adapter called a cable modem termination system (CMTS), which consists of one or more cable modem line cards (CMLC). CMLCs convert the IP data stream into downstream (to the home) and upstream (from the home) RF signals. The downstream signals are sent through an up-converter, which puts them on a specific channel and combines them with the other standard TV signals.

The upstream signals are collected from the subscribers. In a properly designed system, there are several upstream channels for each downstream channel. This is because the upstream data is transmitted at a slower rate than the downstream data. The traffic must be engineered to provide adequate service to users. Voice, streaming video, and gaming services use the greatest amounts of bandwidth. The downstream data rate—approximately 30 to 40 Mbit/s—can be shared by some 500 to 2,000 users. The upstream data rate is approximately 8 Mbit/s per channel. Proxy servers and data cache servers are also employed at the local hub to improve system performance.

Cable service providers in North America and Europe have different transmission requirements for channel width, upstream bandwidth, downstream bandwidth, channel center frequencies, and power limits per channel. After the data-over-cable service interface specification (DOCSIS) had been successfully certified in North America and Europe, cable operators adapted it and most of the specifications set by DOCSIS for Europe. Euro-DOCSIS was formed to address European cable operators, Internet service providers (ISP), and end-user needs (for a

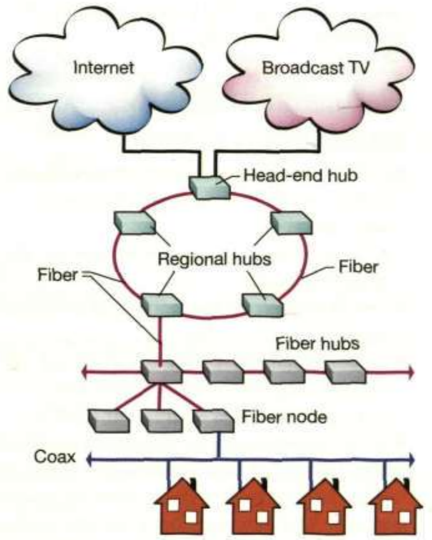


Figure 1
Architecture of the regional hub HFC network.

Figure 2
Architecture of the cable system.

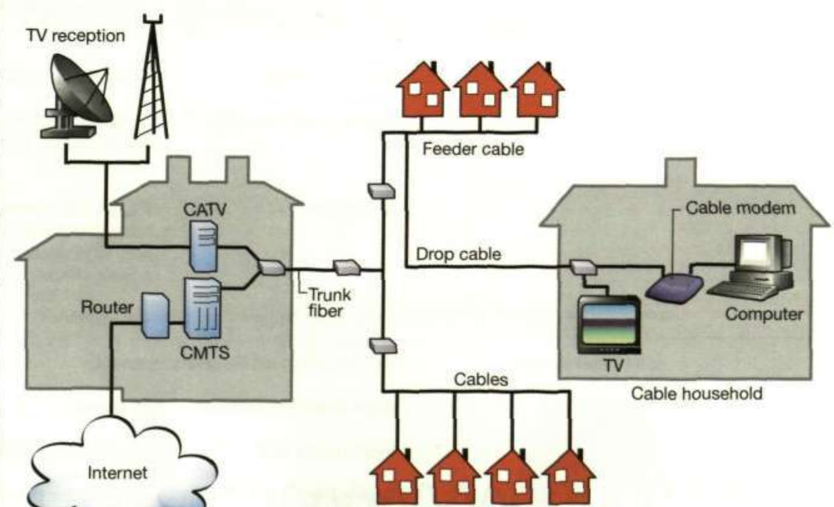


TABLE 1, PROPERTIES OF DOWNSTREAM SIGNALS

Frequency	42 to 850 MHz in North America; 65 to 850 MHz in Europe
Bandwidth	6 MHz in North America; 8 MHz in Europe
Modulation	64 QAM with 6 bits per symbol (normal) 256 QAM with 8 bits per symbol (faster, but more sensitive to noise)

TABLE 2, DATA RATES PER MODULATION SCHEME

	64 QAM	256 QAM
6 MHz	31.2 Mbit/s	41.6 Mbit/s
8 MHz	41.4 Mbit/s	55.2 Mbit/s

comparison of DOCSIS and Euro-DOCSIS, see Tables 1, 2 and 3).

The cable modem termination system

A CATV network consists of six major parts:

- a cable modem termination system;
- a trunk cable;
- a distribution system or feeder in the neighborhood;
- the drop cable to the home and in-house wiring;
- a cable modem, which is connected via coaxial cable and optical fibers; and
- customer premises equipment (CPE).

Figure 3 shows a simple diagram of the data traffic through a data-over-cable system as well as the external interfaces of the key components. The interface between the wide-area network (WAN) and the CMTS is called the network-side interface (NSI); the interface between the CMTS and the cable modem is called the radio frequency interface (RFI); and the interface between the cable modem and customer premises equipment is called the cable modem-to-CPE interface (CMCI). The user connects to the Internet service provider through the WAN interface.

The system shares media for upstream and downstream transmissions. The two most common problems in this architecture are

- the need to control access to shared resources, particularly upstream bandwidth; and
- the injection of ingress noise in the upstream direction.

Traditional coaxial cable systems typically operate with 330 or 450 MHz of capacity, whereas modern hybrid fiber-coax systems have been expanded to 850 MHz. The terms CMTS and head-end are commonly interchanged in contexts that refer to the equipment responsible for communications in the cable network. In reality, the CMTS could be part of the head-end—the equipment from which multiservice operators broadcast television content.

Logically, downstream video programming signals—that is, from the CMTS to the cable modem—begin around 50 MHz, the equivalent of channel 2 for over-the-air television signals. The portion of the spectrum between 5 and 42 MHz is usually reserved for upstream communication—that is, from the cable modem to the CMTS. In North America, each standard television channel occupies 6 MHz of spectrum (in Europe, 7 or 8 MHz). Thus, a traditional cable

TABLE 3, ELECTRICAL INPUT TO CABLE MODEM

Parameter	North American value	European value
Center frequency	91 to 857 MHz ± 30 kHz	112 to 858 MHz ± 30 kHz
Level range (one channel)	-15 dBmV to 15 dBmV	43 to 73 dBµV for 64 QAM 47 to 77 dBµV for 256 QAM
Modulation type	64 QAM and 256 QAM	64 QAM and 256 QAM
Symbol rate (nominal)	5.056941 Msym/sec (64 QAM) and 5.360537 Msym/sec (256 QAM)	6.952 Msym/sec (64 QAM) and 6.952 Msym/sec (256 QAM)
Bandwidth	6 MHz (18% square root raised cosine shaping for 64 QAM and 12% square root raised cosine shaping for 256 QAM)	8 MHz (15% square root raised cosine shaping for 64 QAM and 15% square root raised cosine shaping for 256 QAM)
Total input power (40-900 MHz)/ (80-862 MHz) for Europe	<30 dBmV	< 90 dBµV
Input (load) impedance	75 ohms	75 ohms
Input return loss	> 6 dB (88-880 MHz)	> 6 dB (85-862 MHz)
Connector	F-connector per [IPS-SP-406] (in common with the input)	F-connector per [IPS-SP-406] (in common with the input)

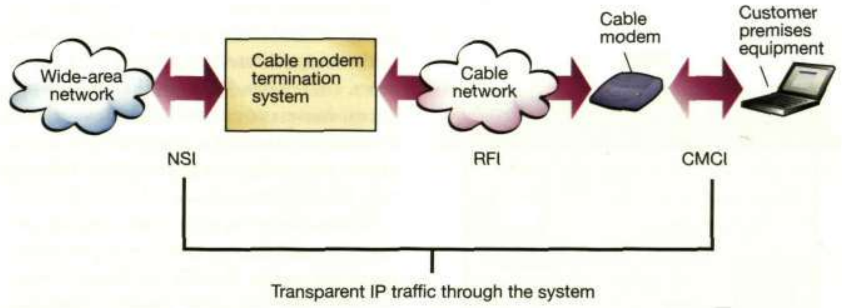


Figure 3
Data traffic through data-over-cable system.

system with 400 MHz of downstream bandwidth can carry the equivalent of 50 to 60 analog TV channels, and a modern hybrid fiber-coax system with 700 MHz of downstream bandwidth has capacity for some 80 to 110 channels.

To deliver data services over a cable network, one television channel (in the 50 to 865 MHz range) is typically allocated for downstream traffic to homes, and one or more channels (in the 5 to 42 MHz band) are used to carry upstream signals. Depending on the availability and the business viability of other channels, the number of cable modem users supported by a head-end can be incremented by commandeering other channels for data and IP transmission. When a channel is used for data, it cannot be used for other conventional, revenue-generating broadcasts, such as commercial TV or pay-per-view services. However, new Inernet-based revenue-generating services can now be offered on that channel.

Figure 4 shows the modulation and demodulation protocols as well as frequency ranges for the CATV system. It also shows the bandwidth and effective bit rates. Figure 5 compares upstream and downstream transmission.

Using 64 quadrature amplitude modulation (QAM) transmission technology, a single downstream 6 MHz television channel can support up to 27 Mbit/s of downstream data throughput from the cable head-end. Speeds can be boosted to 36 Mbit/s using 256 QAM. Depending on the spectrum allocated for service, upstream channels from the home can deliver 0.5 to 10 Mbit/s using 16 QAM or quadrature phase shift key (QPSK) modulation techniques. The upstream and downstream bandwidth is shared by active data subscribers who are connected to a given cable network segment, typically 500 to 2,000 homes on a modern HFC network.

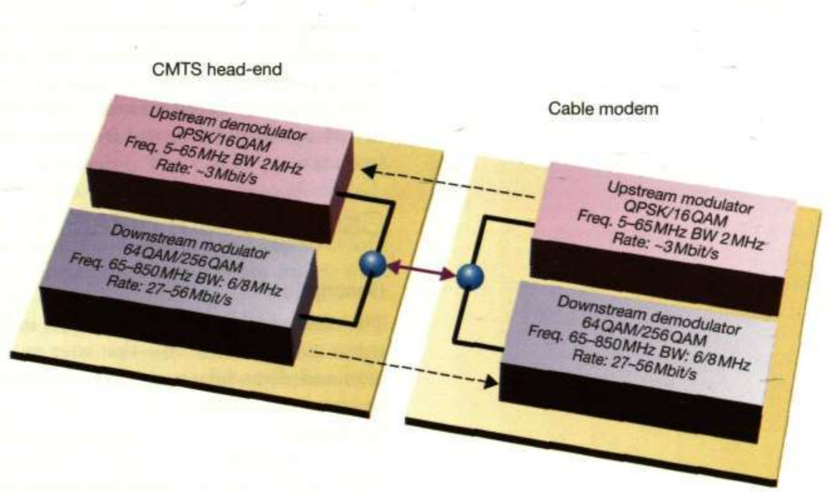
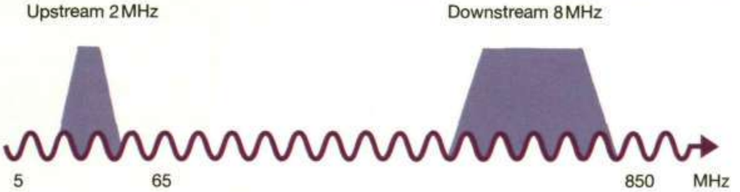


Figure 4
Functional block diagram of the cable modem termination system.

Figure 5
Frequency domains of the upstream and downstream channel positions.



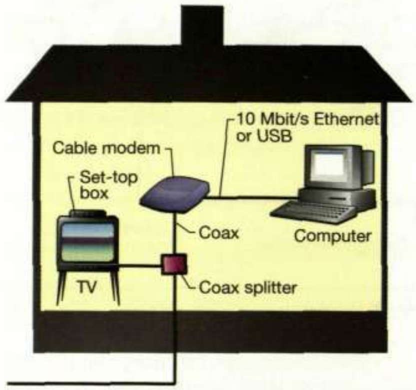


Figure 6
Home environment.

Depending on the network architecture and traffic load, an individual subscriber might experience access speeds from 500 kbit/s to 1.5 Mbit/s or more. Compared to dial-up alternatives, this is blazing performance.

Cable modem details

Description

The cable modem (CM) is a modem in the truest sense of the word—that is, it modulates and demodulates signals. Among its key components are

- a tuner;
- a demodulator;
- an encryption/decryption unit; and
- an upstream modulator.

Cable modems typically send and receive data in two slightly different ways. In the downstream direction, digital data is modulated and then placed on a 6 MHz (North America) or 8 MHz (Europe and PAL system) channel somewhere between 65 and 850 MHz. Upstream transmission is more challenging, since it tends to be very noisy in the 5 to 65 MHz region. Noise or interference is generated by amateur radio operators, citizen band (CB) radios, home appliances, loose connectors and poor cabling. Since cable networks form a tree or branch architecture, noise is aggregated as the signals travel upstream.

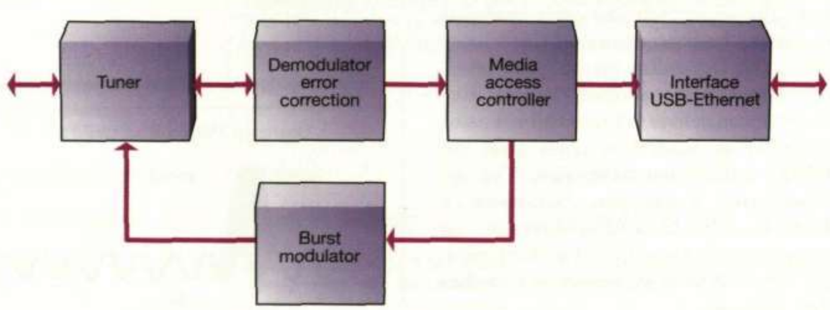
When a cable modem is installed, a power splitter and a high-pass filter might be necessary to isolate the TV set from "strong" signals from the cable modem. The filter also blocks upstream ingress noise in the low frequency band. Figure 6 shows how the signal is split from the main cable to TVs and cable modems in the home.

Figure 7 shows the major components of a cable modem implementation. The tuner, including the diplexer, connects directly to the CATV outlet, which provides upstream and downstream traffic to the rest of the cable modem. The tuner solution, which integrates the diplexer into a dual-conversion tuner and digital signal processor, demodulates 64 and 256 QAM signals. The output is a 44 MHz intermediate frequency (IF) signal that is fed into the analog-to-digital converter (ADC) inputs for demodulation and error correction. The media access controller then extracts data and sends it to the customer premises equipment via a universal serial bus (USB) or Ethernet interface.

The tuner also includes a line amplifier for the transmit function, controlled by the media access controller (MAC), which sends upstream signals at the level negotiated by the cable modem and head-end. According to DOCSIS, Reed-Solomon forward error correction (FEC) is recommended. This adds robustness as well as physical layer overhead, which translates into delay.

The heart of the protocol implementation resides in the MAC. The CMTS has control over the allocation of upstream and downstream bandwidth in a dynamic mix of con-

Figure 7
Cable modem architecture.



tention- and reservation-based transmission opportunities. Each cable modem has a unique 48-bit MAC address that is used for registration and authentication. This is entered in a table with the primary service ID (SID) and IP address, once the IP address is assigned. The CMTS then communicates to the modem with the assigned ID.

Significant processing power is needed for

- converting encoded signals into the format needed by the customer premises equipment; and
- conveying messages from the end-user to the destination point through the CATV system.

All available silicon solutions currently implement digital signal processors (DSP) for demodulating downstream transmissions, which are expected to have low bit error rate (BER). The DSPs handle demodulation, whereas regular processing power is devoted to implementing the key components of the cable modem transmission protocol. Powerful RISC processors enforce protocol functions.

Downstream transmission takes place in one of the 6 or 8 MHz channels between 65 and 850 MHz at 25 to 56 Mbit/s, as determined by the modulation scheme (Table 1). The raw data rate depends on the modulation and bandwidth (Table 2).

A symbol data rate of 6.9 Msym/s is used for 8 MHz bandwidth and 5.2 Msym/s is used for 6 MHz bandwidth. Due to error correction, framing and other overhead, the effective data rate is somewhat slower than the raw data rate. Since downstream data is received by all cable modems, the total bandwidth is shared by all active cable modems on the system. This is similar to Ethernet, except that the wasted bandwidth is much greater in Ethernet.

Operation

Initialization

Cable modem communications are set up through a series of initialization steps. After power-up, the modem scans for a downstream channel with which it can synchronize. The CMTS sends synchronization packets to generate a timing reference. Cable modems are synchronized and ranged so that they know when to begin transmission—in order to hit a specific minislot provided by the head-end. The CMTS controls access to slots by assigning specific "transmit opportunities" to ranges of minislots (a transmit opportunity can be contention- or reservation-based). A reserved slot is a timeslot that is reserved for a particular cable modem—that is, no other cable modem may transmit in that timeslot. The CMTS allocates timeslots through a bandwidth-allocation algorithm. The algorithm is vendor-specific and might differ considerably from vendor to vendor. Reserved slots are generally suited for longer data transmissions. After synchronization is complete, the cable modem receives the upstream parameters it needs to inform the CMTS of its presence on the network. The cable modem receives the upstream allocation information—which it uses to start the ranging process.

Due to the physical distance between the head-end and cable modem, the time delay (in milliseconds) can vary significantly. To compensate for delay, each cable modem employs a ranging protocol that effectively adjusts its internal clock. To do so, a number of consecutive timeslots (normally three) is reserved for each ranging process. The cable modem is instructed to transmit in the second timeslot. The CMTS measures the transmission and instructs the modem to adjust its clock as necessary. The two timeslots before and after create a gap to ensure that the ranging burst does not collide with other traffic.

Ranging is also used to coordinate the transmission power level of all cable modems, so that the upstream bursts arrive at the CMTS at the same level. A balanced transmission power level is essential for maintaining optimum performance of the upstream demodulator in the head-end. The attenuation from the cable modem to the head-end can vary by more than 15 dB.

Contention slots are open for all cable modems to transmit in. If two cable modems transmit in the same timeslot, their packets collide and the data is lost. When this occurs, the lack of positive ACK from the CMTS serves to indicate that the CMTS did not receive any data, and the cable modems retransmit at another, randomly selected time. Contention slots are normally used for very short data transmissions, such as a request for reserved slots, in order to transmit more data.

Ranging registration allows the cable modem to identify itself to the CMTS. It also allows the head-end

- to assign downstream and upstream frequencies;
- to set power levels; and
- to distribute other administration information necessary to manage the network.

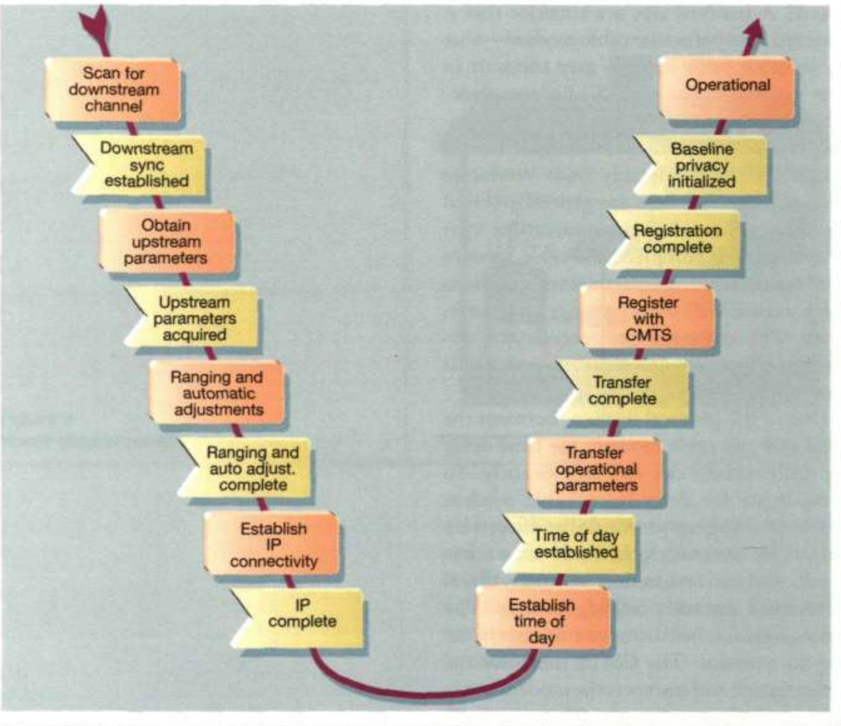


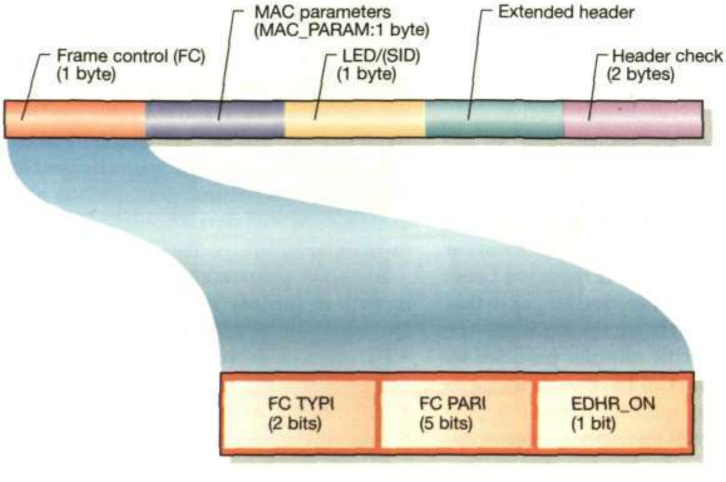
Figure 8
Initialization process.

The time-division multiple access (TDMA) transmissions from cable modems at various distances from the head-end must be coordinated so that all transmissions align with the boundaries of the head-end minislots. Discrepancies between individual modems and the head-end are caused by propagation delays in the CATV plant, the FEC interleaving function (a variable-depth interleaver supports latency-sensitive and latency-insensitive data), and processing time.

The CMTS informs the modem of the propagation delay after it receives the ranging request. Upstream frequency assignments can change at any time. The head-end ensures that the cable modem receives the new frequency assignment before listening for the modem's transmission on the frequency. After ranging is complete, the cable modem must invoke dynamic host configuration protocol (DHCP) mechanisms to obtain an IP address and continue the registration/configuration process (provisioning process).

The DHCP server responds with IP addresses, the name of the configuration file, and server addresses. After establishing a security association, the cable modem must download (via TFTP) a file with configuration parameters from the configuration server. The CMTS then checks that the configuration was obtained from a legitimate configuration server. Finally, the cable modem registers (Figure 8) with the CMTS and be-

Figure 9
Generic frame format.



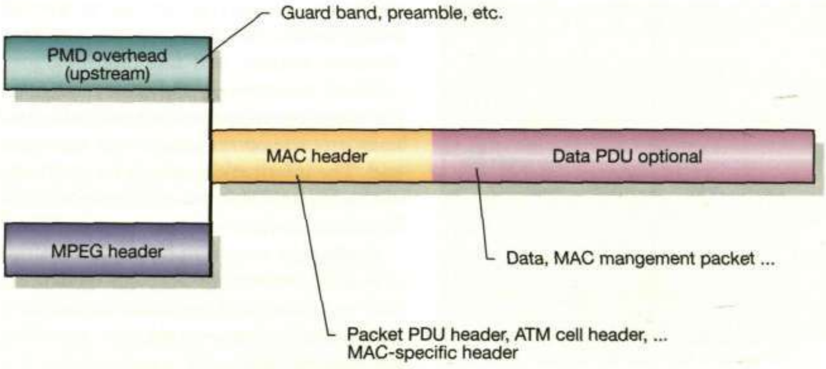


Figure 10
Generic MAC frame format.

comes operational. The CMTS must authorize the cable modem before it can send traffic into the network.

Each cable modem should contain a unique 48-bit MAC address (IEEE 802) that is assigned during the manufacturing process, and security information necessary to authenticate the cable modem.

When the initialization process has been completed, the customer premises equipment can communicate with the outside world through the CATV installation. The communication is bidirectional—downstream and upstream.

Upstream traffic

The term upstream is used to indicate the signal transmitted by the cable modem. Upstream traffic always occurs in bursts, so that many cable modems can transmit in the same frequency. Of the two modulation schemes (QPSK and 16 QAM), 16 QAM (four bits per symbol) is the fastest, but it is also the most sensitive to ingress noise. The upstream direction is characterized by

- a flexible and programmable cable modem under the control of the CMTS;
- frequency agility;
- time-division multiple access;
- QPSK and 16 QAM modulation formats;
- support for fixed-frame and variable-length packet data unit (PDU) formats;
- multiple symbol rates;
- programmable Reed-Solomon block coding; and
- programmable preambles.

When an Ethernet packet arrives at the cable modem from customer premises equipment, it is encapsulated in a MAC packet with a PDU header. The cable modem assesses bandwidth allocation for transmission opportunities and sends the MAC frame when allowed. Transmission opportunities have two basic components: number of mini-slots, and physical layer characteristics.

At the other end, the CMTS receives the packet, removes the header, and then forwards the packet to another cable modem through the radio frequency interface or to the WAN through the network side interface. Figure 9 shows a generic frame format; Figure 10 shows a generic MAC frame format. There are three types of MAC header: packet PDU MAC header, ATM cell MAC header, and MAC-specific header.

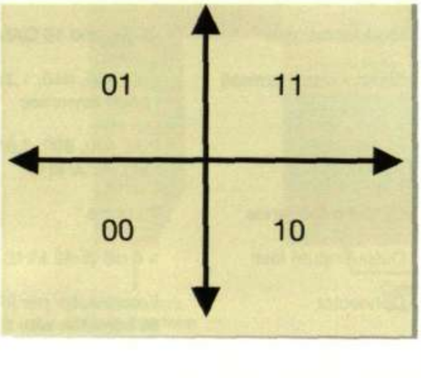


Figure 11
QPSK symbol mapping.

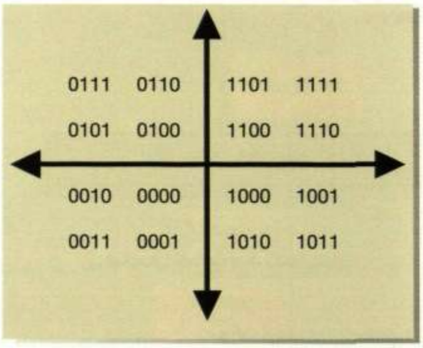


Figure 12
16 QAM symbol mapping.

In the upstream direction, the digital signal is encoded to QPSK or 16 QAM, converted into an analog signal, and sent to the tuner for transmission in the 5 to 65 MHz bandwidth. Figures 11 and 12 show the upstream symbol mappings. Figure 12 represents differential-coded symbol mapping.

Table 4 summarizes the characteristics of the upstream packets generated by the cable modem.

Downstream traffic

The RF signals from the CMTS to the cable modem are encoded in either 64 or 256 QAM. The symbol mappings resemble those shown in Figure 12, but with six-bit encoding for 64 QAM with 16 codes per quadrant, and eight-bit encoding for 256 QAM with 64 codes per quadrant.

There are several stages in the downstream transmission of packets from the CMTS to the cable modem.

When powering up, the cable modem extracts information that regulates when and how it is to communicate with the head-end. The information needed for the initial contact is available in the frames that are broadcast on the network.

During the ranging process, information is transmitted to help the cable modem adjust its timing and upstream transmission levels. After the modem has been registered, the downstream traffic will contain the download data and administrative information that the cable modem requested or information that the CMTS wants to distribute. Every cable modem can listen to the signals broadcast from the CMTS, but only the modem with the correct destination address can access the information contained in the payload section of the packet.

Ordinarily, one downstream channel is paired with multiple upstream channels to achieve the requisite balance in data bandwidth. Each modem transmits bursts in timeslots (reserved, contention or ranging). The cable modem must accept a modulated RF signal (Table 4).

Basic operation of the CMTS

A head-end cable modem termination system communicates with cable modems located in subscribers' homes, to create a virtual local area network (LAN) connection.

TABLE 4, ELECTRICAL OUTPUT FROM THE CABLE MODEM

Parameter	North America	Europe
Frequency	5 to 42 MHz edge to edge	5 to 65 MHz edge to edge
Level range	+8 to 55 dBmV (16 QAM) +8 to +58 dBmV (QPSK)	+68 to 115 dBµV (16 QAM) +68 to +118 dBµV (QPSK)
Modulation type	QPSK and 16 QAM	QPSK and 16 QAM
Symbol rate (nominal)	160, 320, 640, 1,280 and 2,560 ksym/sec	160, 320, 640, 1,280 and 2,560 ksym/sec
Bandwidth	200, 400, 800, 1,600 and 3,200 kHz	200, 400, 800, 1,600 and 3,200 kHz
Output impedance	75 ohms	75 ohms
Output return loss	> 6 dB (5-42 MHz)	> 6 dB (5-65 MHz)
Connector	F-connector per [IPS-SP-406] (in common with the input)	F-connector per [IPS-SP-406] (in common with the input)

Signals from various sources, including broadcast transmissions, satellite-delivered programming, and local television broadcasts, are received and processed in the head-end. Each television signal travels on a different frequency that acts as a self-contained spectrum inside the cable.

Network-layer requirements for the CMTS extend beyond transparency to IP traffic. The CMTS must also support

- variable-length subnet masks;
- classless addressing;
- IP multicast addressing and forwarding;
- Internet group management protocol (IGMP);
- proxy ARP; and
- the filtering of DHCP downstream-bound broadcast packets to protect against BOOTP server spoofing.

The data-over-cable protocol relies heavily on the CMTS for its implementation. Each node on the head-end is capable of supporting between 1 and 2,000 cable modems. The average number of cable modems per node is expected to be around 500. The CMTS is responsible for the initialization, ranging and maintenance of the network formed by the cable modems. The initialization process, which is managed by the CMTS, can be divided into the following phases:

- synchronization—the CMTS sends timing and frequency information to the cable modem, to establish synchronization;
- ranging—the CMTS guides the cable modem through the ranging process;
- IP connectivity—the CMTS establishes IP connectivity;
- time—the CMTS establishes the time of day;
- settings—the CMTS transfers operational parameters; and
- baseline privacy initialization (BPI)—the CMTS initializes baseline privacy if the cable modem can run it.

The CMTS and the cable modem operate as forwarding agents and as hosts. Figure 13 shows the protocol stack used by these components. The data forwarded through the cable modem is *link-layer transparent bridging* and supports multiple network layers.

The main function of the CMTS is to transmit IP packets transparently between the head-end and end-user. Management functions, such as support for spectrum management and software downloads, are sent as IP packets as demonstrated in Figure 13. Both the CMTS and the cable modem operate as IP and logical link control (LLC) hosts according to the IEEE 802 standard for communication over the cable network. The CMTS must support the transport of IP traffic and must be able to restrict the network layer to a single protocol, such as IP.

Conceptually, the CMTS forwards data packets at two interfaces between the CMTS-RFI and the CMTS-NSI, and between the upstream and downstream channels (Figure 14).

The cable modem access network operates at layers 1 (physical) and 2 (media access control/logical link control) of the open systems interconnection (OSI) reference model. Thus, layer 3 (network) protocols, such as

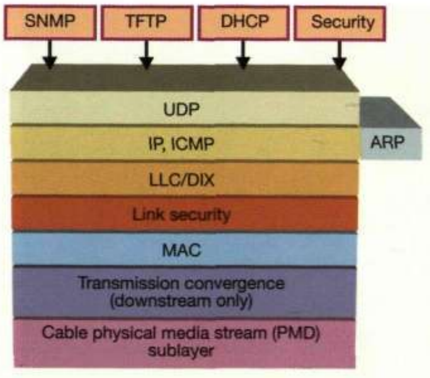


Figure 13
Protocol stack on the RF interface.

Figure 14
Data forwarding through the CMTS and cable modem.

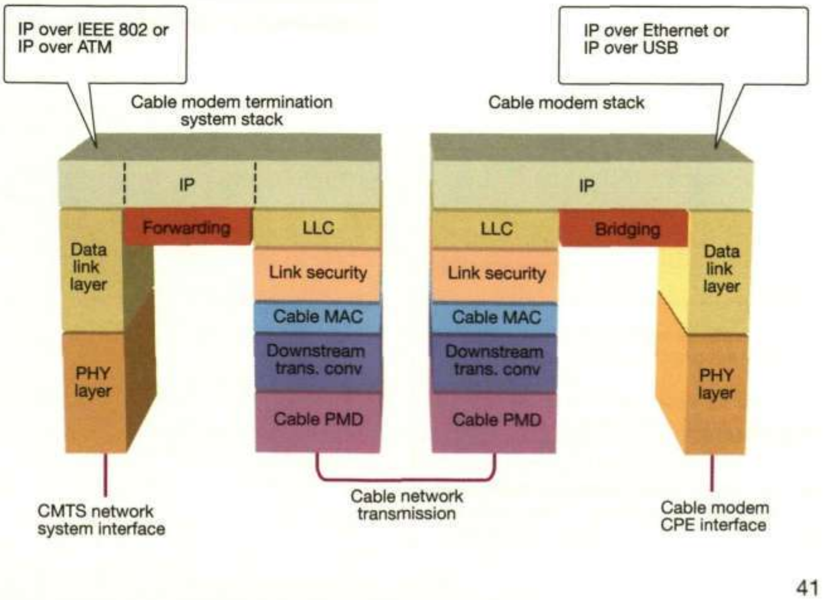


Figure 15
Ericsson currently offers two cable modem products: the PipeRider HM200c and HM201c. The HM200c is DOCSIS 1.0 certified and the HM201c is Euro-DOCSIS certified. The HM200c has received conditional @Home certification. The PipeRiders connect to customer premises equipment via USB and 10 Mbit/s Ethernet interfaces. A PipeLock security button allows the user to isolate the LAN and WAN sides of the cable modem when it is not in use, but keeps the modem attached to the system and available for instant access. In accordance with a specification in the DOCSIS standard, the modem software can be upgraded remotely by the cable operator. The HM200c/HM201c has been certified by regulatory bodies for use in the US, Canada, the EU, Japan, Korea, Australia, and China.

IP traffic, can be seamlessly delivered to end-users over the cable modem platform.

Standardization

In the US, cable operators formed a consortium called CableLabs to accelerate the development of standards within the cable industry. In Europe, tComLabs performs a similar role. Other efforts are currently underway in Europe and North America to standardize voice over IP (VoIP) over cable and home networking.

One of CableLabs' recent standards regulates the sending of data over cable. This effort resulted in the development of the data-over-cable service interface specification (DOCSIS) which was adopted as IEEE 802.14. An appendix to this standard applies to European television standards. The standard was also adopted by tComLabs, a consortium of European operators who coordinated it as Euro-DOCSIS 1.0. (ETSI standard, ES 201 488 V1.1.1). It is based on the 1.1 RFI Version IO6, which is included in tComLabs' appendix to Euro-DOCSIS.

Every cable modem has to go through DOCSIS certification by CableLabs. To date, DOCSIS has gone through seventeen certification waves. Each wave consists of a set of lab tests and documentation to ensure cable modem design compliance with established specifications and standards. Euro-DOCSIS requirements are certified by tComLabs. They are currently in their third wave of compliance testing.

VoIP

Many cable operators are eyeing voice over IP as a major source of revenue and have asked CableLabs to develop a standard, known as PacketCable, which includes voice as well as provisions for other kinds of media. The standard prompted several enhancements to the DOCSIS standard (DOCSIS 1.1), including quality of service (QoS) and additional remote-management features.

TRADEMARKS

PacketCable and DOCSIS are trademarks and/or trade dress of CableLabs or third parties.

Some of the enhancements are

- configuration and registration for cable-modem-based QoS;
- fragmentation of upstream packet data;
- payload header suppression; and
- dynamic establishment of QoS-enabled service flows by the CMTS.

The PacketCable standard stipulates how quality of service should be implemented for different kinds of multimedia flow. The security part of the standard also incorporates provisions for baseline privacy and encryption protocols, for encrypting packet data across the cable network.

The future of cable modems

In the future, cable modems will probably be bundled with other networking features and LAN technologies to facilitate the distribution of broadband IP into the home. At the low end, simple versions of cable modems will be incorporated into personal computers. This evolution into a home gateway is already starting to take place. For convenience and security, the gateway will provide basic software features, such as network address translation (NAT), and function as a DHCP server for local devices, allowing users to set up an extensive home network. Firewalls and content filtering can also be incorporated into the modem to enhance the security of the family computing environment. Operators can simplify the implementation and generate additional revenue by offering these features in the form of a managed service.

Universal plug and play is being investigated to make implementation easy. Other wired and wireless LAN interfaces are being considered, including Bluetooth, HiperLAN/2, HomePNA and powerline networks. Specialty devices, such as set-top boxes and game machines, might also come with built-in cable modems in the future. These devices might even serve as home gateways.

Regardless of the form these devices take, the most interesting and exciting part will be the additional capabilities, services, and entertainment that broadband access will bring into our homes.

Conclusion

A tremendous opportunity exists with broadband over cable. Cable industry standardization is driving exponential growth as evidenced by sales of DOCSIS-standard modems. Cable modems with wireless interfaces will form a foundation step to increased opportunities for mobile Internet services in the home. Soon Internet protocol solutions for various broadband-to-home methods (cable, DSL, fiber) will converge to similar systems. These solutions will not only include infrastructure, but also a suite of additional revenue-producing services for operators.

The full set of services available from Ericsson to cable operators will be a key differentiator among equipment providers. Much of the work for fixed and mobile IP access already developed by Ericsson can be applied to the cable end-to-end solution.

Ericsson has the right mix of capabilities to make broadband over cable a success—cable modems, infrastructure and services. This combination constitutes the bridge that links the worldwide Internet to the home.

REFERENCES

1. High Speed Cable Modems; Albert Azzam, McGraw Hill Series on Computer Communications, 1997
2. Cable Modems: The Answer for Increased Bandwidth or Just a Stop Along the Way?, Analyst Richard Wawrzyniak, Semico Research Corporation, April 1999
3. Data Over Cable Service Interface Specifications; Radio Frequency Interface Specification; SP-RFI-105-991105

The R380s—The first smartphone from the Ericsson–Symbian partnership

Steve Bridges

The Ericsson R380s is the first GSM mobile phone to use the Symbian EPOC32 operating system, which gives the phone the functions that make it a smartphone. The original concept for the R380s was to provide mobile phone and personal organizer functionality in a single phone-sized device. The device was characterized as a business tool rather than a lifestyle accessory. As development progressed, the specification was modified to include a WAP browser, unified messaging (e-mail and SMS), and secure access to restricted-access systems, such as a corporate intranet.

The author explains why EPOC was chosen, and describes the work required to develop the R380s software from EPOC32.

What is a smartphone?

The term *smartphone* has been coined to describe devices which, while being primarily a mobile phone, incorporate elements of functions found in paper-based personal organizer systems or in modern electronic personal digital assistants (PDA). Typically, a smartphone contains a calendar, an address book, e-mail and messaging functions, and a browser for the wireless application protocol (WAP), together with a range of ancillary functions. It is also expected to work simply as a mobile telephone. A smartphone is roughly the same size as a standard business mobile phone—for example, the Ericsson R320—and like a standard phone, it allows users to operate its basic functions in a one-handed fashion. Extra functionality is accessed via the large touchscreen, which differentiates it from standard phones.

What is the R380s?

The Ericsson R380s (Figure 1) is the first GSM mobile phone to use the Symbian EPOC32 operating system, which gives the phone the functions that make it a smartphone.

Why EPOC?

EPOC32 is an ideal operating system for smartphone devices. This real-time, multitasking, multithreaded operating system was created specifically for use in handheld devices, and has been optimized for the constraints of such an environment, the chief of which are

- low power consumption;
- small display;
- limited input and output methods;
- slow processor speed; and
- stand-alone operation.

The creation of Symbian—a joint venture between Ericsson, Matsushita, Motorola, Nokia and Psion—meant that a suitable in-house operating system became available as the R380s project was starting up. However, the fit between EPOC32 and the R380s was not a perfect one. This article describes the work required to develop the R380s software from EPOC32.

The R380s concept

The original concept for the R380s was to provide mobile phone and personal organizer functionality in a single phone-sized device. The device was characterized as a business tool rather than a lifestyle accessory and was specified accordingly. As development progressed and the concept of the mobile Internet took shape, the specification was modified to include features that would take advantage of this mobile business environment:

- a WAP browser would take advantage of the mobile services expected to be offered for this new wireless standard;
- a unified messaging feature would allow standardized control of all available text-transmission services, such as e-mail and short message service (SMS); and
- secure access with built-in security algorithms would allow users to dial up a secure link with restricted-access systems—for example, a corporate intranet.

To provide good functionality both as a smartphone and as a basic phone, two modes of operation were specified. This design affected the hardware of the phone and the user interface (UI).

The solution to the product requirements

The product would include features and functions not previously found in mobile devices. In hardware, the device would have to operate in two significantly different modes: normal phone mode, and smartphone mode. The device must also control two different basic functions with maximum efficiency:

- GSM functionality and the air interface; and
- PDA-type operations.

The requirement to support two different modes of operation while having the largest screen possible for the size of device resulted in the design of a hinged, passive flip. The flip holds the normal phone keypad and covers 60% of the touchscreen. In flip-closed (FC) mode, the R380s looks and operates

BOX A, TERMS AND ABBREVIATIONS

API	Application program interface
ARM	Advanced Research Machines (the manufacturers of the processor used in EPOC devices
ECK	Ericsson component kit
ETSI	European Telecommunications Standards Institute
FC	Flip closed
FO	Flip open
GSM	Global system for mobile communication
OS	Operating system
PC	Personal computer
PDA	Personal digital assistant
QWERTY	Standard typewriter keyboard: first six letters, top-row, left
ROM	Read-only memory
UI	User interface
VGA	Video graphics array
WAP	Wireless application protocol

like an ordinary GSM mobile phone. Key presses are physically transmitted through the flip via plungers attached to the keys. The resultant contact with the underlying touchscreen is interpreted by the operating software. In this mode, the exposed part of the screen does not respond to touch, and the phone looks and feels like an ordinary mobile device, with some added display functionality to help users. The flip is a purely mechanical device that contains no trouble-prone electronic parts or flexible electrical connections. In the event of damage, it can easily be replaced by an approved Ericsson service center.

When the flip is opened, a sensor signals the operating software, which alters the display. In flip-open (FO) mode, the entire touchscreen is exposed, the touch input is enabled, and the screen orientation is rotated through 90 degrees to give a landscape-style display. This mode maximizes the potential for data display and input with a look and feel that more closely resembles established PDA design than that of a mobile phone. Many PDA-type functions are solely available in the flip-open mode, since they are best served by static, two-handed user operation.

Given the requirements for data-input methods, and that in software the device had no proven user interface for the size and form-factor of its screen, the user interface had to be designed from the ground up. A two-processor design was implemented to provide efficient execution of the two main groups of functions in the smartphone, maximizing the possible reuse of software and hardware from earlier successful development work.

The GSM (phone-side) functionality is controlled by a processor similar to that used in standard Ericsson GSM phones. This processor, which runs the OSE real-time operating system from Enea Data, handles all operations related to the air interface and controls audio operation.

The organizer (PDA-side) functionality is controlled by an ARM-based processor design, which manages PDA-side functions and controls the phone-side processor at a high level, initiating and terminating calls on user command. The ARM processor runs the EPOC32 operating system, which has been optimized for this processor family.

The need to close-couple the two different sets of functionality without causing loss of performance on either side led to the design of a specialized communications channel. The two processors communicate via a high-speed software backplane, which provides an efficient interchange of data and commands. Because standard GSM voice calls are handled entirely by the phone-side processor, the PDA-side processor is free to manage other activities during a call.

Figure 1
Image of the R380s.

EPOC32—the starting point

Before Symbian was formed, EPOC32 was associated with the popular Psion 5 organizer (Figure 2), a PDA with a half-VGA-sized touchscreen, a full QWERTY key-

Figure 2
Image of the Psion 5 personal digital assistant.

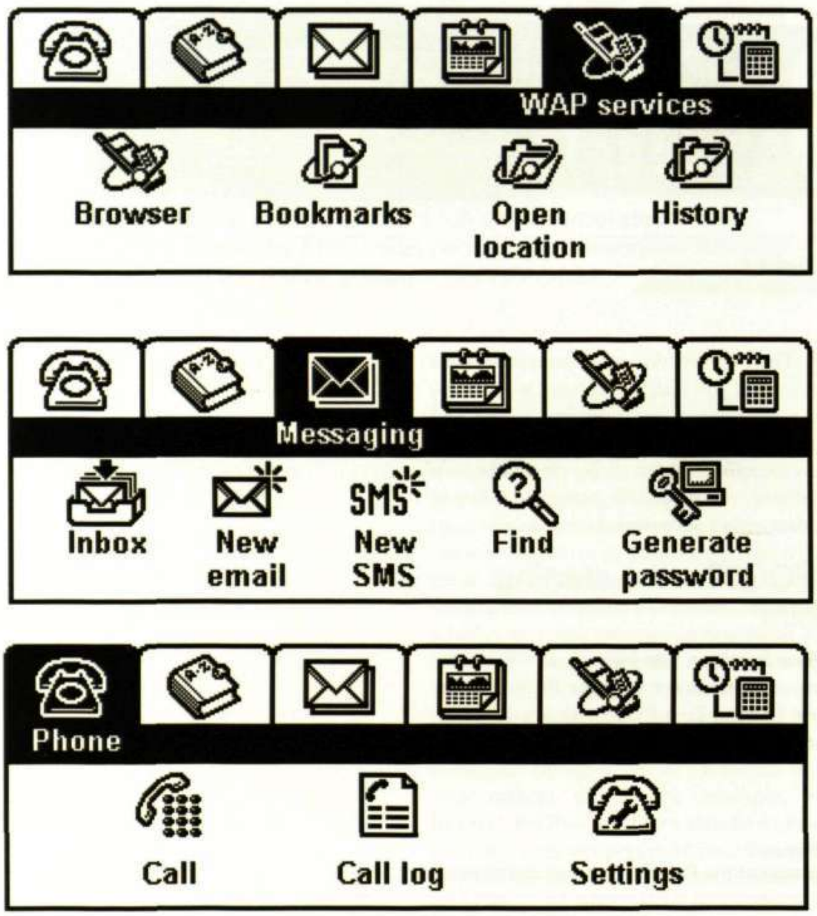


Figure 3
Screen shots of the R380s user interface.

board, and the ability to add extra software packages via PC download or plug-in cards.

EPOC32 offered several benefits to R380s developers, which simplified the task of producing the PDA-side software:

- EPOC is a real-time operating system that permits the support of time-dependent operations, such as telephony;
- the user interface is effectively decoupled from the underlying operating system and application engines—it can thus be modified to accommodate the look and feel of a new device; and
- the software is modular, which simplifies parallel development at multiple sites.

New features needed in EPOC32

Depite its obvious suitability for use in the R380s project, the version of EPOC32 used in the Psion 5 had certain drawbacks for Smartphone use:

- Psion devices adhere to an *always-on* concept, with both a main and a backup battery. The expectation is that the device is always powered and that data in RAM is persistent. However, if both batteries are allowed to run down simultaneously, all user data is lost. Mobile phones adhere to an *often-off* concept. Consequently, to be persistent, user data must be backed up to permanent storage;
- Psion PDAs have no concept of telephony. No engine or user interface was available in EPOC32 to support GSM telephony;
- EPOC32 had been designed to support half-VGA devices with real keyboards and multiple on-screen windows. The user interface and engine software had been optimized for this sort of design;
- the concept of dual-mode operation (FO, FC) was completely alien to EPOC32; and
- software internationalization had been considered a ROM-swap function, but a mobile phone must be able to change languages on the fly.

To overcome these obstacles, and numerous other minor issues, the software was modified as follows:

- A flash file system was introduced, to accommodate persistent data. The challenge was to produce a system that made regular updates—to guarantee an acceptable level of security—without noticeably affecting performance in the rest of the device.
- A telephony support module, known as ETEL, was produced to support telephony functions (from the processor on the PDA side). This module serves as an interface between the phone-side software and PDA-side applications and services that require telephony access.
- The entire user interface was deconstructed and redesigned to function appropriately for the R380 (Figure 3). Instead of being file-based, the concept for the new user interface was task-based.
- Application engines were modified where necessary, and an entire unified messaging application was constructed.
- The operating system was modified to support dual-mode (FO and FC) operation, with
  - seamless switching between defined equivalent views in both modes; and
  - persistence of viewable data.
- The operating system was modified to allow users to control language/locale without the need for external hardware or

without having to modify the hardware. Not all available language/locale packages can be contained in the memory of a single R380s, so provision has been made for downloading the language population not shipped with a particular phone.
- The Symbian software component kit, known as EIKON, was unsuitable for the new user interface. A new component kit—the Ericsson Component Kit (ECK)—was developed to give the correct functionality of components, such as on-screen buttons, icons, dialog boxes, and so on.

Each of these modifications was required to make a truly effective smartphone. The omission of any of them would have resulted in a significantly less attractive, less useful device. However, the primary motivation for this work was usability.

The EPOC32 architecture on the R380s

A detailed description of the R380s software architecture is beyond the scope of this article, but a glance at Figure 4 gives some idea of its complexity.

Usability and the user interface

Since no proven designs existed for this type of device, the usability team had to create a new user interface from the requirements. The team took this opportunity to incorporate continuous usability testing into the design process, to ensure that users would end up with a worthwhile and desirable product. As its guideline, the team took the European Telecommunications Standards Institute (ETSI) definition of usability—in particular, the requirement which stipulates that at least 75% of novice users should be able to perform a given task correctly on their first attempt. All major use cases on the device were tested in this manner. Reiterative testing applied to every aspect of the design, from the on-screen appearance of details, such as buttons, to the overall navigation concept of the operating system.

In this respect, the R380s software can be said to have been designed by users rather than engineers—the activities of the usability team, which acted as users by proxy, ensured that the user interface is the best fit for the size, form factor, and functionality of the R380s. The software started life as Psion 5 software, but as Figure 2 shows, the changes to the user interface were significant.

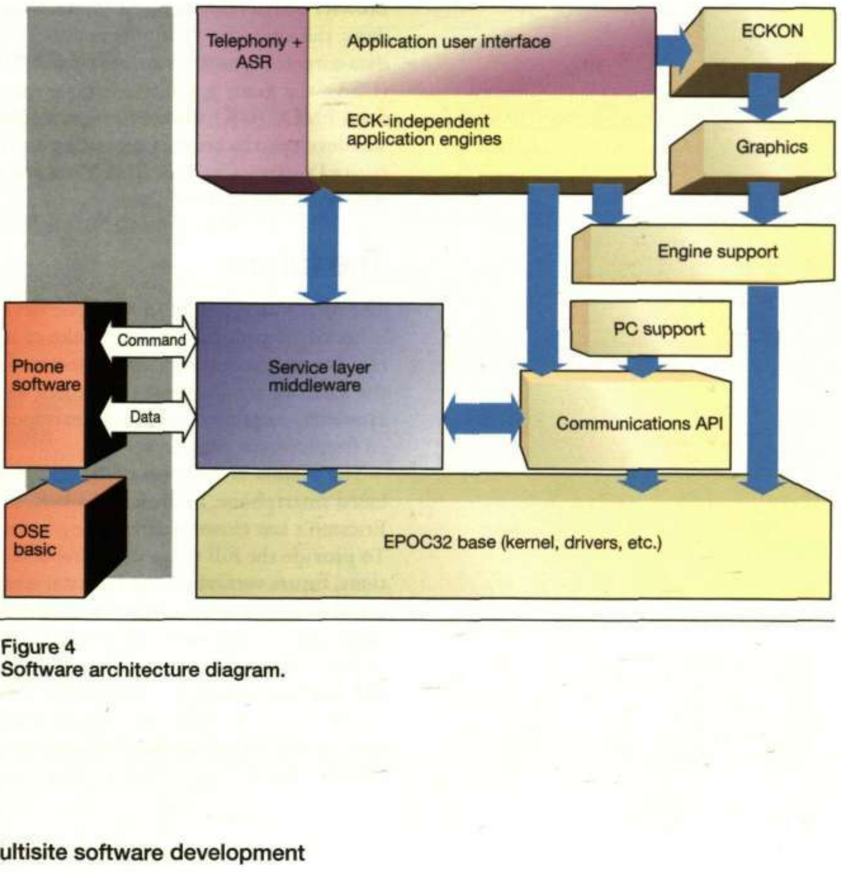


Figure 4
Software architecture diagram.

Multisite software development

The commercial requirement for rapid development dictated that the software would have to be developed at multiple sites. Obviously, this meant that the development team needed a robust system of configuration management and quality control, as well as a strong suite of common development and test tools. Three main sites were involved in software development:
- Ericsson in Stockholm produced the functional specifications and lower-layer PDA software, modified the phone software and verified the system;
- Ericsson Software Applications Laboratory in Warrington (England) produced the user interface specifications (scenario documents), the ECK, the telephony application, and the user interface code for most of the organizer applications. This team also performed the functional test on much of the software; and
- Symbian in London carried out the base port to the ARM processor, supplied software development kits, and modified some of the organizer applications.

In addition, some software development was subcontracted. For instance, the WAP browser was developed by AU System (Sweden); the character-recognition part of the data entry component was developed by CIC (USA); the game application was acquired from EMCC (UK); authentication software was developed by Secure Computing and Security Dynamics; and the flash filing system was developed by Enea Data.

The future

The future direction of smartphone devices is heavily dependent on the uptake of mobile Internet services by consumers and on the kinds of service these consumers want. However, certain trends can be extrapolated from current data.

The R380s is Ericsson's first EPOC32-based smartphone, but it is also likely to be Ericsson's last closed smartphone platform. To provide the full range of desirable functions, future versions of the Ericsson smartphone platform will be open to third-party developers, in the same way as the Psion 5 or Palm Pilot devices are. The ability to add applications to a smartphone, either via the PC or directly over the air will be vital to the success of these devices in the next decade.

To support development, Ericsson has launched the Ericsson Developers' Zone (www.ericsson.com/developerszone)—a one-stop shop for developers looking to get information, support, training and tools related to Ericsson platforms. Test and certification services for externally developed software are also available via this site.

The modular and scalable nature of EPOC32 will make it the operating system of choice for a wide range of devices, thus ensuring a common software architecture. Future smartphones will support multimedia functions as well as traditional mobile data services.

Conclusion

The R380s, which is the first smartphone on the market to incorporate Symbian's EPOC32 operating system, opens up the future of devices by matching hardware and software to give a full range of desirable functions in a user-friendly, efficient, and portable package. The R380s points the way to a community of devices whose common architecture and open platform allow user requirements to be met by specialized developers.

TRADEMARKS

Previous issues

No. 4, 2000

Experiences of the live WCDMA network in Stockholm, Sweden
Media Gateway for mobile networks
Ericsson's IP-based BSS and radio network server
Control servers in the core network
Ronja—A Java application platform
Wireless *ad hoc* networking—The art of networking without a network
Ericsson's WebOnAir Information Server

No. 3, 2000

ENGINE network server
ENGINE Access Ramp—The next-generation access architecture
Speech recognition technology for mobile phones
GPRS support nodes
Family of RBS 3000 products for WCDMA systems
Ericsson's network-based IP-VPN solutions
Fixed wireless access for broadband services

No. 2, 2000

Evolving from cdmaOne to third-generation systems
Third-generation TDMA
Business solutions for mobile e-commerce
The RBS 2206—A flexible ticket to third-generation wireless systems
Communications security in an all-IP world
HIPERLAN type 2 for broadband wireless communication

No. 1, 2000

IPv6—The new generation Internet
WAP—The catalyst of the mobile Internet
The challenges of voice over IP over wireless
Openness in AXE
Management solutions for IP networks

No. 4, 1999

Network evolution the Ericsson way
The AXI 540 router and the public IP network edge
Real-time routers for wireless networks
Ericsson's Bluetooth modules
Mobile Internet—An industry-wide paradigm shift?
Ericsson's mobile location solution

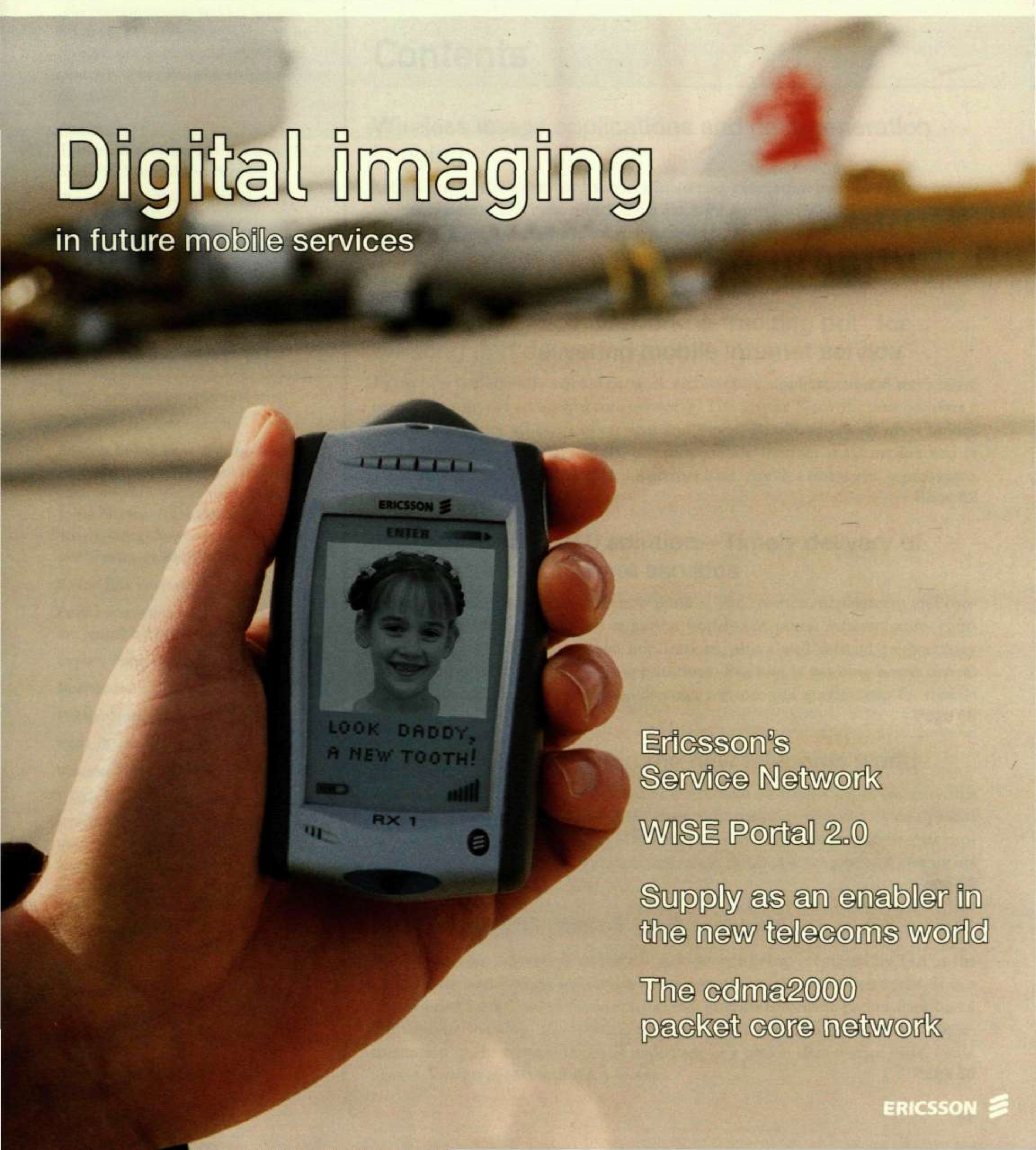
Telefonaktiebolaget L M Ericsson
SE-126 25 Stockholm, Sweden
Phone: +46 8 7190000
Fax: +46 8 6812710

ISSN 0014-0171
Ljungföretagen, Örebro 2001

Ericsson REVIEW

2/2001

THE TELECOMMUNICATIONS TECHNOLOGY JOURNAL



Digital imaging

in future mobile services

Ericsson's Service Network

WISE Portal 2.0

Supply as an enabler in the new telecoms world

The cdma2000 packet core network

ERICSSON

Ericsson
REVIEW

THE TELECOMMUNICATIONS TECHNOLOGY JOURNAL

The purpose of Ericsson Review is to report on the research, development and production achievements made in telecommunications technology at Ericsson. The journal is published in English and Spanish and distributed quarterly to readers in more than 130 countries.

Address:
Telefonaktiebolaget LM Ericsson
SE-126 25 Stockholm, Sweden

Phone: +46 8 719 00 00

Fax: +46 8 681 27 10

Internet:
http://www.ericsson.com/review

New subscriptions:
http://www.ericsson.com/review

Address changes and distribution:
Fax: +46 8 681 27 10
birgitta.blom@lme.ericsson.se

Publisher: Jan Uddenfeldt

Editor: Eric Peterson

Production manager: Eva Karlstein
eva.karlstein@lme.ericsson.se

Layout: Paues Media, Stockholm

Illustrations: Claes-Göran Andersson, Hara

Printer: Ljungföretagen, Örebro

ISSN: 0014-0171

Volume: 78, 2001

Cover: With multimedia messaging service (MMS), users can compose, exchange and display/play back images, audio and text in the same message. Ease of use will allow users to create MMS messages nearly the same way they as they create SMS today. The message content will solely be limited by content developers' imaginations. In this example, for instance, a young girl is excited to share the news of her first permanent tooth with her father.

Contents

Wireless image applications and next-generation imaging

SMS is one of the most-used services in second-generation mobile telecommunications systems. And now, multimedia messaging service (MMS) is taking the mobile messaging evolution a giant step forward by giving users the power to become content developers and content consumers. It will be the first mobile application that can handle several different kinds of media in the same message. **Page 54**

Ericsson's Service Network: a "melting pot" for creating and delivering mobile Internet service

In the new horizontally layered network architecture, applications and services are separated from the access and core networks. The Service Network thus becomes a "melting pot" for all types of services and service combinations. It gives service providers a way of managing the complex mobile Internet mass market and of maintaining "ownership" of their subscriber base. And for end-users, it provides a personalized service environment. **Page 62**

The WISE Portal 2.0 solution—Timely delivery of tailored mobile Internet services

Mobile Internet growth requires new portals. The creation, deployment and continued enhancement of portals require a world-class portal infrastructure, combined with attractive and creative applications, plus a well-defined methodology for delivering the portal and the applications. Ericsson is working across several industry sectors to develop mobile Internet services and applications for mobile operators. **Page 68**

Supply as an enabler in the new telecoms world

The supply processes that have brought the mobile communications industry this far will not be adequate to cope with the extreme market demand for equipment for second- and third-generation infrastructure. Ericsson is thus modeling its supply concepts on large-volume businesses, such as the automotive and consumer-electronic-goods sectors. **Page 80**

The cdma2000 packet core network

The packet core network is a network architecture being promoted by TIA as the packet-data standard for upcoming third-generation cdma2000 networks. It is a collection of logical and physical entities that provide IP-centric packet-data-based registration, roaming, and forwarding services for mobile nodes. Ericsson's implementation in the cdma2000 PCN makes use of a packet data service node, home agents, foreign agents and AAA servers. **Page 88**

Contributors

In this issue

Joel Askelöf

Lars Boman

Charilaos Christopoulos

Andreas Grundsell

Karin Mallmin

Tim Murphy

Thomas Papanikolaou

Joel Askelöf, who joined Ericsson in 1998, is employed at the AudioVisual unit, Ericsson Research, where he works on still-image compression, universal multimedia access and augmented reality. He has been involved in the standardization of JPEG2000 since 1998, both as co-editor of JPEG2000 part 5 (reference software) and currently as the head of the Swedish delegation. He holds an M.S. in engineering physics from Uppsala University.
joel.askelof@era.ericsson.se

Lars Boman is currently head of Ericsson Internet Applications. From 1983 to 1985, he was research and development manager at Ericsson Business Network AB. From 1985 to 1990, he served as director of research and development. In June 1990, he was appointed vice president, general manager and business unit manager of Ericsson Business Network AB, which was responsible for packet-data communication equipment and large PBXs. In September 1995, Lars became president of Ericsson Infotech AB, and from May 1997 to February 2000, he served as vice president, products and technology for Ericsson Mobile Systems in Japan.
lars.boman@eip.ericsson.se

Charilaos (Charis) Christopoulos manages the MediaLab at Ericsson Research. He joined Ericsson in October 1995, and has worked in numerous image- and video-coding-related projects, including MPEG-4 and H.263 video transcoder architectures, and multipoint control units. In 1997, he began working with JPEG2000, in part as head of the Swedish delegation in ISO/SC29 WG1 JPEG (until March 2000); as editor of the verification model; and as co-editor of part I of the standard. Since 1999, he has been working with MPEG-7 (he is co-chairman of the mobile profile *ad hoc* group) and virtual/augmented reality. He is a senior member of IEEE and associate editor of IEEE Transactions of circuits and systems, part II: Analog and digital signal processing. He holds a B.S. in physics from the University of Patras in Greece, an M.S. in software engineering from the University of Liverpool in the UK, and a Ph.D. in image and video coding from the University of Patras and the Free University of Brussels in Belgium.
charilaos.christopoulos@era.ericsson.se

Andreas Grundsell currently works for Supply and Sourcing at Supply and IT within Ericsson's Mobile Systems division (DMS). Before this, he served as the head of UMTS supply development and as the supply strategy advisor at DMS Supply and IT. He has many years' experience of working as a project manager in production and in supply support system development. He holds an MBA and an M.S. in mechanical engineering.
andreas.grundsell@era.ericsson.se

Andy Johnston is a master software designer at Ericsson Eurolab Deutschland in Aachen Germany. He is currently a system manager for the Ericsson WISE Portal at Ericsson Radio Systems in Kista as well as a design manager for Mobile Portal Technologies at Ericsson

Andy Johnston

Mathias Larsson Carlander

Mike Slssingar

Fredrik Öijer

Eurolab Deutschland. Andy has been with Ericsson since 1995. Prior to joining Ericsson, he was the chief developer of the widely used CAMPUS® Database & CAE system. He has a B.S. in computer science from Dublin City University, Ireland.
andrew.johnston@eed.ericsson.se

Mathias Larsson Carlander joined Ericsson and the Compression Lab in April 1998. He is now a senior research engineer in the Visual Technology unit at Ericsson Research. He mainly works with still-image coding projects, but also works with projects related to objective and subjective image-quality measures and video coding. Since 1998, he has been active in ISO/IEC JTC1/SC29/WG1 and JPEG2000 standardization; he was coeditor of JPEG2000, part II, and chairman of the region of interest subgroup. In November 2000, he became chairman of AG29, the standardization group that is responsible for multimedia, coding, and representation—including JPEG and MPEG standards—at Informationstekniska Standardisering (ITS) in Sweden. He holds an M.S. in computer science and engineering from the Linköping University.
mathias.carlander@era.ericsson.se

Karin Mallmin, who joined Ericsson in 1994, is currently manager of supply chain planning at the GSM, TDMA, EDGE Systems business unit, Ericsson Radio Systems. Before this, she worked with GSM supply chain management (focus on the Chinese market) and later as the program coodinator of global supply chain management, Vodafone global accounts. Karin studied industrial engineering and management at Linköping University, Sweden.
karin.mallmin@era.ericsson.se

Tim Murphy is an integration specialist at Ericsson Research Canada for the cdma2000 Packet Core Network. Since joining Ericsson, in 1999, he has been involved in the development of third-generation packet-data systems, including GPRS and the PCN. Most recently, his work has focused on Linux application research and PCN complementary technology research. He holds a B.S. in finance from Georgetown University, Washington, D.C., and is completing an M.S. in telecommunications from the University of Colorado at Boulder.
timothy.murphy@emc.ericsson.se

Thomas Papanikolaou joined Ericsson in 1999, working as technical architect and chief developer, and later as system manager of the WebOnAir™ - Information Server. He is currently a system manager of the Ericsson WISE Portal Presentation Layer. Prior to joining Ericsson, he was a researcher in computational number theory and cryptography at the University of Technology in Darmstadt, Germany. In 1998, he received a postdoctoral fellowship from the Centre National de la Recherche Scientifique (C.N.R.S.) in France. Thomas has a Ph.D. in computer science from the University of Saarbrücken, Germany.
thomas.papanikolaou@eed.ericsson.se

Mike Slssingar joined Ericsson in 1997. Since early 1999, he has managed the development of demonstration WAP applications for CeBIT and Telecom 99, and worked with WAP and wireless ISP consultancy for Ericsson Business Consulting. This work led to his role as systems manager for the WISE 2.0 Portal. Mike was responsible for the presentation, session, subscription, and J2EE application server definition studies and led the initial technology partnering before the requirements for professional services deployment were defined. He holds a postgraduate degree in software engineering and an M.S. in computing for commerce and industry.
mike.slssingar@etl.ericsson.se

Fredrik Öijer is a strategic product manager in the GSM/UMTS business unit at Ericsson Mobile Communications. He has been with Ericsson since 1997, and is currently responsible for messaging technologies and applications, and for the market-related design issues of general user interfaces. He holds a B.S. in informatics (emphasis on human-computer interaction, minor in business administration) from the Lund University in Sweden.
fredrik.oijer@ecs.ericsson.se

Wireless image applications and next-generation imaging

Joel Askelöf, Charilaos Christopoulos, Mathias Larsson Carlander and Fredrik Öijer

Ericsson constantly conducts worldwide consumer research to find out what kinds of application consumers want and how they want to use them. Besides making phone calls, imaging and messaging are on the tops of consumers' wish lists for future mobile applications.

Until now, imaging has mostly consisted of images and graphics that are displayed on Web pages. But because imaging will become a natural part of mobile services, image flow and digital-imaging requirements will change.

This article briefly discusses third-generation-imaging services, including multimedia messaging services, which is perhaps one of the most important services. It also discusses some requirements being put on new imaging applications and formats, and describes JPEG2000—the newly finalized ISO/IEC standard.

TRADEMARKS

GIF and Graphics Interchange Format are trademarks or registered trademarks of CompuServe, Inc. in the US and other countries.

Java is a trademark or registered trademark of Sun Microsystems, Inc. in the US and other countries.

Imaging in third-generation messaging systems

Imaging is not an application, but it will be an integrated part of Ericsson's product offerings and solutions. Messaging is one of the most important applications that use images. For example—using a terminal, such as the one in Figure 1—envision taking photos of a hotel room when you are on a business trip, and sending them to your family, along with messages that say "miss you." Or how about taking photos of family members as they lie on a sunny beach, attaching a message, such as "don't want to come home," and sending the text and photos to your neighbors?

From consumers' perspectives, these applications will symbolize the third-generation regardless of underlying technology. The roadmap toward imaging in the third-generation starts with MMS in generation 2.5.

When we talk about imaging as a very important part of future mobile applications, we have to acknowledge that mobile phone users must somehow capture images. A natural way of doing this is to integrate a digital still-image camera into mobile phones, which would enable them to capture and send images and messages from the same phone. But not all mobile phones will have integrated cameras. To accommodate many different consumers and provide them with the most flexible solution, mobile phones must be able to transfer images from external cameras. In most cases, image transfer will occur via Bluetooth, which has the still-imaging profile—a dedicated profile that is standardized for this purpose.

Ericsson is actively participating in standards initiatives and thus ensuring that its phones will be compatible with external cameras. Ericsson also works with Canon Inc., to ensure that complete imaging systems will be available for mobile consumers. The Ericsson-Canon cooperation basically deals with

- Internet-based imaging services;
- image transfer from the camera to the mobile phone; and
- the development and promotion of world-

Figure 1
This future, pocket-sized, mobile multimedia concept device features voice, imaging, and video. A wireless, hands-free earphone connects via a Bluetooth connection.

wide standards for imaging and image communication.

Multimedia messaging service—the next generation of SMS

Short message service (SMS) is one of the most-used services in second-generation mobile telecommunications systems—more than 15 billion messages are sent monthly. Multimedia messaging service (MMS) is currently being defined and specified for generation 2.5 and third-generation implementation. MMS takes the mobile messaging evolution (which began with SMS) a giant step forward by giving users the power to become content developers and content consumers. It will be the first mobile application that can handle several different kinds of media (text, images, animation, video, audio, or any combination thereof) in the same message. The message content will solely be limited by content developers' imaginations.

Regular mobile phone users will create MMS messages nearly the same way as they currently create SMS messages. The big difference will be opportunities to include different kinds of media—besides text. As mentioned above, mobile phones will be able to import digital images from digital still-image cameras. Some phones might even have built-in cameras, or cameras as accessories, that facilitate image capture. If video cameras are used, then video clips can be recorded and inserted into messages. Users then add text and send messages using recipients' phone numbers as addresses. Simple and straightforward! This type of messaging will probably be the most common use for MMS.

Another interesting use occurs when content providers send teasers to consumers, for example, excerpts from new music CDs. The messages can contain images of a new CD and a short video clip or sound sample that enables users to listen to parts of the CD and then buy the complete album via their mobile phones.

MMS will be launched on the market at the end of 2001 and will spread widely in 2002. This service, which has been standardized within the Third-generation Partnership Project (3GPP), has strong, broad industry support. Among the companies participating in the standardization, we find the biggest operators and terminal manufacturers in the world. We all recognize the importance of MMS, and we are jointly developing and enhancing MMS specifications.

Figure 2
A typical MMS message might consist of text, an image, and a 20-second music excerpt.

Imaging using Bluetooth

Bluetooth connectivity will also be an important part of the mobile future. The ability to transmit images, video, and audio wirelessly between cameras, MP3 players, phones and smartphones will allow users to extend their communications with friends and family members. Within the Bluetooth special interest group (SIG), a working party was established to create the Bluetooth still-imaging profile.

Bluetooth connectivity will enable friends to exchange images and graphics, which can then be sent to other friends—using MMS. Mobile phones can also be used for other tasks, such as browsing images stored in digital cameras. For example, after a day of sightseeing, users can browse their digital cameras, select the images they like, and send them to their private, online photo albums. They can also use their mobile phones as remote controls for camera shutters. At the press of a phone key, an image is captured and sent to the phone's display. When the user arrives at home, the image can automatically be transferred from the camera to a PC—as soon as the camera is within range of the PC.

BOX A, TERMS AND ABBREVIATIONS

3GPP	Third-generation Partnership Project
Bluetooth SIG	Bluetooth special interest group
DCT	Discrete cosine transform
GIF	Graphics interchange format
GSM	Global system for mobile communication
IEC	International Electrotechnical Commission
ISO	International Standardization Organization
ITU-T	International Telecommunication Union – Telecommunications Standardization Sector
JPEG	Joint Photographic Experts Group
JTC1	Joint Technical Committee 1
MMS	Multimedia messaging service
MP3	MPEG, 1 and 2, layer 3
PNG	"Ping", portable network graphics
SC29	Subcommittee 29
SMS	Short message service
UMTS	Universal mobile telecommunications system
WAP	Wireless application protocol
WG1	Working Group 1 of ISO/IEC JTC1/SC29

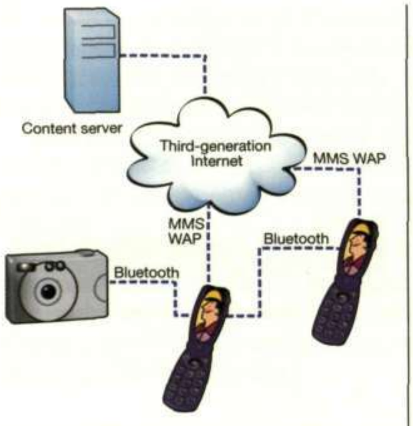


Figure 3
The new image flow.

New requirements

As image flow (Figure 3) changes, and the scope of mobile communication begins to embrace multimedia functions, new requirements will appear. Images will be exchanged between digital still-image cameras, phones and printers. As discussed above, images will also be uploaded to private photo servers or to online photo albums on the Internet, using Bluetooth, MMS, or WAP. Images and graphics will also be sent to and exchanged between family members and friends.

As new multiservice networks evolve, many different clients (phones and smartphones, high-power home PCs, game consoles and TVs) will access content on the Internet and mobile Internet (Figure 4). The convergence of services and networks has changed requirements for forthcoming imaging formats.

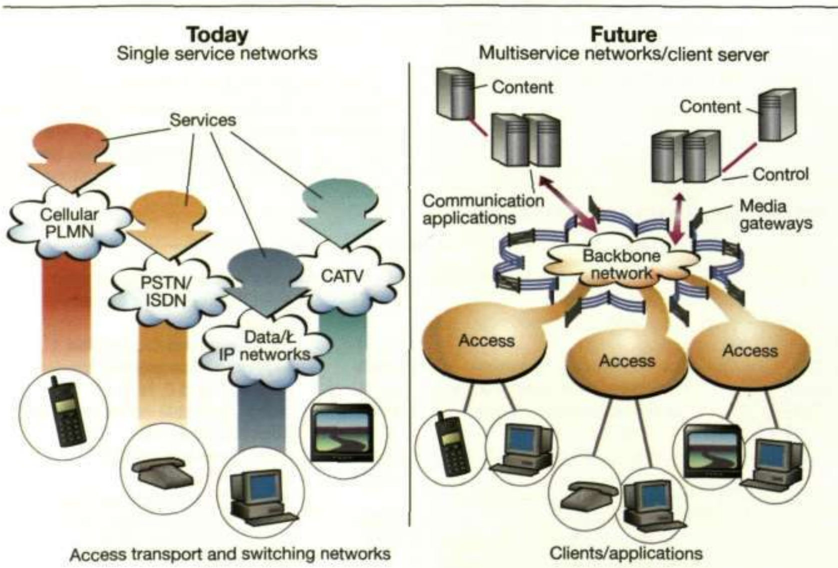
To enable image communication, the imaging formats must be widely used—that is, most applications and users must accept them. New requirements will be put on multimedia services and applications, because many different clients with different capabilities will access the same content. This creates a need to adapt content to the client. Traditionally, the adaptation of images to clients' capabilities has not been an important issue, because most clients that access content on the Internet have similar capabilities. The trend among new imaging formats is that they are designed to ease transcoding and adaptation to client capabilities.

To illustrate the need for adaptation, let us consider an online photo album service. Imagine that you take a photo of your favorite car with your digital camera and then upload the photo to your online photo album. At the website for the online service, the image can be accessed in the following formats:

- low resolution (thumbnail images, typically 160-by-120 pixels)—to be viewed, say, on your mobile device;
- medium resolution—for display on a PC or TV screen; and
- high resolution—to be used for high-quality printouts. Many of today's photo album services allow you to order high-quality printouts from the nearest print shop (so far, mainly in the US).

This scenario requires that content providers can deliver three different versions of the same image. A simple solution is to store three separate images with different resolution, but this increases the need for storage space. Another solution is to store one, high-quality version and then adapt the image to the client's requirements using media gateways in the network. However, operations at the gateways could be computationally complex if the imaging format is not properly designed for this type of application.

Figure 4
Evolution of network architecture.



Imaging formats

The adoption of digital imaging on the Internet depends on factors such as bandwidth, processing capacity and the number of users connected to the Internet. In the context of the mobile Internet, imaging is still in its infancy. As the number of services and users increases, the adoption rates and demand for imaging applications will grow—compared to the evolution of imaging applications on the fixed Internet. Several different imaging formats currently exist:

- GIF87a and GIF89a are among the most widely used formats for graphics compression and computer-generated images.
- JPEG (ITU-T Rec. T.81 | ISO/IEC 10918-1) is used for the compression of photographic images.

Some new formats (such as PNG, mainly for graphics) are becoming increasingly popular. Traditionally, imaging formats were designed for specific types of image, such as graphic or photographic images. So far, no format has been able to cover different types of image using the same compression engine. However, a new standard, JPEG2000 (ITU-T Rec. T.800|ISO/IEC 15444-1) leads the way in the new, imaging-formats area. The standard

- enables a unified system solution for as many imaging applications and image types as possible;
- introduces a way of compressing graphic and photographic images with the same compression engine;
- provides functions, such as random access, for the image code-stream and thus enables client-server applications to take advantage of the fact that users have limited viewing areas but want access to the full resolution of the stored image via commands such as image zoom and pan. This way, information that is solely required on the client side can be sent directly from the server; and
- provides different progression modes, superior compression efficiency, and region-of-interest capabilities. It will also offer easy transcoding with low computational complexity, which makes adaptation to client capabilities simple compared to other existing formats.

JPEG2000

Since 1997, Ericsson has been actively involved in JPEG2000 standardization (Table 1, see also Box B). As part of this initiative, Ericsson Research personnel held several positions, such as co-editors of JPEG2000—parts I, II, and V; editor of the JPEG2000 verification model, and chairman of the region-of-interest subgroup.

Ericsson Research also developed one of two official reference software applications for JPEG2000 together with Canon Research Centre France S.A and École Polytechnique Fédérale de Lausanne.[1] Ericsson will also host the JPEG meeting in July 2001, in Stockholm.

Requirements

When work on JPEG2000 first began, the goals for the new standard were high. Working Group 1 (WG1) wanted to design and create a flexible image-coding system that would last for many years. One important issue was that JPEG2000 was intended as a complement to the JPEG standard, not as a replacement. The aim of the new standard

TABLE 1, THE SEVEN PARTS OF JPEG2000

Part	Description
I	JPEG2000 image-coding system. The core coding system was finalized in December 2000. JPEG2000 part I was designed to contain a limited set of technologies, covering as wide a range as possible for the targeted applications of the standard.
II	JPEG2000 image coding system. Extensions to the coding system will be finalized in July 2001. Part II contains additional tools that are tailored to specific imaging applications.
III	Motion-JPEG2000 (to be finalized in November 2001). This part will be a standard for creating compressed video sequences, each video frame of which is compressed as a still image.
IV	Conformance testing (to be finalized in November 2001). This part describes what is required by a compliant implementation of a JPEG2000 decoder.
V	References software (to be finalized in July 2001). There are two reference implementations of JPEG2000 part I: the first implementation is written in C; the other is written in Java.
VI	Compound image file format (to be finalized in March 2002). In compound imagery, different parts of the image can be encoded separately using different encoding algorithms.
VII	Technical report (to be finalized in October 2001). Guideline of minimum functionality support of part I.

BOX B, BRIEF HISTORY OF THE JPEG2000 STANDARDIZATION INITIATIVE

Mid-1980s	The lack of an international standard for compression of continuous-tone still images becomes an issue.
1986	A working group is formed; the current name of this group (2001) is ISO/IEC JTC1/SC29/WG1, or the Joint Photographic Experts Group (JPEG).
1992	Work with the continuous-tone still-image standard results in the publication of ITU-T Rec. T.81 | ISO/IEC 10918-1—the JPEG standard.
1996	Extensions to JPEG ITU-T Rec. T.84 | ISO/IEC 10918-3 are published. During this work, it becomes evident that a new standard is needed that offers greater flexibility and functionality than JPEG.
1996	The original proposal for the new standard is submitted to ISO.
1997	WG1 issues a call for proposals for JPEG2000; Ericsson submits two proposals (the second is refined and subsequently included in parts I and II of the JPEG2000 standard): 1. An embedded DCT-based still-image-coding algorithm, which is based on the discrete cosine transform (DCT) algorithm, is used in the JPEG standard (called embedded DCT). 2. Efficient methods of encoding regions of interest in the upcoming JPEG2000 still-image standard, for a wavelet-based encoding of regions of interests.
1997	During an evaluation of the proposed algorithms, WG1 selects a scheme that is based on the Wavelet Transform—as a starting point for the development of JPEG2000. Ericsson's embedded DCT proposal is judged to be the best DCT-based proposal.

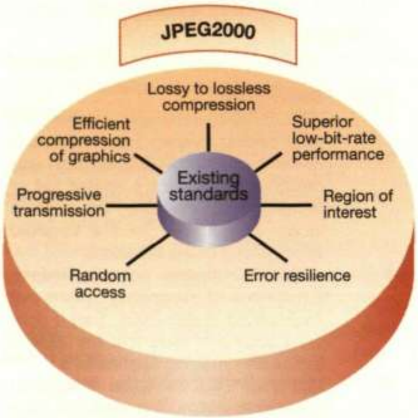


Figure 5
Requirements for JPEG2000 compared to existing standards.

was thus to provide (preferably within one unified system) an image-coding scheme for different image types (for example, bi-level, gray scale, color) with different characteristics, such as natural, medical, remote sensing, and rendered graphics.

The goals for JPEG2000 were to create a standard that was comparable to existing standards and that included all functions in one coding system (Figure 5). Moreover, the standard would

- allow for different imaging models, so the standard could work, for example, in client-server applications, real-time transmission, image archiving, and in situations with limited buffer and bandwidth resources;
- offer numerous functions—to suit it to a wide range of applications and markets;
- provide capabilities to markets that currently do not use compression;
- enable superior, low-bit-rate performance, because no existing imaging format provided good visual quality at high compression ratios;
- create a system that allows lossy and lossless compression. JPEG provides lossy and lossless compression, but uses different technology. It is impossible to decompress a lossy image with JPEG's lossless mode. Image formats (such as GIF and PNG) support only lossless compression;
- enable good compression of photographic and computer-generated images and other types of image, such as medical and remote sensing;
- create a flexible, image-coding system for the wireless environment that is robust against noisy channels; and
- enable random access to the image file and different types of progressive transmission (important requirements and tools)—this enables an image format to be used in as many applications as possible.

Important markets and imaging applications identified during the requirements work were the Internet, mobile communication, digital photography, e-commerce, digital library, printing, scanning (consumer and prepress), medical, and facsimile.

Features and functions

After drafting the requirements, the standardization group has worked to include as many features and functions as possible,

Figure 6
Comparison of JPEG and JPEG2000, at a compression ratio of 1:96.

without sacrificing too much in compression performance and without increasing the complexity of the algorithm.

The resultant compression algorithm offers very competitive compression performance for a wide range of image types, while still offering a large set of features and functions. The standard only addresses decoders, which means that all JPEG2000 decoders will be able to use those features and functions, whereas encoders can choose to implement only those features and functions that are necessary for a particular application. The main JPEG2000 features and functions of the JPEG2000 standard are

- compression efficiency;
- lossy to lossless;
- progressive transmission;
- error resilience;
- region-of-interest coding; and
- random access.

Compression efficiency

Arguably the most important feature of a still-image compression system is its compression efficiency. The more an image can be compressed while retaining acceptable image quality, the faster it can be transmitted and the easier it can be stored. One of the goals of JPEG 2000 was that it should perform better than JPEG at high compression rates. The comparison in Figure 6 illustrates that this goal has been achieved to a very large degree. Although it would probably be possible to design a compression system that is superior to JPEG2000 for some types of image, JPEG2000 currently performs well on more image types than any other standard.

Lossy to lossless

JPEG2000 can be used to compress images with and without loss of information. The lossless capability is very important for the compression of, say, medical and satellite images. But when a digital camera is used to take vacation photos, it might be sufficient to store them with acceptable quality—greater compression permits users to store more images in their cameras.

Progressive transmission

Another important feature of JPEG 2000 is the embedded nature of the compressed file. The most important information pertaining to the image is placed first in the compressed file. Thus, when the image is being transmitted, the recipient receives this information first.

The degree of importance of information in the image is determined when the image is encoded. Three progression modes are available:

- Progressive by resolution (Figure 7)—the image is encoded so that the recipient first receives a low-resolution version of the image followed by the information needed to increase the resolution (step by step).

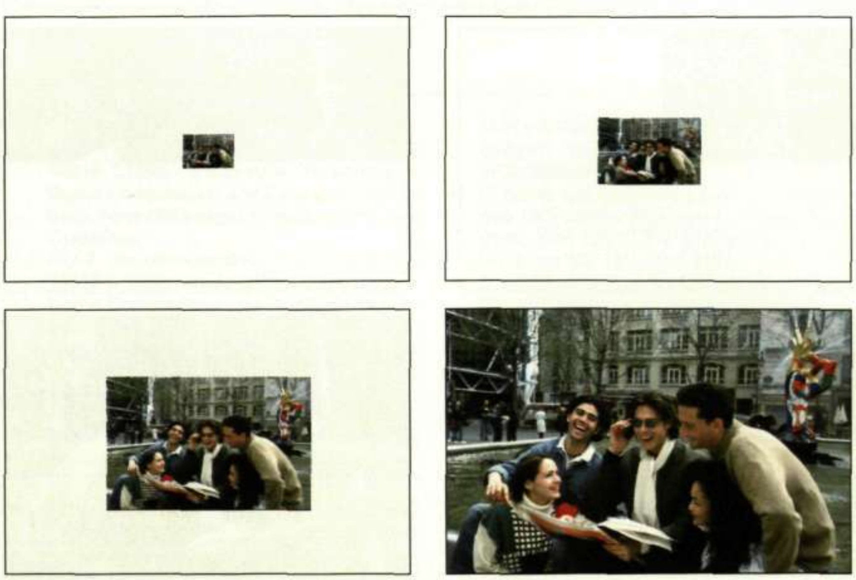


Figure 7
Progressive by resolution. The image is encoded so that the recipient first receives a low-resolution version of the image followed by the information needed to increase the resolution.

Figure 8
Progressive by quality. The image is encoded so that the recipient first gets a low-quality version of the image at full resolution and then receives information needed to improve the image quality.

- Progressive by quality (Figure 8)—the image is encoded so that the recipient first gets a low-quality version of the image at full resolution and then receives information needed to improve the image quality (step by step).
- Progressive by position (Figure 9)—the image is encoded in scan-line order so that the recipient first gets the top-left region of the image at full quality and resolution; the recipient then receives the remaining parts of the image from left to right and from top to bottom.

At any point, the recipient can choose to stop receiving the image and to decompress the information received thus far. For example, a device with a small display that receives an image that has been compressed according to the progressive-by-resolution mode can choose to stop receiving information when the desired resolution has been received.

Error resilience

In JPEG2000 files, there are several ways of increasing resilience against bit errors that occur in the file. Similarly, there are several ways of restarting the decoder at certain intervals, to prevent errors that occur in one part of the image from propagating into other parts of the image when the image is decoded. It is also possible to collect the most important information, such as header data, in one part of the file. This information can then be protected during transmission.

Region-of-interest coding

One requirement for JPEG2000 was the ability to encode different parts of the image at different qualities. Another requirement was earlier placement of information in the compressed file for regions of the image with

Figure 9
Progressive by position. The image is encoded in scan-line order so that the recipient first gets the top-left region of the image at full quality and resolution; the recipient then receives the remaining parts of the image from left to right and from top to bottom.

higher quality (regions of interest). Both of these functions have been implemented in the JPEG2000 algorithm. So when a JPEG2000 image is being received that contains region-of-interest information, the important parts of the image will be received before the background (Figure 10).

Figure 10
Region of interest.

Random access

When a JPEG2000 image is compressed, it is divided into several levels of subdivisions. Each subdivision is encoded separately and can easily be found within the file. It is thus very easy to extract and decode only a desired region of the image.

A standard for the future

As described above, the JPEG2000 standard is intended to be a standard for the future. It will probably not replace existing standards, but it can serve to complement them with a wide range of features and functions. The flexibility and performance of JPEG2000 could make it a good candidate for future use in wireless image applications and next-generation imaging.

Conclusion

Imaging will be a regular part of future mobile services. MMS will probably be one of the most important services in third-generation systems. The use of digital imaging will change as multiservice networks evolve—thus putting new requirements on future imaging formats. Because many different clients will access images, the new format must be flexible, so that it can easily adapt to different clients' capabilities. One such format is JPEG2000, recently finalized by ISO/IEC. JPEG2000—which is one of the first formats that can deal with different types of image and imaging application—is a good candidate for future wireless image applications.

REFERENCES

1 JJ2000 Project, http://jj2000.epfl.ch
2 ITU-T Recommendation T.81 | ISO/IEC 10918-1:1992, Information Technology—Digital Compression and Coding of Continuous-Tone Still Images—Requirements and Guidelines.
3 ITU-T Recommendation T.84 | ISO/IEC 10918-3:1996, Information Technology—Digital Compression and Coding of Continuous-Tone Still Images: Extensions
4 Portable Network Graphics, PNG (Portable Network Graphics) Specification, Version 1.2, http://www.libpng.org/pub/png/spec/
5 ITU-T Recommendation T.800 | ISO/IEC 15444-1:2000, JPEG2000 Image Coding System: Core Coding System
6 ITU-T Recommendation T.801 | ISO/IEC CD 15444-2:2000, JPEG2000 Image Coding System: Extensions, ISO/IEC JTC1/SC29/WG1 N851
7 ITU-T Recommendation T.xxx | ISO/IEC CD 15444-5:2000, JPEG2000 Image Coding System: Reference Software, ISO/IEC JTC1/SC29/WG1 N1851
8 D. Nistér and C. Christopoulos, An Embedded DCT-based Still Image Coding Algorithm, IEEE Signal Processing Lett., vol. 5, No. 6, pp.135-137, June 1998.
9 C. Christopoulos, J. Askelöf and M. Larsson, Efficient Methods for Encoding Regions of Interest in the Upcoming JPEG2000 Still Image Standard, IEEE Signal Processing Lett., vol. 7, No. 9, pp.247-249, September 2000.
10 C, Christopoulos, A. Skodras and T. Ebrahimi, The JPEG2000 Still Image Coding System, IEEE Trans. on Consumer Electronics, Vol. 46, No. 4, pp. 1103-1127, November 2000.
11 WAP Forum, Wireless Application Environment Specification, March 29, 2000, http://www.wapforum.org/

Ericsson's Service Network: a "melting pot" for creating and delivering mobile Internet service

Lars Boman

The Ericsson Service Network, a new solution for creating and delivering mobile Internet service, is based on the new horizontally layered concept—one that separates applications and services from the access and core networks. In this way, the mobile Internet and other communications services, which are accessed by users from any device and any network, converge at the application level. The Service Network becomes a "melting pot" for all types of services and service combinations. It gives service providers a way of managing the complex mobile Internet mass market and of maintaining "ownership" of their subscriber base. For end-users, it provides a personalized service environment, independent of access type.

In this article, the author outlines the need for the Service Network and describes how it will work.

The mobile Internet combines the power of the Internet with the convenience of mobility. Instead of connecting a personal computer or finding an Internet café, anyone with a mobile phone can access the Internet or other online information anytime, anywhere. But the mobile Internet means more than giving mobile access to the Internet. It is more personal. It provides access to services that are based on personal preferences, location, and current circumstances—content context action.

The difference between the Internet and the mobile Internet is akin to the difference between the cinema and television. A whole new genre of services is being opened up by the mobile Internet, in much the same way as TV delivered an enhanced viewing experience to a mass audience in their own homes. The mobile Internet is different from the Internet as we know it because it offers services that

- are relevant in a mobile environment;
- can be efficiently and neatly presented and used in this environment;
- have critical factors that are location-based;
- provide immediacy of reach and response; and
- offer the same basic service set and profile data in all access environments, even if they are presented and sometimes executed differently.

What makes the mobile Internet unique is the fact that the mobile terminal is closely tied to the individual user: terminals are normally switched on and carried by individuals wherever they go. Users can send and receive e-mail instantly. Important news can be "pushed" to users as it occurs. Localized Yellow Pages or street maps are immediately available. The mobile terminal provides secure transactions for online payments, banking and stock trading.

But for mobile Internet services to attract a mass market, certain critical success factors must be in place. First, a variety of suppliers must quickly and easily be able to develop new and attractive services. Similarly, users must quickly and easily be able to tailor services to suit their own individual requirements. Obviously, the services must be very easy to use. And user integrity must be secured at all levels, especially for location-based services. It is not enough that services are safe, they must also be perceived as being safe, particularly for mobile commerce.

An increasingly complex picture

Ericsson predicts that once a mass market starts to develop, the number of mobile Internet users will grow exponentially and each user will demand more and new services. While mobile Internet services will have a role to play in simplifying everyday life, they will also become increasingly complex to deliver. Increased personalization will add complexity; billing will also be more complex; and each application will have to be able to use many different technology platforms (for example, positioning, messaging, and e-commerce systems) in an integrated way.

To be successful, operators and service providers will need to meet increasing user

BOX A, TERMS AND ABBREVIATIONS

3GPP	Third-generation Partnership Project
AAA	Authentication, authorization and accounting
API	Application program interface
ASP	Application service provider
ASUS	Application support server
CAMEL	Customized applications for mobile network-enhanced logic
CRM	Customer relationship management
E-commerce	Electronic commerce
GSM	Global system for mobile communication
HLR	Home location register
IDAE	Integrated distributed application environment
IP	Internet protocol
ISP	Internet service provider
M-commerce	Mobile electronic commerce
MExE	Mobile execution environment
MPS	Mobile positioning system
MVNO	Mobile virtual network operator
O&M	Operation and maintenance
OSA	Open service architecture
OTA	Over-the-air
PKI	Public key infrastructure
PSEM	Personal service environment management
SAG	Service accounting gateway
SCS	Service capability server
SDK	Service development kit
SIM	Subscriber identity module
SMS-C	Short message service center
SNOS	Service Network operation system
USC	User service center
VHE	Virtual home environment
WAP	Wireless application protocol
WISE	Wireless Internet solution
WTA	Wireless telephony application

demand with a constant stream of sophisticated—and complex—new services. They must quickly be able to implement new services and make them available to the mass-market. Operators will increasingly become service brokers, and being first to market with new attractive services will be vital to success. The ability to personalize information by building up individual user profiles and then target information to specific individuals will also be key.

In the long term, there will most likely not be a separate mobile Internet: the mobile terminal will be just one means of accessing the Internet—the home of applications for fixed and mobile users. New and existing players (operators, service providers, content providers and application service providers) have ample opportunities to establish partnerships and gain first-mover advantage. The major players in the mobile Internet market are likely to be

- existing mobile operators—this group of players is already starting to offer mobile Internet services over second-generation networks, thereby gaining vital experience in preparation for third-generation systems;
- greenfield operators of third-generation systems—these players can start the business with a "clean sheet"; that is, without any pre-existing service obligations. They can implement third-generation systems and services at once, although a lack of operator experience might be a disadvantage;
- mobile virtual network operators (MVNO) are new players who do not need to make heavy investments in infrastructure. Instead, they can concentrate on services—the most important part of third-generation networks. A disadvantage is they do not control the network and, by implication, the quality of their services; and
- Internet service providers/application service providers (ISP/ASP)—these players are already well experienced in offering Internet-based services. One problem might be that they will have to charge for services that are currently offered for "free" on the fixed Internet.

In the mobile Internet, new streams of revenue are being created from content provision, advertising, e-commerce transactions, and so on. These revenues will be split between the different players (for example, network operators, content providers, portals and banks).

Shift from vertical integration to horizontal layers

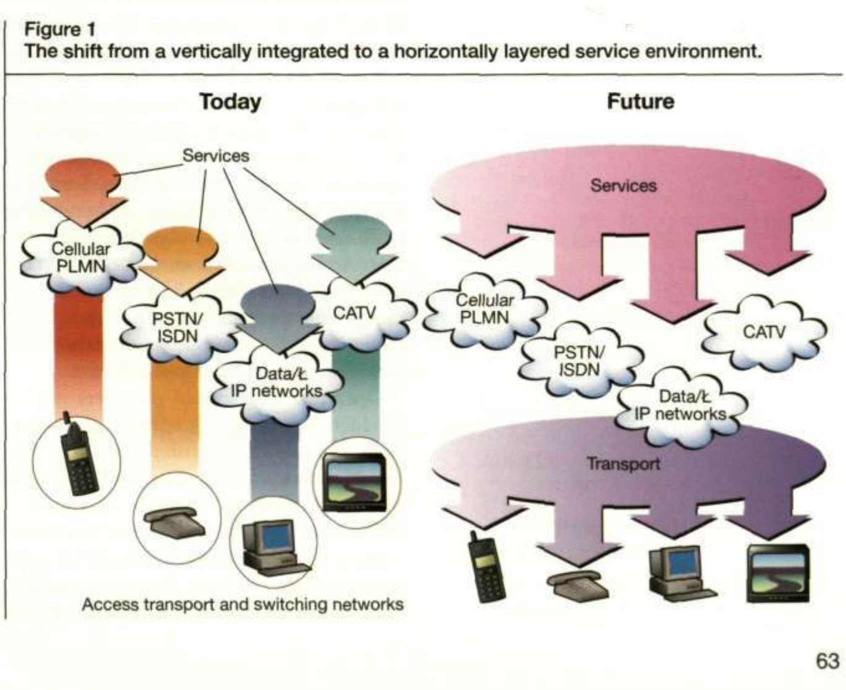
Traditionally, the telecommunications market has been vertically integrated, with applications and services closely tied to the delivery channel—whether the channel was GSM or the fixed circuit-switched network—which resulted in strong vertical segmentation of the supplier market and the associated value chains.

Unfortunately for end-users, Internet services are not uniformly managed or standardized, which makes global access to a set of Internet-based services difficult to achieve.

Perhaps paradoxically, the effect of the mobile Internet and third-generation systems is to improve this situation, by decoupling applications from the underlying infrastructure and by forcing the development of open standards.

The traditional vertical business and technology segmentation is beginning to tilt, so that the value chain is segmented horizontally, where virtually any application or service can be provided over any underlying network technology (Figure 1). Consequently, the mobile Internet business will be driven from the operator or service provider toward the enterprise or content provider.

Figure 1
The shift from a vertically integrated to a horizontally layered service environment.



supports the WAP gateway, WebOnAir filter proxy, SIM application toolkit, OTA SIM management, and a messaging solution, it does not support all enablers. It will thus be further developed to host the WISE Portal 2.1, together with some applications and additional enablers.[2] The user service center will migrate toward an open and flexible architecture based on the open-standard architecture specified by the 3GPP.

Benefits for operators and service providers

The Service Network gives operators, service providers and third-party developers an attractive and easy-to-use interface for developing applications. It has been designed to help operators and service providers to build successful mobile Internet services by offering

- a range of new applications and services, for example, through the Developers' Zone alliance program;
- efficient operation and maintenance with a common database and uniform subscription interfaces;
- end-user monitoring and satisfaction through customer relationship management (CRM) tools;
- a wide range of reliable service enablers; and
- expert consulting services to support strategic decisions, including advice, systems integration and facility management for operators, service providers and enterprises.

The Ericsson Service Network is designed to help operators and service providers to bring interesting and advanced services to market in the fastest, most convenient way, and to keep integration costs as low as possible. For example, the Service Network SDK will use standardized APIs, such as Parlay, Jain and other open-industry interfaces. The application developer can thus concentrate on the creative part—features—and leave the database integration, site management, user interface and access to the right core network functions to the Service Network.

The cost of introducing large quantities of new services rapidly needs to be low.

The Service Network will become a plug-and-play environment, helping operators to achieve short time-to-market and to establish a reputation for being the first to offer advanced services.

It will also be important to keep track of which services are successful, and which are not. Some services will be long-runners, while others might be an enormous but short-lived hit. One way that operators and service providers will be able to keep track of the success and usage of services is through the Service Network's CRM system, which helps identify users' personal preferences, thereby improving service targeting and enabling bundling and niche marketing campaigns. It will also show how successful individual services are and indicate when they should be removed to make way for new services.

How the Service Network works: an example

Ericsson has produced many different scenarios to illustrate the Service Network. What we believe will be one of the most important issues for success is how new services can be developed, operated and launched.

Imagine a third-party application developer who works in close partnership with a mobile operator. The developer can enter the Developers' Zone solution of the operator, hosted within the Service Network, and log on to his own Developers' Zone page.

Once logged on, the developer can see how existing services are doing—for example, how many people are using them, revenue earned, and so on—and create new applications.

To create an application, the developer uses a service development kit—made available by the operator—which can be likened to a toolbox for application development. The developer downloads the open APIs he needs to create the application, for example, for positioning, WAP, mobile e-commerce

TRADEMARKS

and video streaming. When he has finished developing the application, he can test it on a simulator or test site provided by the operator. Once the developer is satisfied with the application, he can upload it to an application server.

Finally, the application developer "signs" the new application to confirm that it is ready, perhaps also sending some information to the operator to indicate how he wants it to be marketed or how the operator should charge for its use.

When this is done, a message is automatically sent to the operator's O&M system. The operator then tests the application to ensure that it meets standards for, say, quality and decency. The operator can then make the application available to his customers.

The application can be categorized according to the operator's own model—for example, according to application type (sports, movie, or news) and functionality (positioning, messaging, or m-commerce).

Finally, the application is put on a publicly available application server. The new application can be marketed in a highly targeted way, by automatically sending a message to the website or mobile phone of customers whose profiles match the application categories.

The operator could also use the Service Network capabilities to send messages to members of the Developers' Zone, launching a competition for, say, the best sports application.

The Service Network enables the operators to have close or loose partnerships with application developers, perhaps giving certain developers special opportunities to test applications live and to stipulate how use of the applications is to be charged.

Conclusion

A highly flexible and open architecture is needed to deliver the mobile Internet's promise of "personalized services for the masses." Moreover, service providers must be able to create and offer personalized services quickly and easily, which is the role of the Ericsson Service Network.

Mobile Internet subscribers will demand—and get—a service environment that is tailored to their own personal requirements. They will want to access it using whichever device or method is most convenient at the time. The Ericsson Service Network fulfills this criterion, providing a vital layer between connectivity networks and the end-user.

Because its open architecture is standards-based, the Service Network enables operators to easily integrate new service and application ideas from third parties, whenever these ideas come along.

The Service Network is the IP-based "glue" that binds together the many different access media, core networks, content and service providers and user devices for seamless service delivery. It handles virtually any type of service and convergence of telecommunications, data communications, Internet and multimedia services. This convergence will enable new service combinations across numerous kinds of access network and across traditional service domains.

Using the Service Network as a platform, Ericsson can provide

- open, standardized interfaces;
- partnerships with content and service providers;
- support for third-party developers and application centers; and
- integration and consulting expertise.

Ericsson's goal is to enable mobile operators to add value to their services on their own terms, in a highly manageable way.

The Service Network approach is designed to provide performance, reliability and low cost of ownership. The open, scalable architecture of the Service Network is designed to provide economies of scale, while reducing development time and enabling easy integration of new technology as it becomes available. It offers mobile operators a low time and cost threshold for testing and introducing new applications, thereby reducing the business risk.

REFERENCES

1 Johnston, A., Papanikolaou, T. and Sissingar, M.: The WISE Portal 2.0 solution—Timely delivery of tailored mobile Internet services. Ericsson Review, Vol. 78(2001):2, pp. 68-79.

2 Swedberg, G.: Ericsson's mobile location solution. Ericsson Review, Vol. 76(1999):4, pp. 214-221.

The WISE Portal 2.0 solution—Timely delivery of tailored mobile Internet services

Andy Johnston, Thomas Papanikolaou and Mike Slssingar

Two of the hottest issues in mobile communications today are the growth of the mobile Internet and the evolution of new portals. Both are interdependent. Mobile Internet growth requires new portals. The creation, deployment and continued enhancement of new portals require a world-class portal infrastructure, combined with attractive and creative applications. And there must be a well-defined methodology for delivering the portal and the applications in order to capture a viable share of the market.

Ericsson is working across several industry sectors to develop mobile Internet services and applications for mobile operators around the world. The authors describe the WISE Portal 2.0 solution, which is a cornerstone of this work.

Today's mobile communications users have high expectations at a time when deregulation and open systems have made the competitive landscape more aggressive. The mobile Internet offers a golden opportunity for creating new services and new revenue streams. The key to unlocking these opportunities is the ability to deliver services and applications that work, are available at the right time, and offer personalization.

The tool that enables mobile operators and service providers to deliver personalization is the portal. Over the past decade or so, the development and operation of portals over the fixed-access Internet have taught us many lessons, and these have been brought to bear on the development of Ericsson's WISE Portal solution.

In the mobile environment, the portal enables its owner—the network operator or service provider—to tailor mobile Internet offerings to match the market requirements. For the user, the mobile portal provides a familiar, personalized service interface to employ the mobile Internet as a valuable yet simple tool that makes everyday life easier and more fun. An effective mobile portal helps ensure rapid end-user acceptance of mobile Internet services.

Ericsson's WISE Portal 2.0 is designed for the creation and delivery of targeted mobile Internet services that are user-friendly and closely tailored to individual users' needs. The solution has a crucial role to play in helping mobile users to access the services and information they want quickly and with a minimum number of clicks. Moreover, users can personalize the portal to suit their own preferences and needs.

The WISE Portal 2.0 seamlessly integrates the Internet and wireless application protocol-based (WAP) services so that they can be accessed from any personal computer, WAP phone or other mobile device.

Satisfying market expectations

The market has high expectations of portal solutions. First, the solutions must be usable. The WISE Portal 2.0 solution provides a unified user interface, called *bricks*, for WAP and Web access.

The solutions should also be scalable. The WISE Portal 2.0 solution can be scaled, for example, across processors and nodes, and in terms of the number of registered users, concurrent users, and content feeds that it can handle. The solution is available in several standard configurations.

The portals should offer high reliability and availability. The WISE Portal 2.0 solution can be supplied with "four-nines" (99.99%) or better availability.

Flexibility is also important. The WISE Portal 2.0 solution has a component-based architecture that enables flexible configurations as well as integration into portal operators' enterprise systems.

The portals must be global. The WISE Portal 2.0 solution can be deployed internationally and supports wide (16-bit) character standards, such as the universal transform format no. 8 (UTF-8) and unicode character set no. 2 (UCS-2).

BOX A, TERMS AND ABBREVIATIONS

3GPP	Third-generation Partnership Project	OSA	Open service architecture
API	Application program interface	PDA	Personal digital assistant
DNS	Domain name server	RADIUS	Remote authorization dial-in user service
DTD	Document type definition	RDBMS	Relational database management system
EJB	Enterprise JavaBeans	SDK	Software development kit
GPRS	General packet radio system	SMS	Short message service
GSM	Global system for mobile communication	SMS-C	SMS center
GW	Gateway	SSL	Secure socket layer
HDML	Handheld device markup language	UCS	Unicode character set
HTML	Hypertext markup language	UMTS	Universal mobile telecommunications system
HTTP	Hypertext transfer protocol	USC	User service center
ISP	Internet service provider	UTF	Universal transform format
IT	Information technology	WAP	Wireless application protocol
J2EE	Java 2 Enterprise Edition	WCDMA	Wideband code-division multiple access
JS	JavaScript of Java Servlet	WML	Wireless markup language
LDAP	Lightweight directory access protocol	XML	Extensible markup language
MAI	Mobile application initiative	XSL	Extensible stylesheet language
MSC	Mobile switching center		
O&M	Operation and maintenance		

The mobile environment puts additional demands on portal solutions. For example, mobile network operators generally have more stringent quality requirements than traditional portal operators. In all likelihood, mobile portals will have much greater numbers of simultaneous users than fixed portals. They will also have special integration requirements, for example, for charging and pre-payment systems. Similarly, strict conformance testing will be required during network integration. These demands for system quality have an impact on scalability, reliability, availability, security and interoperability—the task of managing the design, implementation and delivery of the portal is thus vital.

Therefore, a key element of Ericsson's WISE Portal 2.0 solution is the delivery methodology, which covers all aspects of the development of mobile Internet services and associated platforms, including the Service Network[1], positioning, transactions, integrated billing, portal core technology, an observational database, and personalization.

Ericsson has a mature solution-development methodology called *Framework*, which focuses on the business implications, design, and development, as well as the implementation and operation of the applications and platform on which the services run. Framework supports the development of customers' business and continues throughout delivery to maintain the momentum of each complex and concurrent element. It also includes management consulting, a project office, a systems design team, and applications development.

BOX B, THE WISE PORTAL 2.0 SOLUTION AT A GLANCE

WISE Portal 2.0 is a complete, end-to-end solution for the creation and development of mobile portals (Figure 1). With it, portal operators can choose to install and manage the portal in-house, or opt for an Ericsson-hosted portal, which comes complete with a generic set of applications and services.

The WISE Portal 2.0 meets the challenges presented by the mobile market by combining a number of key components:

- Ericsson WISE 2.0 Portal presentation layer and presentation layer bricks;
- BroadVision One-To-One enterprise application server;
- J2EE application server;
- Ericsson USC 1.2E; and
- Ericsson mobile Internet enablers.

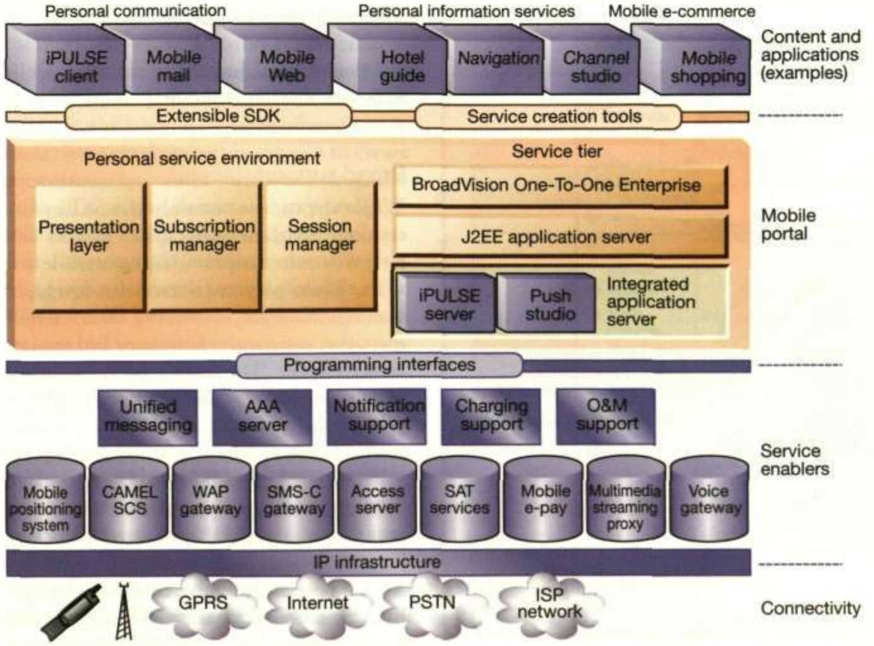


Figure 1
Overview of the WISE Portal 2.0 solution.

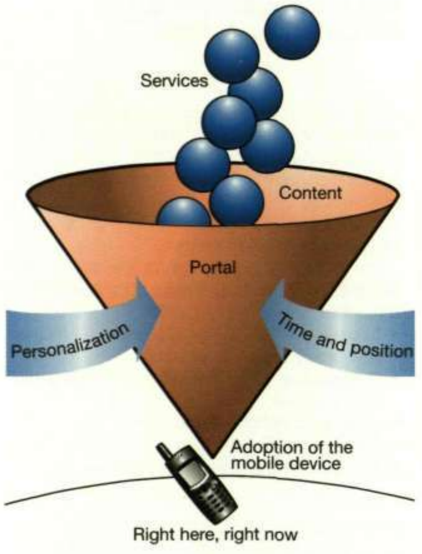


Figure 2
The "anywhere-Internet" concept.

The power of mobility

A solution that supports several important service attributes is needed to realize the "anywhere Internet" concept (Figure 2). These attributes are

- personalization;
- time and position;
- device support;
- services; and
- local content.

Personalization is key to mobile Internet services. Advanced personalization features help users to access information quickly and easily. Personalization and profiling features enable operators to serve their customers better—enhancing the "stickiness" of their offerings, encouraging customer loyalty, reducing churn, and growing the user base.

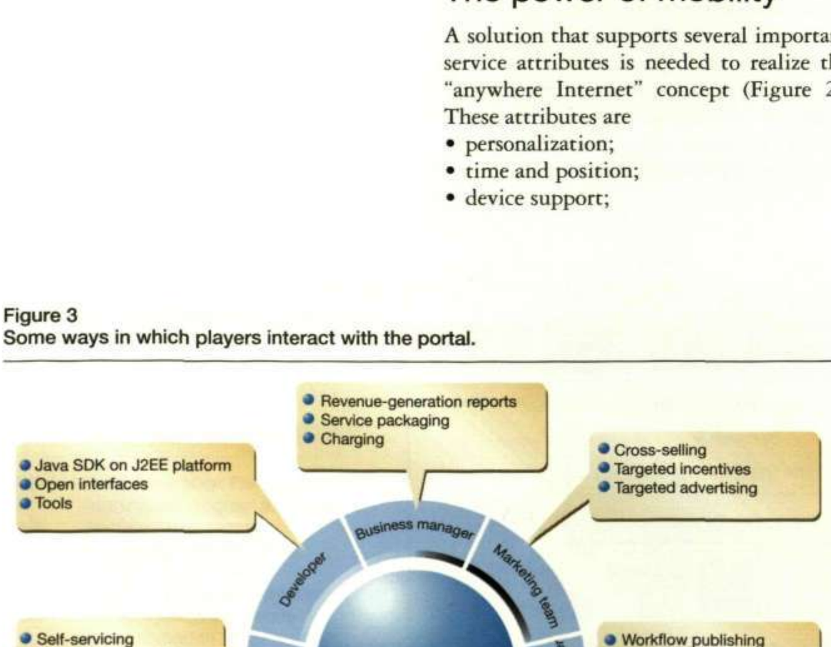
Intelligence in the network gives information on user whereabouts. When combined with time-of-day awareness, this information can be used to deliver highly targeted and relevant services.

The ability to support multiple devices for the delivery of mobile portal services is also a critical factor for success.

Compelling mobile Internet services are key to widespread take-up.

In mobile contexts, local content is of much greater importance than in the fixed Internet. Content must be filtered and delivered in a format that can be understood by the user.

Figure 3
Some ways in which players interact with the portal.



Market requirements

Ericsson has been driving the mobile Internet since its inception, and this has helped the company identify the key success factors for mobile portal-based services. In addition to meeting the market expectations outlined above, the mobile portal requires

- broad support among industry players;
- best-of-breed components;
- personalization;
- content management;
- service development; and
- universal delivery.

Broad support

While the user is certainly the focus of attention, there are many other players who have key roles in provisioning mobile services. These players include the developer, business manager, marketing team, content manager, system administrator and the customer's agent. The mobile portal solutions must support these different players (Figure 3).

Best-of-breed components

Industry-wide open interfaces enable components to be mixed and matched. Excellence in design, performance, and availability is essential, as is choice, because monolithic or proprietary solutions are generally unacceptable today.

The guiding principle in selecting components for the WISE Portal solution has been to choose the best-of-breed components from third-party suppliers in each major product area (directory server, application server, firewall, and so on). This requires evaluating the components against a rigorous set of criteria to ensure that they are feature-rich and of high quality. The components need to be proven market leaders. And the vendor must be reputable and have relevant expertise.

Ericsson then contributes its own business and technical know-how and experience, adding value to *de facto* standard architectures and developing third-party components for use in the mobile environment. The reuse of tried and tested technologies and components cuts time to market and enhances quality.

Personalization

Personalization is a proven way of creating customer loyalty and generating repeat visits to mobile portals. Advanced forms of personalization can be created through new mobile network features, such as positioning (Figure 4).

The WISE Portal allows users to personalize their services in terms of presentation, service and service category selection, service delivery mechanisms, and content, and can provide automatic personalization based on an analysis of user behavior.

A portal solution that offers multi-dimensional personalization must provide concepts as well as features and tools that allow operators to immediately deliver and benefit from personalized services. The WISE Portal provides a range of easy-to-use tools that portal operators can use to create personalization rules without programming. These tools enable editors and content creators to produce and manage content—for example, while an extensive profiling system records user preferences. The WISE Portal solution can also observe user actions and learn from them.

Content management

Content management is key to offering compelling portal services. Portal solutions must provide comprehensive and efficient support for content management, especially when large numbers (100 or more) of content providers are involved.

The main content-management functions are creation, editing, removal, classification, integration, scheduling and staging. Other functions include managing adapter connections and controlling the quality of content feeds (Figure 5).

Several qualitative characteristics are also important. These include the ability to manage and process both static and dynamic feeds, and to process real-time feeds. The portal should be able to accommodate pop-

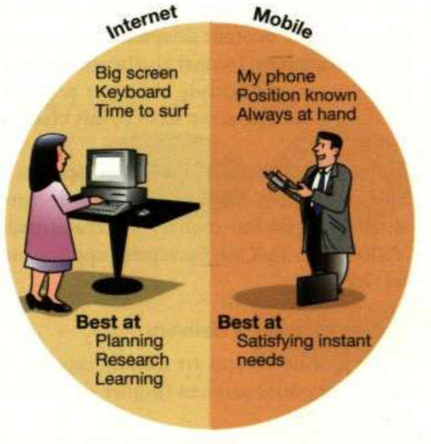


Figure 4
Personalization means a service that immediately satisfies needs regardless of user location.

Figure 5
Content management using uncomplicated procedures.

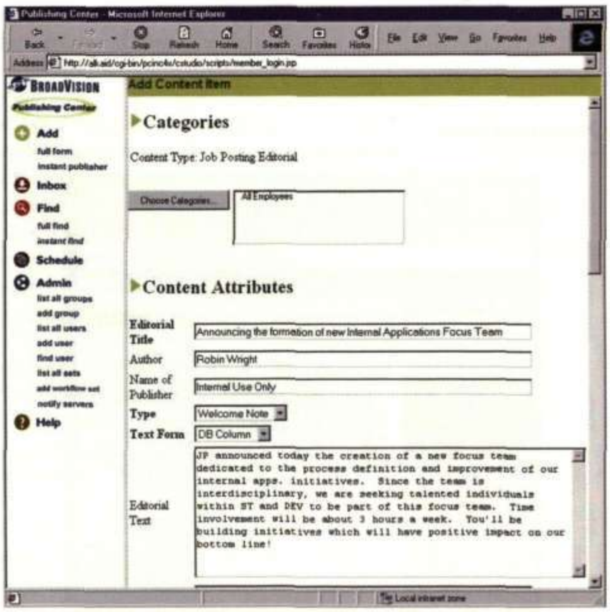


Figure 6
Mobile portals need to be independent of the devices that access them.

ular data feeds (such as newswire services) via packaged content adapters, and secure access to content. It must also be able to handle content in multiple character sets, including wide character sets (16-bit characters).

The WISE Portal 2.0 solution includes a unified content-management architecture with a supporting tool set via the BroadVision One-To-One Enterprise application service.

Universal service delivery

Many mobile devices are already capable of accessing portal services (Figure 6). Examples include

- devices that can handle short message service (SMS) notification;
- WAP-enabled phones;
- smart phones and personal digital assistants (PDA) that can handle WAP, handheld device markup language (HDML) or hypertext markup language (HTML);
- voice-browsing systems; and
- interactive television.

Mobile portals need to be independent of the devices that access them. They must also provide a universal delivery service.

Service development

Portal platforms must offer the rapid development and deployment of compelling services. Portal developers must be able to create simple services without complex programming—for example, through rule-based programming and profile- and content-driven service behavior. They must also be able to create complex services using standard programming languages and environments, such as the Java 2 Platform, Enterprise Edition (J2EE) server-side Java standards. Similarly, mobile portal operators or their subcontractors need a well-documented set of service-development concepts and interfaces with which to create services.

The WISE Portal uses the BroadVision One-To-One enterprise application server to support the creation of release-based profile- and content-driven services. Likewise, it uses the BEA WebLogic J2EE application server as a deployment platform (and any of the myriad Java development tools available on the open market as development environment) to support programmatic service creation.

A Java-based software development kit (SDK) provides additional programmatic service support through key system interfaces to

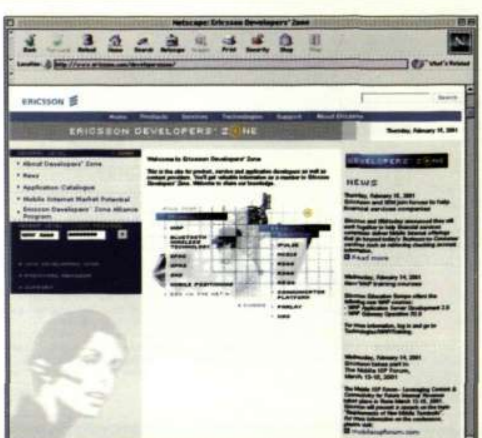
- subscriber directory access;
- charging and billing services;
- SMS notification services;
- presentation layer rendering services;
- operation and maintenance (O&M) and logging facilities; and
- session-management facilities.

Mobile portal solutions must extend service development to areas that add value in a specifically mobile context, for example, through interfaces to the mobile positioning system[2] and terminal characteristics database. In addition, mobile portal solutions must accommodate the interfaces currently being specified by the Third-generation Partnership Project (3GPP) and Parlay for the third-generation open service architecture (OSA).

Ericsson is also encouraging third parties to develop applications and services. The Ericsson Developers' Zone has been created to offer third-party application developers access to a wide range of Ericsson mobile application technologies and product (Figure 7). After registering with the Ericsson Developers' Zone, developers gain access to information and tools that enable them to build and test applications. The Ericsson Developers' Zone also includes an Alliance Program for companies who work closely with Ericsson on the development of mobile Internet applications and content.

Ericsson's Mobile Applications Initiative (MAI) program is designed to drive the

Figure 7
The Ericsson Developer's Zone.



GPRS and WCDMA/UMTS applications industry and to expand knowledge, experience and opportunities in these technology sectors. The aim is to stimulate market interest and to ensure that high-quality applications become available.

WISE Portal feature round-up

The WISE Portal 2.0 contains four key components that combine to create a complete solution for delivering personalized, easy-to-use, portal-based services to mobile users:

- the WISE Portal presentation layer;
- BroadVision One-To-One;
- the J2EE application server; and
- the user service center (USC).

WISE Portal presentation layer

The presentation layer, which is the user interface to mobile portal services, provides a scalable and robust extensible markup language-based (XML) publishing framework for portals and is capable of presenting services on multiple devices using multiple markup languages.

Using XML, the presentation layer clearly separates presentation from business logic. It also provides a specification that allows users to develop new or to integrate existing services into the presentation layer architecture. Services delivering plain HTML or plain wireless markup language (WML) can also be integrated.

The presentation layer provides extensive capabilities for branding, and for specifying the "look and feel" and behavior of the portal (the personality). It can support multiple "personalities" and comes with a default personality (bricks).

Presentation layer components

The connection manager is responsible for receiving a user request, transforming it into the standardized internal format of the presentation layer (Figure 8), and establishing the connection between the end-user and an application. Establishment of the connection usually includes the creation of a user session and additional security checks. This is checked by the user manager. To provide scalability, the connection manager is implemented using Java Servlet technology.

The request broker is responsible for retrieving the data requested from the user, by accessing the application's communication protocol. All information that is relevant to the application (such as the communication protocol) is stored in the application manager.

XML data returned by applications is forwarded to the rendering engine, which parses, validates, and transforms the data appropriately for the user's device, using device information retrieved from the device manager and extensible stylesheet language (XSL) transformations.

Presentation layer repositories store all resources required for validating XML data sent by applications (document type definitions, DTD), transforming XML into a language understood by the end-user's device, and so on. The repositories enable resources to be associated with applications, and provide versioning facilities for the publishing framework.

The basic services are a minimal set of interfaces on which the presentation layer is implemented. The presentation layer can function with any component that provides sufficient functionality for implementing the basic services. In the current implementation, the basic services are implemented using Ericsson's USC user and application manager and the WISE Portal 2.0 session and device manager.

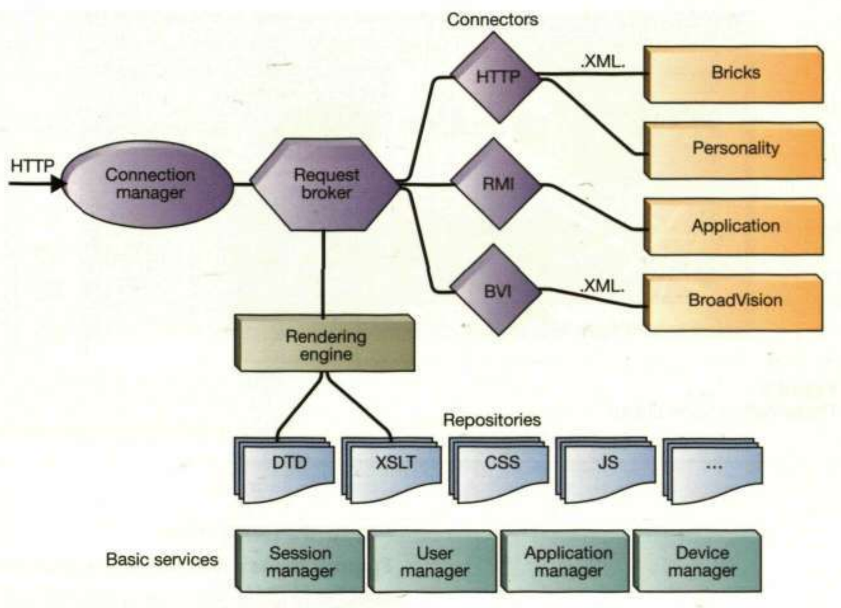


Figure 8
The architecture of the WISE Portal presentation layer.

Figure 9
Presentation layer bricks.

Presentation layer bricks

Figure 9 shows a typical bricks portal entry page for HTML. The portal user is presented with a set of bricks, grouped into six categories (shown as tabs above the bricks grid. In this example, the tabs are Favorites, Finance, David, Travel, Messaging and Leisure).

Figure 10 shows the WAP variant of the interface described above. The tabs are shown on the WAP deck as a set of links. Selecting the Favorites link calls up the presentation of a second WAP deck with a set of links for each brick under Favorites.

Figure 11 shows how the user can personalize the bricks interface by dragging service bricks from the pane on the right to the bricks grid on the left.

Figure 10
WAP screenshot of presentation layer bricks.

Figure 12 shows a single mobile trading brick ant its associated set of controls (*max*, *config*, *help* and *info*).

BroadVision One-To-One

BroadVision One-To-One provides content management, personalization, and rule-based service-development functions within the WISE Portal 2.0 solution (Figure 13). BroadVision and Ericsson have cooperated to design and develop a set of components that bring key BroadVision One-To-One functionality closer to the mobile network. These include interfaces to SMS gateways, charging systems and subscriber databases.

BroadVision One-To-One enables the development of portals that treat users as unique individuals, thereby increasing the level and regularity of user activity. It remembers and tracks data and transactions from visit to visit, stimulates transactions and feedback, and supports continuous improvement in the level and scope of personalized services.

Content management

The BroadVision One-To-One content-management features are the heart of the WISE Portal unified content-management architecture, providing authoring, workflow, scheduling, classification, tagging, persistence, audit trails of uploads and downloads, and previews. There are six types of pre-defined content, each with extensive semantics:

- templates;
- product catalogs;
- editorials;
- discussion groups;
- incentives; and
- advertisements.

A unique feature of BroadVision One-To-One is that it allows content creators to categorize the content in hierarchical folders. Matching attributes can be attached to content to enable powerful personalization. For example, a content creator can attach language, product, job title and level of technicality to each item which is then matched with the user's profile to show only the relevant content.

BroadVision also includes a publishing center (Box B). This is a configurable, Web-based tool designed for all participants in the content-management lifecycle, from full-time content professionals (who create, maintain, schedule, and manage content) to casual content contributors (who occasionally author and submit content.

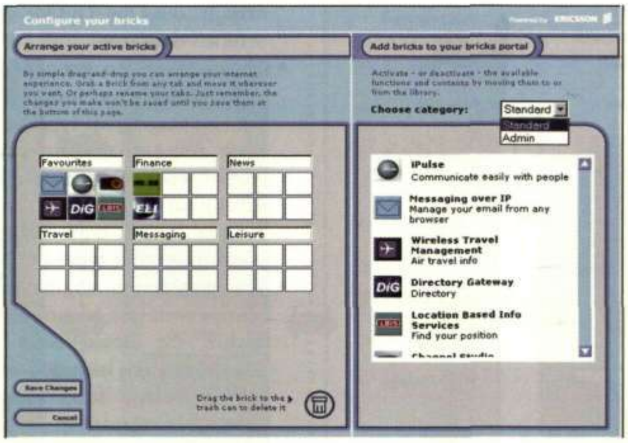


Figure 11
The bricks interface.

BroadVision One-To-One has a comprehensive user-profile concept and architecture, which forms a sizeable proportion of the WISE Portal approach to user profiling. This concept includes an extensive, persistent user profile that can be used by all parts of the BroadVision system as well as externally. There are two main sources of information on user profiles:

- the WISE Portal user information component, which is based on the USC subscription manager and is used to store relevant, general, cross-application information on the user profile; and
- the BroadVision One-To-One profile manager, which is used to store BroadVision application-specific profile data.

User profiling

BroadVision One-To-One tracks users as either members or guests of the site. Members have registered, retrievable profiles; guest profiles are typically limited to the current session. User profiles are stored in database tables (internal or external relational database management system, RDBMS). The contents of a user's profile can be extended and changed by the portal operator.

Session profiling

The BroadVision One-To-One command center can be used to create session rule sets that contain references to session and user-profiles attributes. By creating multiple session rules, it is possible to have a set that assigns different values to the same session profile variable for different visitors.

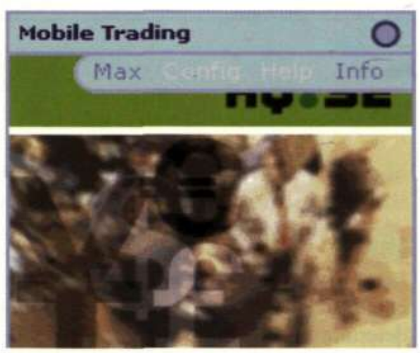
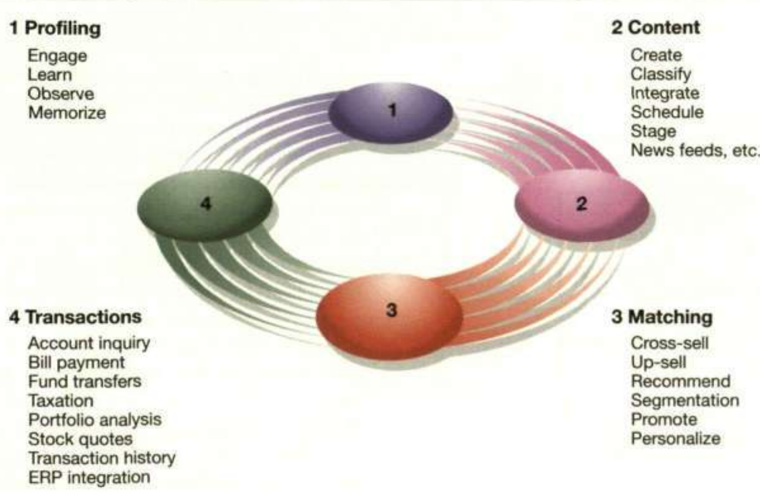


Figure 12
Mobile trading example.

Figure 13
BroadVision One-To-One functionality.



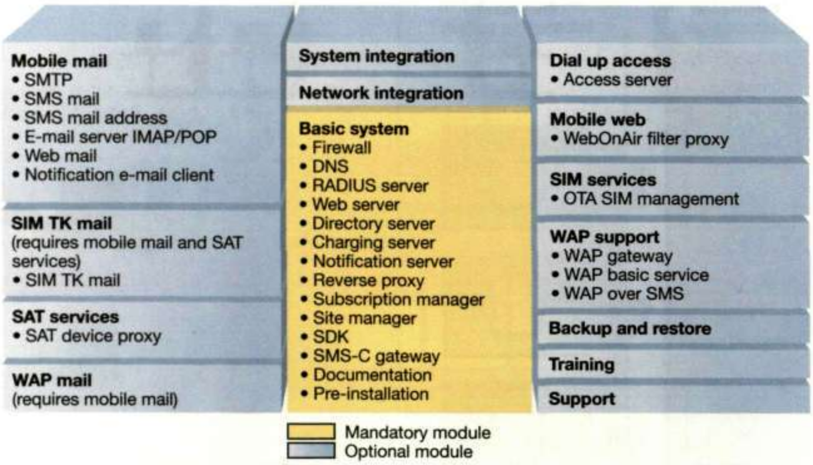


Figure 17
An overview of the USC.

BOX C, KEY FEATURES OF THE BROADVISION PUBLISHING CENTER

Increased publishing process control—content publishers can control the publishing process by
- granting and restricting publishing rights to individual users;
- subjecting content to a workflow process in order of appearance; and
- previewing content before it appears live on the site.

Increased publishing participation—content contributors can submit and edit content using Instant Publisher forms, which are tailored and personalized to each user's access and skill level.

The open solution leverages existing content in the enterprise (for example, in Documentum, Lotus Notes, Microsoft Office) and works with existing authoring tools, such as Macromedia, and versioning tools, such as Interwoven.

Flexible, rule-based interface—Instant Publisher provides personalized and customized forms for casual content contributors and sophisticated content-management tools for content editors.

Content profiling allows authors to profile content, which in turn, enables BroadVision matching agents to deliver personalized content by matching content and user profiles.

Browser compatibility—the Java-based solution supports Web-based publishing access and administration using a Java-compliant browser.

Open content support for content created from any tool, including HTML editors, Microsoft Office products, and Lotus Domino; ability to pull content from external content and document repositories including Documentum and Lotus Notes.

Publishing access control makes it possible to create and populate customizable access-control groups that grant rights (read, edit and approve publishing) to individual publishers.

Publishing workflow (customizable workflow states to support any publishing process) supports multiple start states, state branching, and "pass-back." Each portal can use multiple workflows.

Content scheduling makes it possible to manage a content programming calendar and to schedule content to go online and offline automatically.

Content preview before a presentation is put online.

Instant Publisher forms (simple, personalized, and easy-to-use publishing forms), which are tailored to each user's publishing tasks and skill level, make use of the powerful One-To-One Publishing Center functions, but solely expose what is needed to casual contributors.

The BroadVision One-To-One Publishing Center seamlessly integrates info InterWoven's TeamSite for content versioning and roll-back. Other versioning vendors' products can also be integrated.

The USC can be scaled in terms of performance, availability characteristics, and number of users. It consists of modules (Figure 17) for
- basic IP infrastructure support;
- application support functions;
- operation and maintenance support;
- service capability servers; and
- application and service development support.

The USC is built from standard hardware components, taking advantage of the latest developments in the information technology (IT) industry. This approach ensures that the USC always offers the best performance and provides world-class availability characteristics with trouble-free operator experience at a very competitive price.

In terms of software, the USC consists of a mix of industry-standard software and components developed by Ericsson. This mix gives the right balance between short time to market and suitability for the operator market. In July 2000, USC 1.2 went into commercial operation with operators in Europe and Asia.

Security

Because the USC is connected to external systems, such as the Internet, its services must be protected against unauthorized use. The USC thus provides firewalls, switches, reverse proxies, and remote authorization dial-in user service (RADIUS) servers. Support for the secure socket layer (SSL) protocol and other security features are also provided.

Directory server

Central to the USC architecture is a directory server that maintains all information on subscribers and the services to which they subscribe. The directory server is based on the lightweight directory access protocol (LDAP), to help ensure high performance and conformance to standards. The implementation is currently based on the Netscape Directory Server, which runs on Sun Solaris.

Web server

Many applications require the use of a Web server to host HTML and WML documents and applications. A highly scalable and reliable Web server is tightly integrated into the security architecture of the USC. The Web server can also be used for third-party and operator-specific applications.

Charging service

For services managed by the USC, a built-in charging service offers a variety of charging events. All services can generate charging records by calling the charging API. The subscription manager automatically generates a charging record when an end-user subscribes to or unsubscribes from a service. The charging service software (developed by Ericsson) runs on Sun Solaris.

Site manager

All components within the USC system, as well as the applications it manages, are supervised and managed by the site manager. Operators can thus

- manage complex multi-vendor information systems; and
- control resources, such as network components, computers, operating system, databases, and file stores.

The site manager is based on HP OpenView building blocks and a database whose management console runs on Sun Solaris.

Service capability servers

In current (GSM) mobile networks, an SMS center (SMS-C) is responsible for sending, receiving, storing and forwarding short messages between mobile terminals and servers that use proprietary protocols for communicating with applications.

The USC includes an SMS-C gateway that shields applications from the specific details of each vendor's SMS center. The gateway is based on the Across Wireless Transport Server, which runs on Sun Solaris and supports the major brands of SMS centers.

Application and service development support

The USC is an open system, which—thanks to the service development kit—can serve as a platform for components delivered with the USC, third-party equipment and existing operator equipment. The SDK gives interfaces to charging, subscription-management and notification functions.

Conclusion

More than anything else, the mobile Internet is set to be the most dynamic of markets the industry has ever addressed. Competition encourages users to shop around for services, and fashions change almost overnight. Keeping ahead of the market is one of the secrets of success. This has been a key consideration in the development of the WISE Portal solution.

The open architecture of the WISE Portal enables new services to be created and new technologies to be introduced easily and quickly. *Framework* (the design, implementation and delivery) includes market monitoring and feedback procedures which ensure that changes in taste and demands are identified and responded to immediately.

The WISE Portal 2.0 solution provides a highly adaptable gateway between service providers and users where mobile Internet services can be adapted to meet the changing needs of the market and individual users.

REFERENCES

1 Boman, L.: Ericsson's Service network: a "melting pot" for creating and delivering mobile Internet service. Ericsson Review, Vol. 78(2001):2, pp. 62-67.

2 Swedberg, G.: Ericsson's mobile location solution. Ericsson Review, Vol. 76(1999):4, pp. 214-221.

TRADEMARKS

Supply as an enabler in the new telecoms world

Andreas Grundsell and Karin Mallmin

Building the third-generation network infrastructures needed for mobile Internet services will require the production and delivery of base stations and network nodes in very large volumes.

At the same time, the demand for infrastructure equipment for GSM and other second-generation networks will continue to grow as existing networks expand to cope with the twin pressures of subscriber growth and traffic growth.

The supply processes that have brought the mobile communications industry this far will not be adequate to cope with the extreme market demand for equipment for second- and third-generation infrastructure. This is why Ericsson is developing supply concepts that have a lot in common with the techniques used in large-volume businesses, such as the automotive and consumer-electronic-goods sectors.

In this article, the authors focus on the likely impact of market growth on the supply flow, and profile the supply management initiatives being taken by Ericsson to meet the expected demand from mobile network operators.

Introduction

With more than 650 million mobile phone users in the world today—a number that is expected to reach one billion in the next two years—the mobile phone has clearly become a mass-market commodity (Figure 1). It has become as much a part of people's lives as televisions, cars, radios and washing machines. In fact, if you compare the mobile phone with any other consumer electronic product, the fact that its customer base grew from zero to 600 million in 15 years puts it in a league of its own.

Yet it is interesting to look back to the late 1980s and remind ourselves that the mobile phone was then regarded as an accessory that could only be afforded and justified by business people. In 1988, the industry was forecasting that by the year 2000, the number of mobile phone users might reach 20 million in western Europe.

Every couple of years since then, the industry raised the forecasts in light of growing market demand. Each time this happened, the manufacturers of infrastructure equipment (radio base stations, mobile switching centers, and so on) had to gear up for greater volumes.

As the main supplier of mobile infrastructures, with a global market share of around 40%, Ericsson had a bigger challenge than most companies each time the market growth exceeded expectations, since it also had to increase the production of necessary infrastructure equipment.

The fact that the company managed to maintain its global market share at a fairly constant level throughout this decade of rapid growth says much about the effectiveness of the supply strategies that have been developed. But as the world prepares for the build-out of third-generation mobile network infrastructures, it is clear that even more aggressive supply strategies will be re-

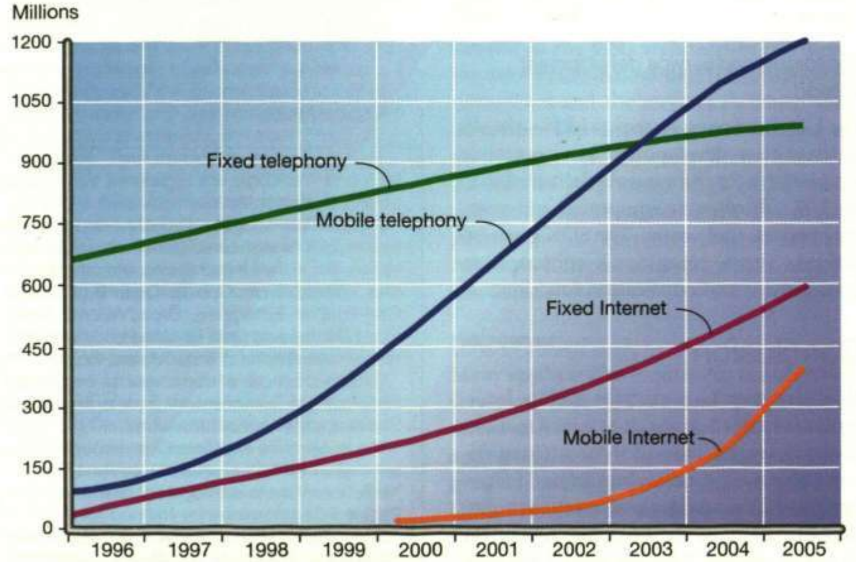


Figure 1
Subscriber growth.

quired (Figure 2). Mobile network operators are planning for

• rapid migration to third-generation resources; and
• rapid increases in traffic in the networks.

One scenario portrays an explosive growth of traffic in mobile networks, due to swift migration from fixed to mobile telephony and rapid growth in data traffic. This scenario represents the consequences of an operator's change in focus from subscriber growth to increased use of services by the subscriber base.

The industry is poised on the brink of another massive expansion phase, but this time there is greater pressure on shorter time to revenue than ever before. And when time and cost are essential, supply flow is critical.

BOX A, TERMS AND ABBREVIATIONS

BSC	Base station controller
FC	Flow control center
GPRS	General packet radio service
GSM	Global system for mobile communication
MSC	Mobile services switching center
OSS	Operations support system
TTC	Time to customer
TTS	Time to service
UMTS	Universal mobile telecommunications system
WCDMA	Wideband code-division multiple access

The pressure for change

Although mobile phones have become standardized mass-market products, the network infrastructures that support mobile phone users are generally still built up on the traditional telecommunications infrastructure supply model. According to this model, network operators specify in detail at the node level what they require, and a supplier, such as Ericsson, makes custom adaptations. This process has delivered what the industry wanted. At least, so far. But special orders mean special handling and special manufacturing and testing. Special orders also promote multiple variants. This is a costly and time-consuming approach that also creates capacity problems right through to the supply chain.

For the future, the emphasis is on achieving large volumes and rapid roll-out, with high quality and at low cost. The stakes are high for network operators, with the costs of licenses and of building out new and en-

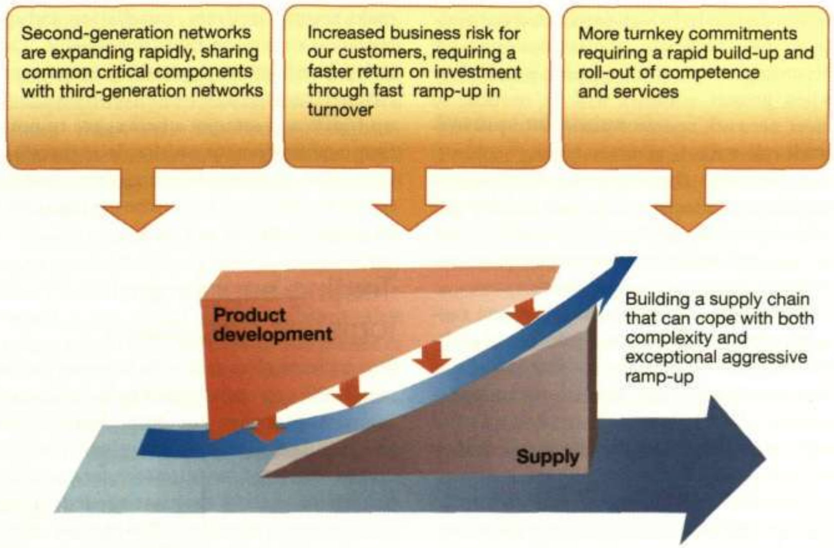


Figure 2
The supply challenge.

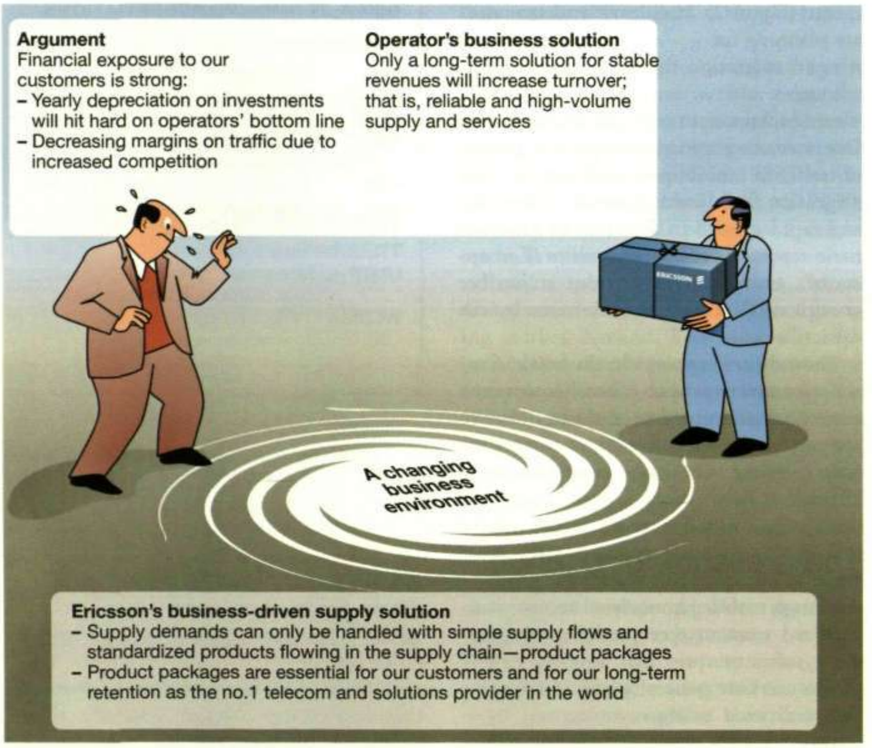


Figure 3
Changing business environments.

hanced network capabilities (Figure 3). Anything that adds unnecessary costs is unacceptable. This focuses attention on certain key areas of the supply chain. Special designs, multiple variants, and inventory all have heavy cost penalties.

Technology, of course, continues to be a vital ingredient. In fact, as we move to third-generation networks, the technologies are becoming even more complex—at a time when greater volumes will be needed, to meet network operator plans for rapid network roll-out.

A further challenge is that third-generation infrastructures will involve the integration of a greater proportion of products from multiple vendors than was the case with second-generation networks.

The process of consolidation among mobile operators has created a number of global operators—another factor that is shaping the supply strategies. Global operators expect standardized and consistent supply and support deals in every market in which they operate.

The challenge for manufacturers is to fit increasingly complex technologies into simpler, faster and less costly supply processes that deliver better quality. The only way to do this is through new supply chain strategies based on greater standardization that enables flexibility.

Large volumes of standard configurations allow a manufacturer to maintain good quality consistently. Standard products also make testing, delivery, installation and integration easier and more dependable. Delivery precision is particularly important. An operator might have to close streets, hold up traffic and arrange a helicopter to position new equipment on site. It is therefore vital that planned delivery commitments can be honored, not just to the day, but often to the precise hour and minute.

Testing supply and logistics processes

Several initiatives taken by Ericsson in the late 1990s, to transform supply chain and logistics processes for second-generation systems, had already highlighted the potential benefits. One initiative, for example, focused on cutting the time needed to fulfill customer orders for GSM mobile infrastructure equipment. This initiative has evolved into the wider TTC Global (time-to-customer) program initiative that operates across the entire Ericsson organization. The focus is on time to service (TTS), since the truly important requirement among network operators is to get equipment quickly installed, integrated and into service, in order to earn revenues.

The primary focus was on taking time and cost out of every stage of the supply chain—from the customer ordering stage through to delivery to the customer site and customer acceptance. This was tackled by introducing significant changes in the ordering process and the supply processes.

One result was that the time needed to build GSM base stations and ship them to the customer site was cut from around 60 to less than 14 days. Cutting the time needed to deliver built-to-order products is only one benefit. Quality and delivery accuracy also improved significantly. Likewise, the whole ordering process has become much easier for customers, thanks to simplified product packages and new e-business processes. This had the important spin-off benefit that customer personnel and Ericsson sales and support personnel spent less time engaged in completing and checking orders.

Greater standardization in the configuration of equipment, such as base stations, meant that there was less scope for errors in ordering, and reduced the number of variants that had to be manufactured. This led directly to better quality and greater production efficiencies.

Ericsson has introduced a logistics concept based on regional flow control centers (FC). Orders flow to the center directly from the customer via Web tools. Completed equipment is then shipped directly from the flow control center to the customer site. This allows traditional warehousing requirements to be reduced or even eliminated, and greatly improves delivery precision (Figure 4).

These supply processes offer clear commercial benefits to network operators. They also represent some big changes in established practices and ways of thinking. One of the main challenges is the acceptance of standard product packages, rather than fully customized products. As Ericsson introduces these supply processes, network operators recognize that a reduction in the "uniqueness" of the equipment they order is a small price to pay for significant improvements in time scales, cost, quality and simplicity.

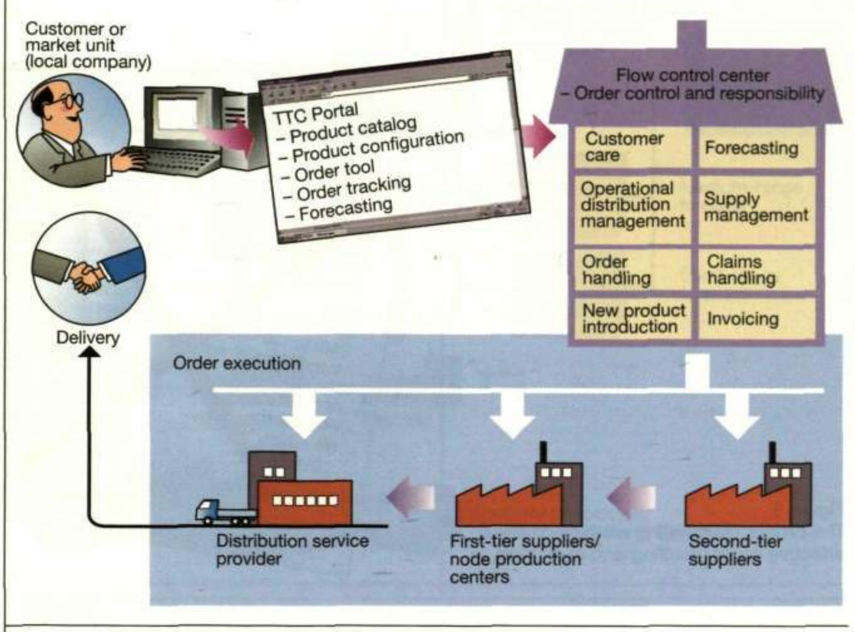


Figure 4
The flow control center concept.

Guiding principles for the supply of equipment

Drawing on experiences of the initial phase of the TTC Global program and other related initiatives, Ericsson has developed a supply and logistics framework that will cope with the combined need to deliver even greater volumes of second- and third-generation products.

Product packages

Product management is about considering the infinite variety of possibilities, and turning them into the actual products that can enter the supply flow. The enabler of this in Ericsson's supply strategies is strict adherence to standard product packages. Ericsson has looked to other industries to see what lessons could be learned, drawing ideas from strategies used in the automotive, personal computing, and consumer goods sectors.

In essence, the idea of product packages is that a customer should only be required to specify a few properties and options in order to get the node that will meet the network and service needs (Figure 5). This represents a major shift away from traditional ordering

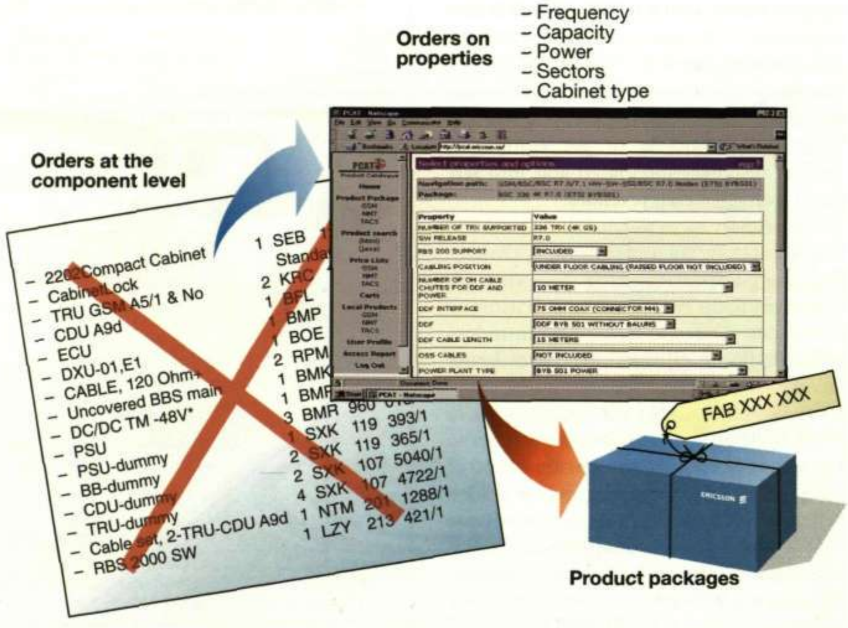


Figure 5
The product packaging enables cost-effective order handling and execution.

practices in the telecommunications infrastructure business, and reduces the operator's scope to specify unique market adaptations. The benefits, however, are considerable:

- because software is the main bearer of functionality, hardware platforms can now be standardized for a mobile network;
- the attraction of standardized product packages is that they promote or facilitate repetitive production, short and fixed lead times, and reduce problems during on-site installation and network integration;
- product packages enable better cost control throughout the chain of value-added activities, which yields configuration flexibility;
- standardized products or modules also enable the supply chain to provide the customer with better and more consistent product quality at lower cost; and
- the ordering procedures are simplified and more resistant to errors.

All new Ericsson product development for mobile network infrastructures is currently managed within this product package concept. Proposed designs are evaluated as much for their impact on supply processes as for price and functions.

The product-package approach has moved the ordering process to a higher level. The emphasis is on the functionality and value of the node in meeting the operator's business plan, rather than on a detailed technical appraisal of what is inside the equipment cabinets.

The product-package approach has strong parallels with, say, ordering an automobile: the product package is the model. Customer choices are simplified (color, engine size, seating, and so on). Additional items that can be specified might include aluminum rims or a stereo.

The product packaging model consists of core product packages—these represent the range of packages that suits all market requirements. On top of the core product packages are the options. There might also be additional materials (special needs to meet, for example, mandatory requirements for a certain market), and finally the installation materials.

Ericsson's goal is that standard product packages should cover over 90% of the market needs. Market adaptations can still be considered, but these will always incur higher cost and longer lead times.

In the case of an Ericsson base station controller, for example, there are now only a few core sizes from which to choose—each size supports a different number of speech channels. There are also a few options, such as interfaces to different generations of base station, choice of overhead or underfloor cabling, choice of 75-ohm or twisted pair connections, and inverter voltage level. Similarly, there are standardized expansion steps and upgrade packages from previous releases.

Site solution

The scope of the product-packaging idea is being broadened. The next step takes the ordering process even higher up in the value chain, to encompass complete site solutions. A site solution contains several node products together with all necessary cables, fixings, power, cooling, services and other items.

The completeness of a site solution can be extremely important in the timely completion of a network project. A missing two-dollar cable, for example, might not be a problem for a base station installation in a city—the installation staff can probably obtain a suitable replacement from a local computer shop. But if the installation is in a very remote location, it could result in several days' delay in getting the network into service. The missing cable could thus result in lost operator revenues running to thousands of dollars.

Forecasting

Forecasting is always a mission-critical activity in planning supply processes. As supply chains get leaner and faster, the role of forecasting will become even more important. Disturbances are more critical when lead times are shorter.

Good forecasting input results in good supply performance. Ericsson is thus developing ways of ensuring the accuracy of the forecasting process. In part, it is a question of education, so that everyone involved in the forecasting process, on the customer side and on Ericsson's side, understands the new context for forecasting. In part, it is an issue of trust on both sides: trust that the figures being presented are accurate, and that the requirements will be met.

Traditionally, forecasting has been more of an art than a science, subject to "adjustments" by people at various stages in the information flow. Part of the supply flow is a Web-based forecasting tool that authorized persons in the local markets use to update forecasts directly. In addition, some customers can currently update their own forecasts in the system. By means of the planning part in the tool, suppliers can immediately get access to updated plans for the products they supply.

The aim is to give everyone in the supply chain the required information at the same time, so that the necessary capacity can be accurately planned. The Internet is a good enabler of this objective (Figure 6).

Ericsson is also investigating a way of bringing greater precision to the process by means of a new software tool which, based on the figures of actual and desired network performance, will automatically calculate a forecast.

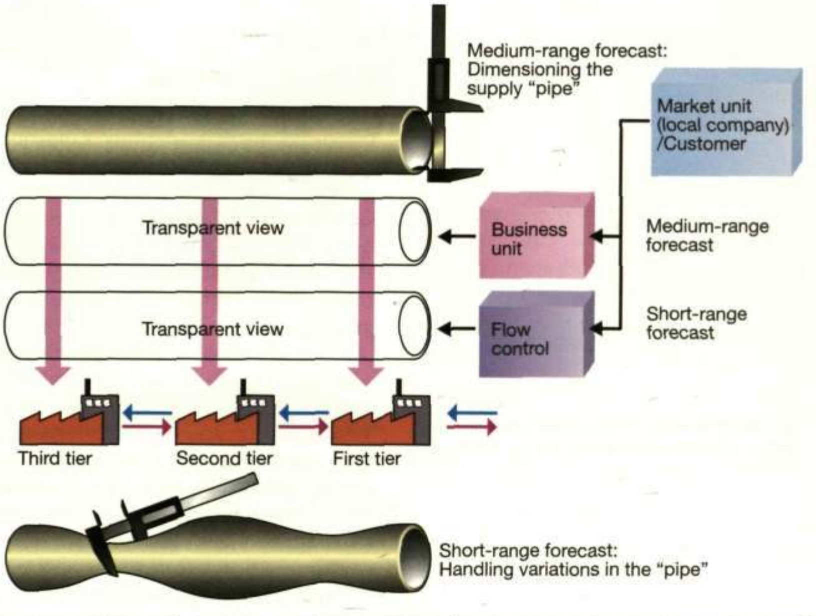


Figure 6
Forecasting information is critical to manage the supply chain.

Direct distribution

Inventory costs money, and as the value of the inventory rises, so do the cost implications. At the same time, product lifecycles are becoming shorter, so there is a need to keep inventory as low as possible to avoid losses through obsolescence.

An important change to supply management is direct distribution. Beginning in 1999, Ericsson started to introduce a technique called *merge in transit*, which can be defined as the coordinated distribution of material from multiple suppliers in order to provide a single, complete, on-time delivery without using any inventory.

Software perspectives

Software delivers the functionality of a mobile communications network. Consequently, because traffic in mobile networks is growing, new types of node element for GPRS and WCDMA/UMTS services are being deployed. Similarly, as the number of services and applications expands, the need to install new software in the network is also growing.

A complication is that many software adaptations have been made to suit the needs of different operators in different countries. This legacy situation is being redressed as the market recognizes that, just as with the product-package concept in hardware, there are enormous benefits in accepting a more standardized, less customized, software strategy.

The traditional approach to a software upgrade is for the network operator to implement the upgrade node by node, with engineers spending time at each site. For an operator with many nodes in the network, this is a heavy, ongoing workload as well as a financial burden. It takes skilled people who might quickly tire of living out of a suitcase and working through the night in order to execute the upgrades.

In 1994, Ericsson began remotely loading software in GSM networks. Today, operators can manage upgrades automatically from a central maintenance center by means of electronic links (as well as satellite links) to the switches in the network. Increasingly, operators are allowing Ericsson to handle these upgrades.

In one case that has been studied, remote upgrading cut the number of man-hours needed by 90% compared to traditional methods. Remote upgrading also lays the foundation for new upgrade strategies. For example, in the past, the emphasis was put on big annual releases of new software, with monthly packages and corrections. With remote software loading, it is possible to install smaller packages at shorter intervals and with less disturbance to the network. Additional benefits are

- fewer human errors;
- operators can more easily retain the expertise of software specialists who no longer have to travel extensively to install software in individual nodes; and
- operators can tap into Ericsson's resources, if necessary, by giving Ericsson online access to individual nodes in the network.

A similar approach has also been adopted to the flow of software updates to customer networks. The automated update-deployment concept—which is based on the operations support system (OSS), Ericsson's network management system in use by the majority of Ericsson's mobile operator customers—facilitates a very aggressive roll-out of updates. The first mobile operator to apply this approach has reported that it could reduce its manpower resources for update handling by more than 80%.

As part of these new software supply processes, Ericsson is moving toward the licensing of software on a right-to-use basis. This business logic for mobile infrastructures fits well in an environment in which operators need to remain flexible and move quickly with their services. It also plays to the strengths of an industry that is moving toward high-functionality software running on non-proprietary hardware platforms.

E-business

Web-enabled ways of working are another integral part of the supply chain management program. The Internet is the main carrier of information along the supply chain, giving everyone the same information at the same time. It is also used as the central communication channel between Ericsson and customers, not only for product ordering, but also as a shared platform for information that can be used by Ericsson and customer personnel.

When customers use the "fast-track" supply processes based on standard product packages, they can place orders via a Web portal directly to the flow control center. Customers can check the status of the order, tracking it all the way to site via a global tracking system that is used by all main global distribution service providers.

Ericsson's first customer to switch to direct ordering of standard product packages over a Web portal did so in 1999.

The prime advantage of putting the product catalog on the Web is that it eliminates any handovers of paperwork or information between the customer and the flow control center. It also prevents customers from ordering non-standard product packages, since only standard packages are included on the Web portal.

For operators with a global business, Web ordering is useful for reinforcing a consistent ordering strategy across the entire organization.

Once the standard product information is on the Web, customers can easily look up relevant technical information on products and related subjects, such as training.

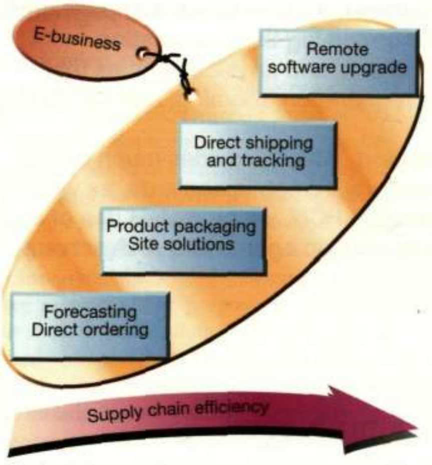


Figure 7
Enhancing business competitiveness by increasing efficiency in the supply chain.

Conclusion

The supply and logistics concepts presented in this article are part of a larger trend toward new ways of working between suppliers—in this case, Ericsson—and their customers.

Many operators are struggling to cope with shortages of skilled and experienced people. Consequently, many of them are reevaluating the role of their procurement functions. As a result, the language of procurement is changing. Instead of discussing bits and pieces, the participants

- want to discuss more strategic business issues, such as coverage, capacity and consultancy; and
- are seeking solutions that support their business objectives.

Operators are increasingly willing to give the technology supplier access to the network. That way, for key activities, such as software upgrades, updates, and planning for the expansion of infrastructure, the operator can benefit from the supplier's know-how and resources.

Obviously, these changes in the supply flow must be planned and implemented jointly by the operator and the supplier, because there are many related implications to be considered. In some cases, the supply processes will also require operators to change their own internal processes—for example, how should an e-procurement process be integrated into the operator's procurement and computer-support systems? However, the overall trend is clear: with the advent of third-generation products, the supplier-customer relationship is moving into new territory, putting new demands on the supply chain as a business-critical function.

The cdma2000 packet core network

Tim Murphy

The packet core network, which is a network architecture being promoted by TIA as the packet-data standard for upcoming third-generation cdma2000 networks, is a collection of logical and physical entities that provide IP-centric packet-data-based registration, roaming, and forwarding services for mobile nodes.

The author describes IP mobility support, which provides much of the basic functionality for proper IP connectivity in mobile contexts, as well as Ericsson's implementation in the cdma2000 PCN, which makes use of a packet data service node, home agents, foreign agents and AAA servers (datacom AAA and telecom AAA).

Introduction

The first commercial wireline telephone switchboard, an ungainly manual contraption, began operation in New Haven, Connecticut on January 28, 1878. It served twenty-one phones on eight separate lines, so the majority of users had to make do with what was referred to as a "party line." By 1903, the estimated number of independent (consumer/household) telephones in place in the US had risen to approximately two million.

Personal transmission of analog audio data was rapidly becoming a ubiquitous and indispensable feature of the technology landscape. Jumping ahead to the 1970s, the number of telephones installed, both household and business, had risen to well over 100 million plus all the commensurate infrastructure. With tongue in cheek, one might say that our reliance on access to a dial tone is now only slightly less than that on access to oxygen.

The first North American trial deployment of a wireless cellular phone system took place in Chicago, Illinois, in July 1978. Illinois Bell and AT&T rolled out a ten-cell advanced mobile phone service (AMPS) system that covered 21,000 square miles. A market trial with paying customers began on December 20, 1978. The first truly commercial service, after the dust of the Bell breakup had settled, was once again fielded in Chicago by Ameritech on October 12, 1983. By 1985, there were approximately 204,000 cellular subscribers. Within three years, this base had grown to 1,600,000 (incidentally, 1988 also saw the publication of TIA's IS-41 specification). In 1990, driven by capacity concerns, the US cellular network headed down the digital path with the introduction of IS-54B, or digital AMPS (D-AMPS), a TDMA-based dual-mode cellular standard. Three years later, the subscriber count had swelled to more than 13 million. By 1999, the combined North

BOX A, TERMS AND ABBREVIATIONS

3GPP/3GPP2	Third-generation Partnership Project
AAA	Authorization, authentication and accounting
ABR	Available bit rate
ADSL	Asymmetrical digital subscriber line
AMPS	Advanced mobile phone service
ANSI	American National Standards Institute
ATM	Asynchronous transfer mode
BSC	Base station controller
BSD	Berkeley Software Distribution (UNIX)
CBR	Constant bit rate
CDMA	Code-division multiple access
CHAP	Challenge-handshake authentication protocol
COA	Care-of address
CORBA	Common object request broker architecture
CPE	Customer premises equipment
CTIA	Cellular Telecommunications Industry Association
D-AMPS	Digital AMPS
FA	Foreign agent
GRE	Generic routing encapsulation
GSM	Global system for mobile communication
HA	Home agent
HDLC	High-level data link communication
HLR	Home location register
HTTP	Hypertext transfer protocol
IANA	Internet Assigned Numbers Authority
ICQ	"I seek you"
IETF	Internet Engineering Task Force
IP	Internet protocol
ISP	Internet service provider
LCP	Link control protocol
LDAP	Lightweight directory access protocol
MAAE	Mobility agent advertisement extension
MAP	Mobile application part
MIP	Mobile IP
MIPS	Million instructions per second
MN	Mobile node
NEBS	Network equipment building standards
OA&M	Operation, administration and maintenance
PCF	Packet control function
PCI	Peripheral component interconnect
PCN	Packet core network
PDSN	Packet data service node
PPP	Point-to-point protocol
RADIUS	Remote authentication dial-in user server/service
RBS	Radio base station
RFC	Request for comments
RP	Radio packet
RRC	Radio resource control
RTT	Radio transmission technology
SDRAM	Synchronous dynamic random access memory
TCP	Transmission control protocol
TIA	Telecommunications Industry Association
TPS	Transactions per second
TSP	The server platform
UBR	Unspecified bit rate
UDR	Usage data record
UTP	Unshielded twisted pair
VBR	Variable bit rate
VPN	Virtual private network
WWW	World Wide Web

American subscriber base of the remaining analog AMPS systems and digital TDMA/CDMA systems was (according to figures released by CTIA in April 2000) in the neighborhood of 86 million. Indisputably the phone, either fixed or mobile, and its services have become a necessary fixture in everyday life.

Roughly contemporaneous with the development of wireless phone service, the growth of the "capital-I" Internet proceeded apace. In 1983, 4.2 BSD UNIX with IP support appeared as the conceptual ancestor to the PC—the workstation. By 1987, there are approximately 10,000 interconnected hosts. This number jumped to 100,000 by 1989. The years 1991-1993 were also very big:

- Tim Berners-Lee "created" the World Wide Web (WWW) by introducing the hypertext transfer protocol (HTTP) for simple, atomic, client-server data requests;
- Mosaic offered an integrated "window" into the Internet (the distinction between the Internet and the WWW began to blur, at least in the popular mind) and overall data traffic grew by 341,634% (!) due to Mosaic-fueled browsing.

As of July 2000, there were over 93 million hosts and over 1 billion indexed Web pages, offering search, commerce, specialty communications, such as ICQ ("I seek you"), and purely informational services. Daily reliance on these services and multimedia data communication features made possible by connection to the Internet, now arguably rivals that on fixed/mobile phone services. However, packet-data network access, as exemplified by the Internet, is still "land-locked." Most end-users still connect their home systems to the Internet over analog phone lines via modem, or (increasingly) via newer, higher-speed purely digital technologies, such as ADSL or a cable modem. Access to mobile packet-data networks still suffers from two limitations:

- low data rates, which make accessing multimedia content and services an exercise in patience; and
- no support for true mobile packet-data service.

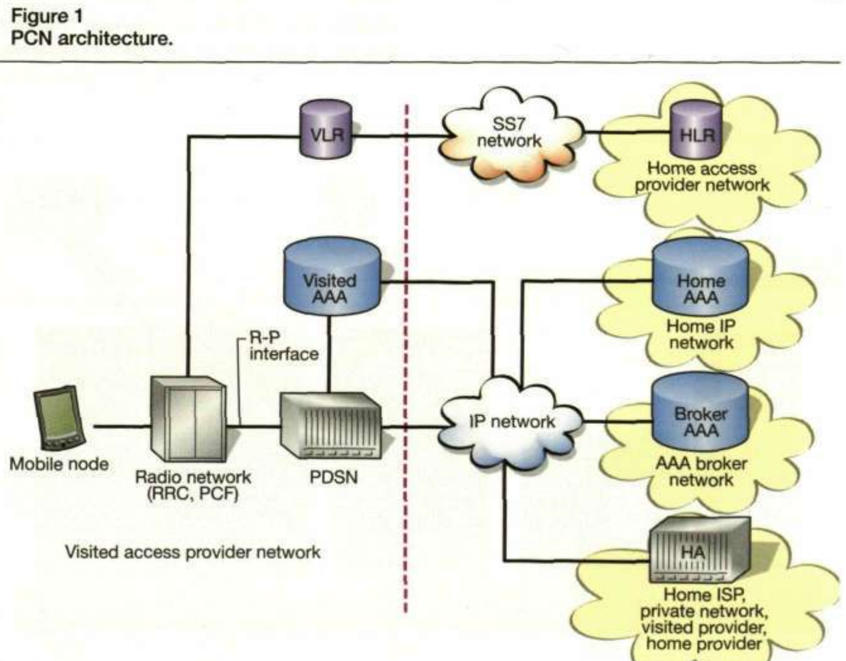
Ericsson's packet core network (PCN) solution, coupled with the higher data rates provided by their cdma2000 radio base station, heralds the convergence of the mobile circuit-switched and existing packet-switched worlds, to provide seamless voice and data service.

Background

The wireless evolution to third-generation technologies and services reflects a convergence of the circuit-switched world of "traditional" telephony and the packet-switched world of contemporary computer networks. Hardware and software which belong to packet-data networks and which support IP-centric protocols must be cleanly integrated into third-generation mobile core networks. The proposed architectures for mobile packet-data services are predicated on existing IP-related standards, which—in light of the unique requirements of mobile nodes (which constantly and sometimes rapidly) change their point of connection to a packet-data network—must be extended or completely developed "from scratch" to permit proper IP connectivity and operations.

As implied by its appellation, the packet core network is a collection of logical and physical entities that provide IP-centric packet-data-based registration, roaming, and forwarding services for mobile nodes. At a high level, the packet core network is analogous in operations and entities to an ANSI-41 or GSM mobile application part-based (MAP) mobile core network. The PCN is a network architecture being pro-

Figure 1
PCN architecture.



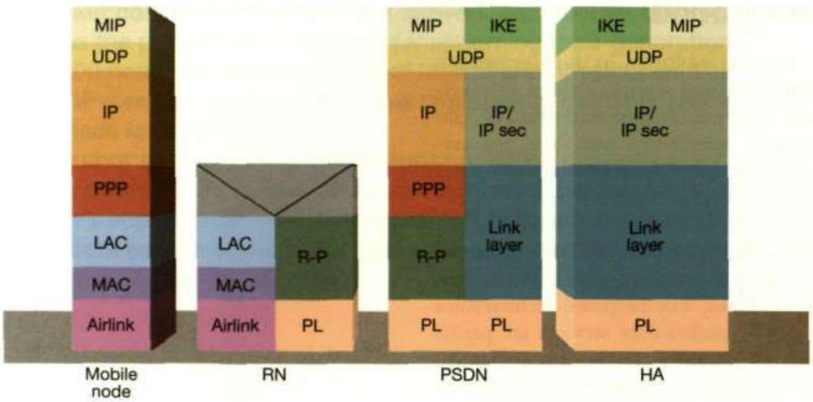
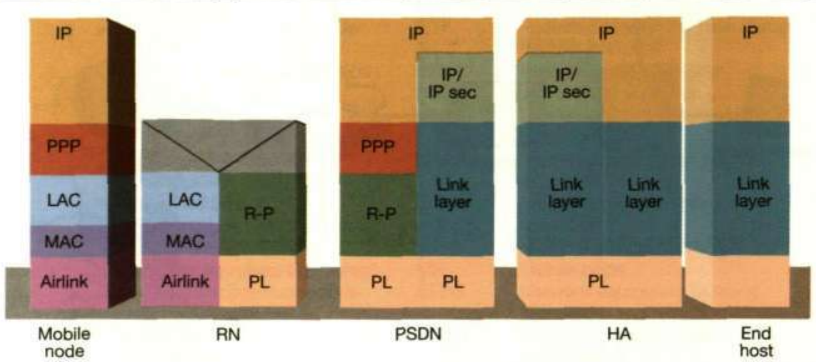


Figure 2
Protocol reference model for mobile IP control and IKE.

moted by group TR45 of the Telecommunications Industry Association as the packet-data standard for upcoming third-generation cdma2000 networks. The higher bit rates available from the 1xRTT CDMA air interface (up to 144 kbit/s of user data when radio configuration 3 rate sets are available) will permit the use of new wireless multimedia, voice and data applications that were formerly impractical with AMPS or cdmaOne IS-95A technology. In an effort to avoid reinventing the wheel, TIA relied on several existing IETF RFCs in crafting IS-835, the wireless IP network standard. The key RFC, from which was derived most of the basic functionality of the nodes described in IS-835, is RFC 2002, *IP mobility support*.

Figure 3
Protocol reference model for mobile IP user data.



IP mobility

IP mobility support describes the framework of procedures, messages, and message formats that enables a node to change its point of attachment to a network without requiring alteration to its IP address, which would otherwise disrupt layer 3 and higher operations. Peer nodes can respond to the sending node indicated by the IP source address of incoming packets regardless of the IP prefix of the network through which the sending node obtained connectivity. To do this, RFC 2002

- introduces a key distinction between the home address and the care-of address (COA);
- defines three entities—the mobile node (MN), foreign agent (FA), and home agent (HA); and
- delineates the exchanges between the MN, FA and HA that facilitate seamless roaming in layers 3 and above. RFC 2002 also augments the functionality of an existing standard, RFC 1256 ICMP, *router discovery messages*, and extends it to define a mobility agent advertisement extension (MAAE) that enables nodes to advertise their availability to provide foreign agent services in a manner similar to RFC 1256 provisions for the advertisement of routers.

The home address, which is the IP address assigned to a mobile node by its controlling provider, is assigned in much the same way as some Internet service providers (ISP) give a subscriber a fixed Class C address. This address, which can be assigned dynamically, is the source address value in the IP header of outgoing packets. Because the network prefix is the same as that of the network assigned to the controlling provider, this address will probably be topologically incorrect when the mobile node roams into a new network. The care-of address functions as an indirect pointer to the mobile node. It represents the topologically correct and reachable IP address that corresponds to the mobile node's current network attachment, so that the home agent can tunnel packets to it. The care-of address can be located

- in the foreign agent—this is more efficient, since only one address is required to support multiple mobile nodes; or
- in the mobile node itself—this obviates

the need for a foreign agent. In this case, the care-of address is referred to as a co-located COA.

A mobile node is basically any user device which features a TCP/IP stack and supports the added functionality that allows it to register its new location as it roams from network to network. The foreign agent is the critical entity in a (visited) network, making mobility services available to a roaming mobile node. It is also one endpoint of the HA-FA tunnel that is created when a mobile node successfully registers with its home agent for mobile IP (MIP) services. This might be an IP-IP, generic routing encapsulation (GRE), or minimal encapsulation tunnel. The foreign agent advertises its availability for service when a mobile node roams into its network and accepts, processes and forwards a mobile node's registration request to the home agent. The foreign agent also accepts, processes and forwards registration replies from the home agent. The home agent, which is the other endpoint of the FA-HA tunnel, is responsible for attracting traffic destined for the mobile node (based on network prefix) and for tunneling it to the care-of address associated with a given mobile node (presumably an interface on a foreign agent) for further delivery to the mobile node.

The position of the home agent in attracting and tunneling traffic to the mobile node gives rise to inefficiencies in the routing of traffic. If the roaming mobile node and a peer node are geographically very close to each other (in terms of router hops), possibly even on different subnets homed on the same node, packets from the peer to the mobile node must still travel through the mobile node, which may be many additional hops away. Referred to as triangle routing, this is one of the issues in the MIP standard being addressed by a number of IETF drafts collected under the umbrella term of route optimization (Box B and Figure 4).

IOS 4.0 and TIA IS-835

IS-835—TIA's wireless IP network standard—lays out the detailed architecture and procedures of the packet core network entities. The standard, which is intended as an implementation document, clearly specifies distinct nodes and their relative areas of responsibility for handling packet data service delivery and mobility management. The interoperability standard (IOS) 4.0 defines

- the radio packet (RP) interface, which is an open interface between the base station controller (BSC) and packet control function (PCF). A logical division was made, separating the BSC into the radio resource control (RRC) and packet control function (PCF, the "interface" to the PCN); and
- the packet data service node (PDSN), which is a key feature of Ericsson's PCN solution.

As stated above, IS-835 uses existing RFCs as a starting point, so I will focus on four important additions to the general MIP framework described in the document:

- the authorization, authentication and accounting (AAA) server;
- the packet-data service node;
- packet-data-related RADIUS attributes; and
- node-specific mobility-management procedures.

The radio packet interface defines two logical channels: *A11* for signaling, and *A10* for data. Signaling is based on MIP messages (*registration request* and *registration reply*) with two additions: *registration update* and *registration acknowledge*. Data from the mobile node is encapsulated in generic routing encapsulation and tunneled from the packet control function within the base station controller to the PDSN, where it is decapsulated and further processed.

The AAA server, which is based on remote authorization dial-in service (RADIUS), is analogous to the home location register (HLR) of the voice core network. It contains subscriber packet-data-provisioning information and is used to authenticate and de-

BOX B, TRIANGLE ROUTING

- The mobile node sends data to its peer node through normal means, using the foreign agent as its default router.
- The peer node sends data to the mobile node indirectly—the home agent intercepts packets destined for the mobile node.
- The home agent encapsulates and tunnels packets to the mobile node via the foreign agent's care-of address. This is inefficient, however, because in theory, a peer node adjacent to the mobile node, either geographically or by network segment, will still have its packets directed through the home agent, which might be located many hops away (see Figure 4).

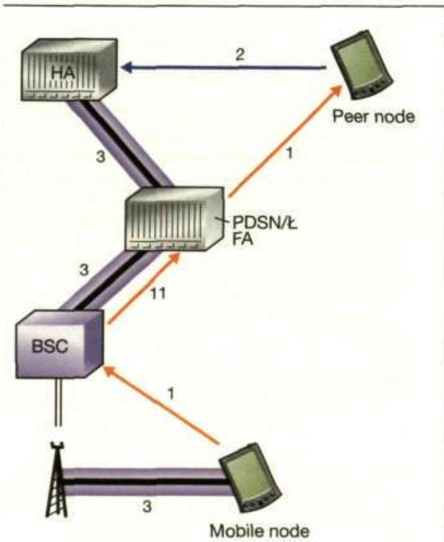


Figure 4
Triangle routing

termine the parameters of a subscriber's packet-data session. It is also a store-and-forward point for accounting data in the form of usage data records (UDR) generated by the PDSN.

The PDSN is the mobility anchor point between the mobile node and PCN. The mobile node point-to-point protocol (PPP) data link terminates at the PDSN, and the FA logical entity resides in this same node. The PDSN

- operates as a RADIUS client toward an AAA server;
- is responsible for aggregating accounting information from the radio resource domain via updates sent from the PCF and for combining this data with its own packet traffic information to form complete accounting records; and
- vets subscriber credentials during PPP CHAP activity where the subscriber's information is verified to allow completion of PPP link establishment.

RADIUS attribute extensions have been defined by TIA in IS-835 to enable information on cdma2000 packet data operations to be exchanged within the RADIUS protocol framework. The extensions are of type "26"—that is, they are vendor-specific (3GPP2). The 3GPP2 vendor ID used with them is 5535, as defined by the Internet Assigned Numbers Authority (IANA). The attributes can be broken down further into accounting and service-related attributes, such as session status, session activity, security status, and differentiated services class option. These attributes are exchanged between the PDSN and AAA in, for example, access request, access accept, and accounting request RADIUS messages.

Mobility management defines two roaming scenarios and the procedures to accommodate them. In the PCF-to-PCF handoff scenario (intra-PDSN), the mobile node moves from one base station controller to another, and both BSCs are connected to the same PDSN. This type of handoff might not require the renegotiation of the PPP session between the mobile node and PDSN, since the PDSN can re-associate the PPP state of the mobile node with the new base station controller. In the PDSN-to-PDSN handoff scenario (inter-PDSN), the mobile node (presumably) roams into a new network with a different PDSN. In this case, the mobile node is required to establish a new PPP link to the new PDSN and then perform MIP registration again.

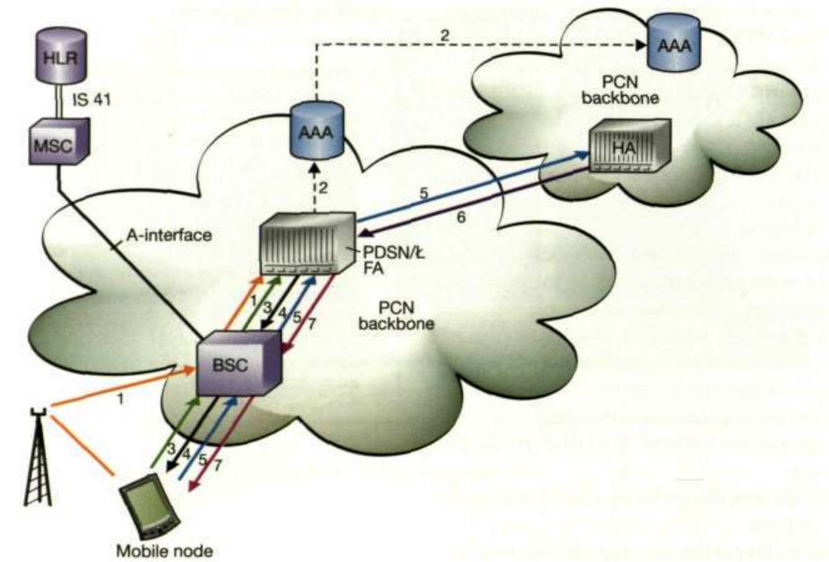


Figure 5
Mobile IP registration.

BOX C, DETAILED MIP REGISTRATION IN ERICSSON'S PCN

- The mobile node (MN) roams into a new (visited) network and sends an origination request to the base station, specifying packet-data service option 33.
- The packet control function (PCF) determines that there is no A10 connection associated with the mobile node and initiates A11 signaling to the packet data service node (PDSN) to set up A10 data connection.
- The mobile node negotiates PPP LCP with the PDSN to establish the data link.
- LCP link parameter negotiation concludes; the CHAP phase begins.
- The mobile node sends CHAP data (CHAP ID, CHAP secret, etc.) to the PDSN, which
  - creates a RADIUS access request, assigning CHAP data to appropriate RADIUS attributes; and
  - sends the attributes to the local AAA server, which proxies the request to the mobile node's home AAA.
- The home AAA returns a positive access accept message through the visited AAA to the PDSN.
- The PPP link is established.
- The mobile node sends agent solicitation to the PDSN/FA looking for an available foreign agent (FA).
- A foreign agent responds with a mobility agent advertisement, listing available foreign agents.
- The mobile node creates a MIP registration request, specifying a willing foreign agent as its care-of address (COA) and forwards the request through the foreign agent to the home agent (HA).
- The foreign agent processes the request and forwards it to the home agent.
- The home agent processes the request and responds with a positive registration reply. It also updates its binding tables with the mobile node's new care-of address association and its remaining lifetime.
- The foreign agent processes the reply, updates its visitor list with a new entry for the mobile node, and forwards the reply to the mobile node.
- A tunnel is established between the foreign agent and the home agent.
- The mobile node can now send packet data using the foreign agent as its default router.
- The home agent can now tunnel traffic sent to the mobile node's home address to the foreign agent, which decapsulates the traffic and transfers it to the mobile node.

An addendum to IS-835, IS-835-A-1, which is currently under review, attempts to clarify "gray areas" in the initial document, such as handoff optimization, security indications for optimized key exchanges, and the removal of data entities (these have been underutilized in emerging implementations of the standard).

Ericsson's packet core network

Ericsson has capitalized on decades of telecom and datacom expertise to deliver a fully open standards-compliant third-generation packet network architecture that incorporates the most recent advances in mobile packet-data transmission technology. The PCN product roadmap includes PDSN, HA and AAA offerings that span the range of subscriber or traffic loads, from equipment that handles from 5,000 to 7,000 simultaneous sessions to platforms that handle up to 400,000 simultaneous sessions typically deployed in a central office.

All Ericsson solutions offer carrier-class network equipment building standards (NEBS) level 3-compliant access platforms for the PDSN/foreign agent and home agent nodes. The AAA server is also NEBS-compliant for use in a carrier operations center.

PDSN

The first router in the Ericsson PCN family to implement PDSN functionality is the AXC, which has been optimized for small-scale deployment. In PDSN trim, up to 50,000 simultaneous PPP sessions are supported. The AXC provides a PCI-derived 1 Gbit/s packet bus coupled with a MIPS CPU and up to 512 MB of SDRAM, split between packet memory (for packet forwarding) and local memory. Up to 64 MB of flash memory is used for operating system activity, such as maintaining IP routing tables, physical interface tables, and so on. Dual control cards provide support for redundant failover, so that the failure of one control card will result in the backup card taking over—a necessary configuration in high-availability carrier environments. This configuration features 10/10BaseT unshielded twisted pair (UTP) redundant ports for network connectivity and uses asynchronous transfer mode (ATM) interface cards to the base station controller. These high-performance interface boards, which are based on the IDT R4700 processor, support OC-3 and STM-1 155 Mbit/s over single- or multi-mode fiber or CAT-5 STP. Standard ATM service class configurations, such as constant bit rate (CBR), variable bit rate (VBR), available bit rate (ABR) and unspecified bit rate (UBR) and their associated metrics are supported, as well as support for RFC 1483, to enable multiprotocol encapsulation over ATM.

The medium- to high-capacity PDSN platform is the AXI, in seven- or fifteen-slot configurations, supporting up to 200,000 simultaneous PPP sessions.

To the base station controller, each OC-3/STM-1 ATM interface card features four single- or multi-mode physical 155 Mbit/s fiber interfaces. The ATM forwarding engine supports both PPP (RFC 1619, *PPP over SONET*) and HDLC (RFC 1662, *PPP in HDLC-like framing*) encapsulation. All major components (the route processor module, switch fabric modules, fans, power supplies, and so on) are one-to-one (1:1) redundant and hot-swappable for zero-downtime operation.

Home agent

The home agent is based on the same hardware platforms as the PDSN. It is also pro-

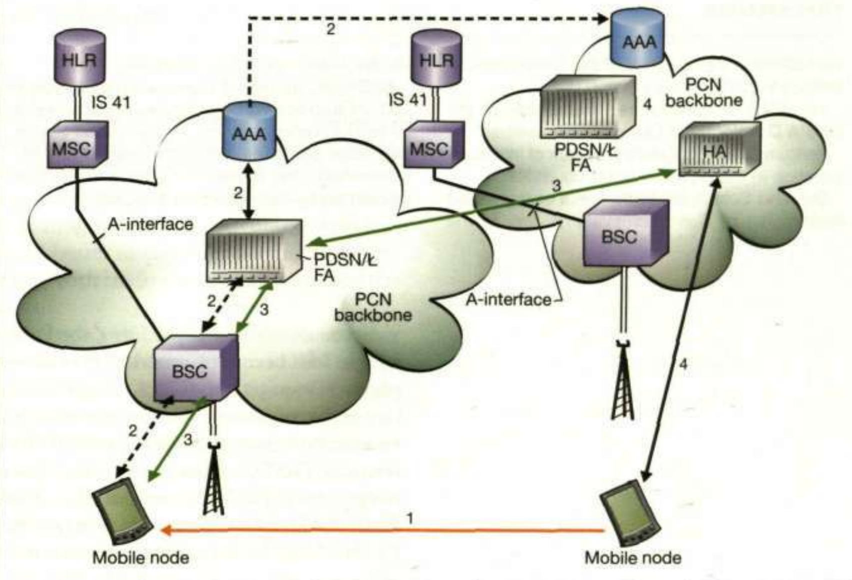


Figure 6
Mobile IP handoff.

BOX D, MOBILE IP HANDOFF

- The mobile node roams into a new access provider.
- The mobile node engages in new PPP negotiation with a new PDSN. As before, the PDSN determines the subscriber attributes during the CHAP authentication phase of PPP via an AAA server.
- The mobile node obtains the care-of address (COA) from the foreign agent and registers with the home agent, updating its mobility binding with the new COA (in Figure 6, the mobile node is shown roaming away from its home network). The foreign agent updates its visitor list.
- The previous PPP link is torn down and the foreign agent visitor list entry is deleted after registration lifetime expires.

vided on an additional platform, the Fraser, which has been optimized for customer premises equipment (CPE) solutions that target the corporate level (that is, virtual private networks) and supports up to 7,000 PPP sessions. The Fraser features two fast Ethernet ports, an IDT R4700 central CPU, 8 MB Flash RAM for configuration storage, and 16 MB DRAM dedicated packet memory.

AAA server

A unique design innovation in the Ericsson PCN places all node configuration information in the AAA server. Thus only the AAA server is provisioned—subscriber and node data is accessed by the PDSN or HA on an as-needed basis. This avoids the redundancy of duplicate provisioning, a situation which arises because IS-835 specifies the data that is required by the nodes for MIP operations but not how the data is to be disseminated. An additional benefit of this approach is the short deployment time for nodes in an Ericsson PCN. If a node currently operating as a PDSN must be rotated out or switched to home agent duty, a reload of the operating system (incorporating appropriate node-specific PCN features) is the only operator activity needed to completely re-provision the node. The remaining provisioning data is retrieved from the AAA server.

Datacom AAA

The AAA applications feature full RADIUS support with cdma2000-specific attributes (RFC 2138, 2139) as well as AAA extensions for MIP authorization and accounting support. The underlying database systems, which support widely accepted open data

TRADEMARKS

BOX E, STANDARDS AND RFC'S

TIA standards

TSB115	Wireless IP Network Architecture
IS-707-A-1.10	Data Service Options for Spread Spectrum Systems: Radio Link Protocol Type 3
IS-707-A-1.11	Data Service Options for Spread Spectrum Systems: cdma2000 High Speed Packet Data Service Option 34
IS-707-A-1.12	Data Service Options for Spread Spectrum Systems: cdma2000 High Speed Packet Data Service Option 33
IS-835	Wireless IP Network Standard
IS-2000-1	Introduction to cdma2000 Standards for Spread Spectrum Systems
IS-2000-2	Physical Layer Standard for cdma2000 Spread Spectrum Systems
IS-2000-3	Medium Access Control Standard for cdma2000 Spread Spectrum Systems
IS-2000-4	Link Access Control Standard for cdma2000 Spread Spectrum Systems
IS-2000-5	Upper Layer (Layer 3) Standard for cdma2000 Spread Spectrum Systems
IS-2000-6	Analog Standard for cdma2000 Spread Spectrum Systems
IS-2001	Interoperability Specification for cdma2000

RFCs

RFC 1256	ICMP router discover
RFC 1661	PPP
RFC 1662	PPP HDLC framing
RFC 1701	Generic routing encapsulation (GRE)
RFC 1702	GRE in IP
RFC 2002	Mobile IP
RFC 2003	IP in IP encapsulation
RFC 2004	Minimum encapsulation
RFC 2138	Remote authentication dial-in user service (RADIUS)
RFC 2139	RADIUS accounting
RFC 2344	Reverse tunneling
RFC 2401	IPSec security architecture
RFC 2402	Authentication aeader (AH)
RFC 2406	Encapsulating security payload (ESP)
RFC 2661	Layer 2 tunneling protocol (L2TP)

processing standards, such as the lightweight directory access protocol (LDAP), are fully distributed and replicated (logically, and if so configured, geographically) for zero downtime, should a particular node or office fail. Very high transactions-per-second (TPS) levels can be sustained, which reflects the large subscriber base and high activity rate currently seen in the circuit-switched world.

Telecom AAA

The telecom AAA node is based on Ericsson's TSP, a multi-CPU and scalable applications platform. The distributed real-time operating system, runs on Pentium boards, each with 512 MB dedicated RAM, and supports open data transfer standards, such as the common object request broker architecture (CORBA). Dual, dedicated UltraSparc boards run operation, administration and maintenance (OA&M) applications, providing access to external databases. The AAA and HLR platforms are derived from the same architecture, providing a single point for packet and telephony subscriber provisioning.

Conclusion

Two forces have driven the development of the packet core network: the evolving consumer demand for convenient access to multimedia data services, as exemplified by the World Wide Web, and the approaching technological convergence of circuit-switching and packet-switching—a packet-based bearer medium supporting and coexisting with traditional circuit-switched services.

Ericsson's PCN solution introduces three nodes—the home agent, PDSN/foreign agent, and AAA—which support high-speed wireless packet-data operations and provide zero-downtime operational characteristics for demanding telecommunications environments. The nodes are evolved from mature hardware architectures with proven track records, such as the AXC, AXI and TSP platforms. With Ericsson's full support of the open standards radio-packet interface, the PCN can be rapidly and smoothly integrated into any cdma2000 core network.

REFERENCES

1 Perkins, Charles: Mobile IP Design Principles and Practices. Addison Wesley 1998
2 Harte, Lawrence: CDMA IS-95 for Cellular and PCS. McGraw-Hill 1999
3 Nilsson, M.: Third-generation radio access standards. Ericsson Review Vol. 76(1999):3, pp 110-112
4 Farley, Tom: Private Line's Telephone History. Independent 2000

Previous issues

No. 1, 2001

Wireless-Trench technology for portable wireless applications
AXE 810—The evolution continues
Open architecture in the core of AXE
Cable modems—Broadband highway to the home
The R380s—The first smartphone from the Ericsson–Symbian partnership

No. 4, 2000

Experiences of the live WCDMA network in Stockholm, Sweden
Media Gateway for mobile networks
Ericsson's IP-based BSS and radio network server
Control servers in the core network
Ronja—A Java application platform
Wireless *ad hoc* networking—The art of networking without a network
Ericsson's WebOnAir Information Server

No. 3, 2000

ENGINE network server
ENGINE Access Ramp—The next-generation access architecture
Speech recognition technology for mobile phones
GPRS support nodes
Family of RBS 3000 products for WCDMA systems
Ericsson's network-based IP-VPN solutions
Fixed wireless access for broadband services

No. 2, 2000

Evolving from cdmaOne to third-generation systems
Third-generation TDMA
Business solutions for mobile e-commerce
The RBS 2206—A flexible ticket to third-generation wireless systems
Communications security in an all-IP world
HIPERLAN type 2 for broadband wireless communication

No. 1, 2000

IPv6—The new generation Internet
WAP—The catalyst of the mobile Internet
The challenges of voice over IP over wireless
Openness in AXE
Management solutions for IP networks

Telefonaktiebolaget L M Ericsson
SE-126 25 Stockholm, Sweden
Phone: +46 8 7190000
Fax: +46 8 6812710

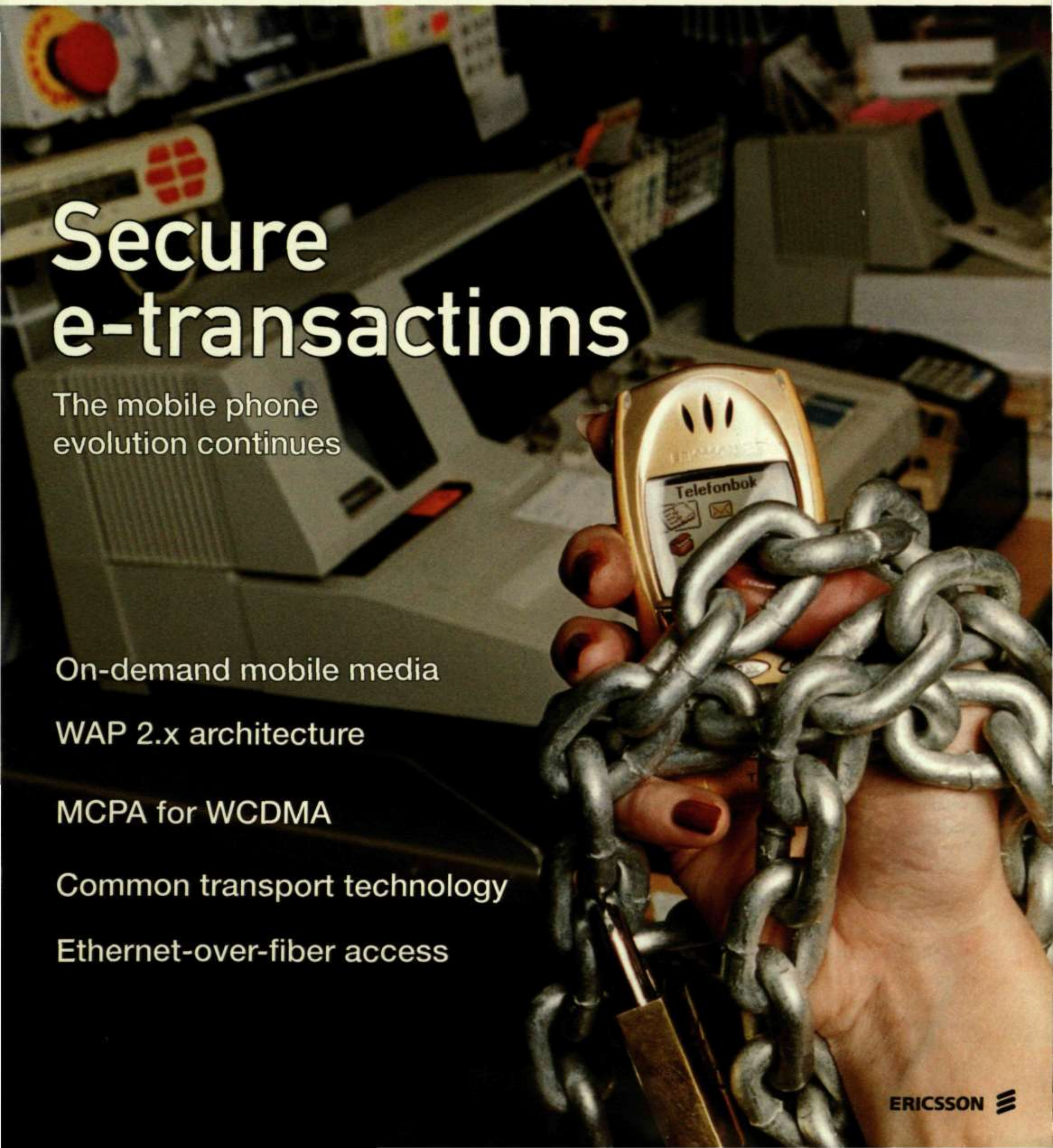
ISSN 0014-0171
Ljungföretagen, Örebro 2001

Ericsson

4/2001

REVIEW

THE TELECOMMUNICATIONS TECHNOLOGY JOURNAL



Secure e-transactions

The mobile phone evolution continues

On-demand mobile media

WAP 2.x architecture

MCPA for WCDMA

Common transport technology

Ethernet-over-fiber access

ERICSSON

Dear subscriber

As part of Ericsson's ongoing efficiency program, Ericsson Review will henceforth be published in English only (it was previously published in English and Spanish). Moreover, in 2002, we will solely publish three issues of Ericsson Review (publication dates: March, July, and November).

We would also inform you of changes, effective September 1, 2001, affecting the administration of subscriptions and distribution. If you have questions regarding your subscription or want to report a change of address, please contact PressData:

PressData
Box 3263
SE-103 65 Stockholm
Sweden

Phone: +46 8 799 6328
Fax: +46 8 285974
E-mail: ER@pressdata.se

Distinguido suscriptor

Como parte del programa de eficiencia aplicado por Ericsson, en lo sucesivo la Ericsson Review sólo se publicará en inglés (antes aparecía en las versiones inglesa y española). Además, en 2002 editaremos únicamente tres números de la revista (fechas de publicación: marzo, julio y noviembre).

También quisiéramos informarle de los cambios, entrados en vigor el 1° de septiembre de 2001, relativos a la administración de las suscripciones y la distribución. Si desea hacer alguna consulta acerca de su suscripción o quiere comunicar un cambio de dirección, le agradeceremos que se ponga en contacto con PressData:

PressData
Box 3263
SE-103 65 Estocolmo
Suecia

Teléfono: +46 8 799 6328
Fax: +46 8 285974
Correo electrónico: ER@pressdata.se

Cover: The mobile electronic transaction (MeT) initiative, established in April 2000 by Ericsson, Motorola and Nokia, is a generic framework for regulating how business-to-consumer mobile electronic transactions can be handled securely on mobile phones, guaranteeing a consistent user experience independent of device, service and network.

Ericsson
REVIEW

THE TELECOMMUNICATIONS TECHNOLOGY JOURNAL

The purpose of Ericsson Review is to report on the research, development and production achievements made in telecommunications technology at Ericsson. The journal is distributed to readers in more than 130 countries.

Address:
Telefonaktiebolaget LM Ericsson
SE-126 25 Stockholm, Sweden

Phone: +46 8 719 00 00

Fax: +46 8 681 27 10

Internet:
http://www.ericsson.com/review

New subscriptions:
http://www.ericsson.com/review

Address changes and distribution:
Phone: +46 8 799 63 28
Fax: +46 8 28 59 74
ER@pressdata.se
Address: PressData
Box 3263
SE-103 65 Stockholm, Sweden

Publisher: Jan Uddenfeldt

Editor: Eric Peterson

Production manager: Eva Karlstein
eva.karlstein@lme.ericsson.se

Layout: Paues Media, Stockholm

Illustrations: Claes-Göran Andersson, Hara

Printer: Ljungföretagen, Örebro

ISSN: 0014-0171

Volume: 78, 2001

Contents

Secure electronic transactions—The mobile phone evolution continues

Ericsson is convinced that future mobile phones will be equipped with unique digital credentials to be used for a multitude of services—the phone is thus becoming a personal trusted device, not only because of technical advances but also because consumers are increasingly willing to use their phones for non-traditional services, such as transaction-based services. **162**

On-demand mobile media—A rich service experience for mobile users

The introduction of standardized mobile messaging services, such as EMS and MMS, will enrich person-to-person messaging and pave the way for content-push services. On-demand mobile media services will be delivered to users via media streaming and download techniques that enrich mobile browsing and general content access. The building blocks for on-demand mobile services have been standardized in the 3GPP. **168**

WAP 2.x architecture—Features, services and functions

Web services adapted to the wireless environment have been a tremendous success, especially in Japan, where more than 25 million people use i-mode. However, the architecture of i-mode has some limitations. These will be overcome with WAP 2.x. **178**

RF multicarrier amplifier for third-generation systems

The growing market for third-generation systems requires a base station RF power amplifier designed for large-volume production. Ericsson has chosen the feed-forward technique with analog predistortion, which gives excellent spurious emission values and high production yield. A design for efficiency and cooling guarantees the best MTBF in a small-size amplifier. **184**

Ericsson's family of carrier-class technologies

The horizontally layered network sets the stage for new service applications and for the introduction of a common transport technology. Ericsson's way of developing systems and applications is thus undergoing a transformation: the company strategy involves a family of carrier-class technologies on which to base telephony, server and access applications. **190**

Building a reliable, cost-effective and future-proof fiber optical access network

Ethernet-over-fiber is likely to become the preferred solution for extending optical techniques straight out to the wall socket. By offering symmetric transmission of 100 Mbit/s or better, Ethernet-over-fiber access networks will enable the delivery of advanced services. And given the current and emerging Ethernet standards, we can soon expect to see flexible, multipurpose access networks built entirely on Ethernet technology. **196**

Contributors

In this issue

Göran Ahlforn joined Ericsson in 1971 and has held numerous positions within the areas of systems development, product management and product marketing. He was involved in the early development of AXE, and was responsible for AXE systems management between 1991 and 1995. He has also served as director of R&D for Broadband Systems as well as Enterprise Systems. He is currently director of platform evolution at Ericsson's Core Network Development core unit. He holds an M.S. in electronics from the Lund Institute of Technology.
goran.ahlforn@uab.ericsson.se

Per O. Andersson has been engaged in optical research for 20 years in areas ranging from WDM transmission, optical signal quality, optical components, and optical access systems. He has also represented Ericsson in ITU and ETSI, and served as head of the Optical Network Research Laboratory. He is presently director of business development in Ericsson's Transport and Transmission business unit.
per.o.andersson@etx.ericsson.se

Peter Arnby works for Mobile Internet Applications in Kista, Sweden. In March 2000, he became the project manager of EriWAP, the project responsible for standardization in the WAP Forum and W3C. He also currently serves as a member of the specification committee in the WAP Forum and the document secretary of the organization. Peter has also held assignments in other standards bodies, including 3GPP and Parlay.
peter.arnby@eip.ericsson.se

Bo Berglund is an expert in MCPA technology at Ericsson Radio Access. Since joining Ericsson, in 1996, he has been involved in the development of MCPAs for PDC, TDMA and WCDMA systems. He holds an M.S. in electrical engineering from Chalmers Institute of Technology, Göteborg.
bo.berglund@era.ericsson.se

Erik Ekudden is an expert in the field of audio/visual communication systems at Ericsson Research. Since joining Ericsson, in 1993, he has worked with source coding research and standardization for GSM, PDC, TDMA and WCDMA systems. He is currently involved in the research and development of new mobile multimedia services. Erik holds an M.S. in electrical engineering from the Royal Institute of Technology, Stockholm, Sweden
erik.ekudden@era.ericsson.se

Susanna Friis-Hansen serves as strategic product manager for WAP and secure transactions at Sony Ericsson Mobile Communications AB. She is also Sony Ericsson's representative on the MeT board of directors. Susanna holds a masters degree in engineering and business management from the Linköping Institute of Technology, Sweden.
susanna.friis-hansen@ecs.ericsson.se

Ingvar Fröroth is system manager of Ericsson Telecom's Ethernet Access product development unit. He joined Ericsson in 2001, after having spent five years researching new access networks and broadband communication services for the residential sector. Ingvar holds an M.S. in electrical engineering and a licentiate of technology degree in Teleinformatics.
ingvar.froroth@etx.ericsson.se

Johan Hjelm is project manager for applications research in the Ericsson Research Japan IMT-2000 group at Yokosuka Research Park outside Tokyo, Japan. During the past two years, he has been working with convergence issues between the WAP Forum and the W3C. Johan also chairs the W3C CC/PP working group. In 2001, he was appointed senior specialist, Internet information services and protocol design. He is the author of two books, *Designing Wireless Information Services* and *Creating the Semantic Web with RDF*, published by John Wiley and Sons.
johan.hjelm@nrj.ericsson.se

Uwe Horn works at Ericsson Research as a technical advisor and research coordinator for streaming-related research, standardiza-

tion, and product-development projects. In 1998, he joined Ericsson Eurolab, Aachen, Germany, and in 2001, he was appointed senior specialist, mobile streaming. Uwe holds a diploma in computer science from the University of Bonn, and a Ph.D. in electrical engineering from the University of Erlangen-Nuremberg.
uwe.horn@eed.ericsson.se

Mats Melander is employed as technical manager at Ericsson Mobile Platforms in Lund, Sweden. He joined Ericsson, in 1997, to work as a member of the data communication team, developing software solutions for mobile terminals. In 1999, he became technical manager of multimedia protocol implementations. He has also been active in the standardization of multimedia services. Mats holds an M.S. in computer science from Uppsala University, Sweden.
mats.melander@emp.ericsson.se

Stefan Nilsson-Gistvik works at the fiber access department of Ericsson Network Technology. He joined Ericsson, in 1984, as an instructor of data communication, switching in 1987 to fiber optics. He created and developed Ericsson's fiber-optics training program for customers and is the author of Fiber optic theory for communications networks. In 1999, he served as an advisor on fiber and fiber-optic cables to the Swedish ITC-Commission.
stefan.n-g@eca.ericsson.se

Thorsten Nygren joined Ericsson in 1982 as a radio design engineer for defense communications. He currently works with linearized RF power amplifiers for radio base stations. He holds an M.S. in electronic engineering from the Royal Institute of Technology, Stockholm, Sweden.
thorsten.nygren@rsa.ericsson.se

Jan Olin joined Ericsson in 1996 and has held various positions ranging from operation development to product portfolio management. In 1999, he initiated Ericsson Radio Systems' business development activities for on-demand multimedia for wireless systems. Today, he is strategic product manager of Ericsson's on-demand multimedia distribution solution—Interactive Media. He holds a B.S. in electrical engineering from the Royal Institute of Technology in Stockholm and an MBA from the University of Stockholm.
jan.olin@eip.ericsson.se

Karl-Gösta Sahlman, who joined Ericsson in 1993, is an expert in radio implementations. The main focus of his work at Ericsson is on radio receivers, radio transmitters and RF power amplifiers (analog and digital technologies). He holds an MSEE degree from the Chalmers Institute of Technology, Göteborg.
karl-gosta.sahlman@rsa.ericsson.se

Peter Stark joined Ericsson in March 2000 and currently works for the Ericsson Mobile Platforms business unit in Lund, Sweden. Since 1997, he has been a participant in the WAP Forum—two years in the protocols group, and two years in the applications groups. Similarly, since 1999 he has served in the W3C HTML group, and since 2000, in the W3C XForms groups.
peter.stark@ecs.ericsson.se

Bengt Stavenow joined Ericsson Radio Systems in 1983 after completing post-graduate studies in telecommunications systems at the Lund Institute of Technology. In 1994, after holding various system management assignments, he was put in charge of system management and user interface design at Ericsson Mobile Communications. Bengt is currently responsible for software application development of trusted open execution environments for feature phones.
bengt.stavenow@ecs.ericsson.se

Erik Örnulf is currently vice president of Systems Technology Management at Corporate Technology, Telefonaktiebolaget LM Ericsson, a position he has held since 1998. From 1992 to 1998 he was vice president of research at Ericsson Radio Systems. Before this, he was responsible for the systems architecture group for cellular systems. Mr. Örnulf began his career at Ericsson in 1969. He holds an M.S. in engineering from the Lund Institute of Technology.
erik.ornulf@lme.ericsson.se

Secure electronic transactions—The mobile phone evolution continues

Susanna Friis-Hansen and Bengt Stavenow

Ericsson holds firm to the idea that future mobile phones will be equipped with a single digital ID based on an open, globally accepted standard. The phone will be used for a multitude of trust-based services including e-commerce and secure access to e-mail clients, intranets, and even physical access to property and premises.

In April 2000, Ericsson, Motorola and Nokia jointly established the MeT initiative. The mission of the initiative is to define how certain core technologies should be used in mobile phones to enable electronic transactions. To facilitate acceptance of all parties involved, and applicability in all relevant markets, the strategy for MeT has been to base its framework on existing standards, and to design it to be independent of underlying system standards.

The authors give an introduction to secure electronic transactions, explain the role of the handset manufacturer, and describe MeT and the enabling technologies of the initiative.

Introduction to secure electronic transactions

Being small, handy and increasingly utile, the mobile phone is the central or core device of a mobile society. The addition of novel applications and capabilities make it even more personal and trusted, and thus a regular part of everyday life.

Mobile phones are becoming commonplace products. The number of mobile phone subscribers worldwide is expected to reach 1 billion by 2002. And a significant share of these users will soon be equipped with Mobile Internet-enabled terminals. In fact, by 2003, this group is expected to outnumber users of the fixed-line Internet. Consequently, consumers will have access to content and services at any time and from virtually any geographical location.

At present, the main use of mobile communications is still voice-related service, but several indicators show that this is changing. Larger displays, packet-data and third-generation networks, Bluetooth, and recent releases of the wireless application protocol (WAP)—which offers adequate security for a wide spectrum of applications thanks to more expressive markup language elements in WAP 2.0—are combining to give good user experience of data-centric services. In many markets, where users are becoming experienced and are interested in finding new ways of using their mobile phones, the stage is set for a successful roll-out of data-centric transaction-based services, such as payment services and ticketing. The mobile phone is thus becoming a personal trusted device (PTD) that is increasingly capable of handling transaction-based functions in the physical and online worlds.

Transaction-based services have caught the interest of powerful stakeholders who are eyeing expanded business opportunities. Operators, for example, are active in the mobile-transactions arena, and see big opportunities in expanding their businesses to become a partner higher up the value chain.

Typical transaction-based services include supporting infrastructure services throughout the various stages of the service life-cycle, such as the issuing of WAP identity modules (WIM), certificate authority (CA) services, and payment-clearing services.

Banks, which are the traditional strong players in the financial sector, are determined to maintain their position by offering financial services in the mobile environment. Examples of services are stock trading, the transfer of funds between accounts, and payment services. Certain progressive banks are already involved in limited trials, and the financial sector generally expects the market for mobile financial services to take off in 2002. Obviously, trust is a key issue. Traditionally, the financial sector has not trusted anyone but itself. Therefore, single-application security solutions, such as secure electronic transactions (SET) and Europay, VISA and Mastercard (EMV), were developed for use in the fixed Internet environment and for point-of-sale terminals. Unfortunately, even though consumers expect to have global access to services, SET and EMV have thus far failed to

BOX A, ABBREVIATIONS

CA	Certificate authority
CUE	Consistent user experience
EMV	Europay, VISA and Mastercard
J2ME	Java 2 platform, micro-edition
MeT	Mobile electronic transactions
PIN	Personal identification number
PKI	Public key infrastructure
PTD	Personal trusted device
SE	Security element
SET	Secure electronic transactions
SIM	Subscriber identity module
SWIM	WIM on SIM
TLS	Transport layer security
WAP	Wireless application protocol
WIM	WAP identity module
WML	Wireless markup language
wPKI	Wireless PKI
WTLS	Wireless TLS

take into consideration geographical differences and the limitations of the mobile phone and the mobile environment. In general, the financial sector, which is wary of operator-driven activities, has stated that regardless of the technical solution employed, operator independence is an absolute must. It sees itself as the obvious supplier of the security infrastructure needed for financial services and wants to have as little involvement from operators as possible.

Apart from the companies involved in e-commerce-related applications, numerous security companies offer their own security infrastructure (servers and clients) for e-mail, secure access to, say, a corporate intranet, and physical access to buildings.

This is still very much an emerging market fraught with numerous contending technologies and devices, and a lack of widely accepted standards and global solutions. This fragmentation is serious and potentially threatening to market development. Because content and service providers must take into consideration all possible products that might be used to access services, the development process will be slow and tedious. What is even more serious, however, is that the user experience, which is the basis for creating consumer trust, is inconsistent. If consumers are to turn to the capabilities of their mobile phones for making payments, accessing e-mail clients, and for gaming and ticketing services, then the mobile phone must be easier and more convenient to use than credit cards or some other secure-access device that can be carried in the wallet. In addition, consumers must feel that the services can be trusted and that their integrity is being protected. Otherwise, consumer behavior will not change.

As a whole, the wireless application protocol, its associated WAP security elements, and the consistent user experience (CUE) defined in MeT have been proven to fulfill the requirements of the various stakeholders in multiservice environments. Furthermore, the Mobey Forum has stated its active support of the WAP identity module for remote mobile financial services. Thus we have clear indications that the major players of the financial industry all actively support WAP and its associated security elements as a viable solution for the mobile environment.

Still to be resolved, however, is the issue of how the WAP identity module is to be implemented. For operators, the obvious elements of a technical infrastructure show WAP and WIM implemented on the subscriber identity module (SIM). By issuing WIMs, operators can give customers new business opportunities. This in turn strengthens the customer relationship. On the other hand, the financial sector is reluctant to support this kind of implementation, since it takes the issuing process out of their hands and makes them dependent on operators. What is more, the financial sector is afraid that it will lose business opportunities when operators eventually begin offering payment services. Instead, to retain full control of the WAP identity module, the financial sector wants to issue WIM cards that are physically separate from the subscriber identity module—that is, it wants to see a dual-chip implementation.

The role of the handset manufacturer—Ericsson

Ericsson holds firm to the idea that the mobile phone will be equipped with a single digital ID based on an open, globally accepted standard. The phone will be used for

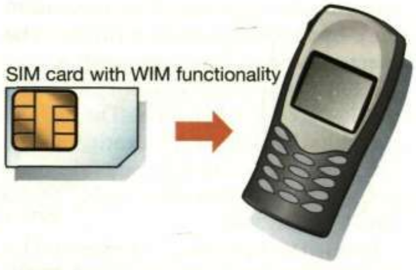


Figure 1
WIM on SIM implementation.

TRADEMARKS

Bluetooth is a trademark owned by Telefonaktiebolaget LM Ericsson, Sweden.

Java is a trademark or registered trademark of Sun Microsystems, Inc. in the United States and other countries.

MasterCard is a registered trademark of MasterCard International Incorporated.

MeT is a trademark owned by MeT and its sponsors.

VISA is a registered trademark of Visa International Service Association.

Figure 2
Dual-chip (separate SIM and WIM) implementation.

a multitude of trust-based services including e-commerce and secure access to e-mail clients, intranets, and eventually—when the Bluetooth infrastructure is in place—physical access to buildings. The basis for the digital ID is the WAP identity module. The mobile phone thus becomes the user's personal trusted device and enabler of the mobile society.

Ericsson will play an active role until the market for secure transactions has matured. Apart from implementing the required technology, to help build the market and create the critical mass of customers and services needed for market take-off, Ericsson is involved in several field trials and demonstrations to give *proof of concept* and to collect feedback. In Hong Kong, for example, all the operators have joined forces to set up a collective certificate authority. The field trial includes content providers, an infrastructure, and users (initially 100). And in the spring of 2001, Ericsson and RSA Security demonstrated secure access to e-mail

BOX B, TERMS AND DEFINITIONS

TERM	DEFINITION
Access PIN	PIN that protects WIM data and authentication keys (called PIN-G in the WIM specification).
Authentication	Verification of identity.
Authentication key	Private key used during authentication.
Authorization by user	Process whereby the user authorizes a transaction to be charged to his account. Involves an application-layer digital signature from the user, and provides proof of commitment.
Brand elements	Name, URL, images, or jingles that relate to a service or service provider.
Certificate database	Storage area in the PTD for service certificates and root certificates.
Consistent user experience	Similar user experience among phones of different makes and types. For example, any Web shopping user experience should be largely similar among MeT-compliant phones. CUE also includes consistency of user experience when using the same core function in different usage scenarios (for example, the user authorization experience should be the same in Web shopping and retail shopping usage scenarios).
Content provider	Provides goods and services to the user by hosting a content server.
Contract	Data object on which the user creates a digital signature when making a commitment (identical to the stringToSign in first realization).
Core functions	Basic functions on which MeT is based: initialization, registration, secure session, authentication and authorization.
Dual-chip	WIM implemented on a SIM-sized smart card separate from the SIM.
Dual-slot	SIM application toolkit-based solution for support of full-sized smart cards, currently supported in Motorola and Sagem phones only.
Initialization	Provides the PTD with one or more public–private key pairs and root certificates.
Issuer	Entity that has issued a service certificate for a key pair in the PTD. Typically a bank or a credit card company. The entity and its supporting infrastructure are used synonymously.
Local environment	MeT-defined environment in which the PTD gains access to content via a local or personal area network.
Personal environment	MeT-defined environment in which the PTD is used with other computational resources in a private (secure) environment.
Personal trusted device	Mobile communications device which can be personalized (through the registration of service certificates) and trusted by the user and service providers that require application-level public key security.
Registration	Provides the PTD with a service certificate that relates to a public–private key pair residing in the PTD.
Remote environment	A MeT-defined environment in which the PTD gains access to content via a public mobile network.
Root certificate	Certificate needed for handshake between server and client to achieve WTLS class 2 (server authentication) or higher.
Secure session	Guarantees confidentiality, data integrity and server authentication.
Security element	Component of the PTD that contains the user's key pairs and root certificates and is responsible for performing encryption and authentication functions.
SE initialization interface	The interface (process) that equips the PTD with key pairs and root certificates.
Service certificate	Certifies that a public–private key pair is valid for a specific service.
Service execution interface	Interface (process) that allows a content server to access MeT functions on the PTD.
Service registration interface	Interface (process) that enters service certificates into the PTD.
SignText	Application-level signature function in WMLScript Crypto Library (a WAP specification).
Sign-up	Process whereby the user subscribes to a service and registers the PTD (for the service).
Signature key	Private key used for creating digital signatures on contracts.
Signature PIN	PIN that protects a signature key (called PIN-NR in the WIM specification).
SWIM	SIM card with WIM application.
Transaction database	Storage area in the PTD for transaction data and secure objects, such as receipts and tickets.
Unblock PIN	PIN that can be used to unblock blocked PINs.
User verification	Process that requires the user to enter a PIN to permit access to a key pair.
WIM	Tamper-resistant device used to perform WTLS and application-level security functions, and to store and process information needed for user identification and authorization functions.
WIM card	WIM implemented on a smart card.
wPKI	PKI optimized for WAP.
WTLS	WAP equivalent of TLS.

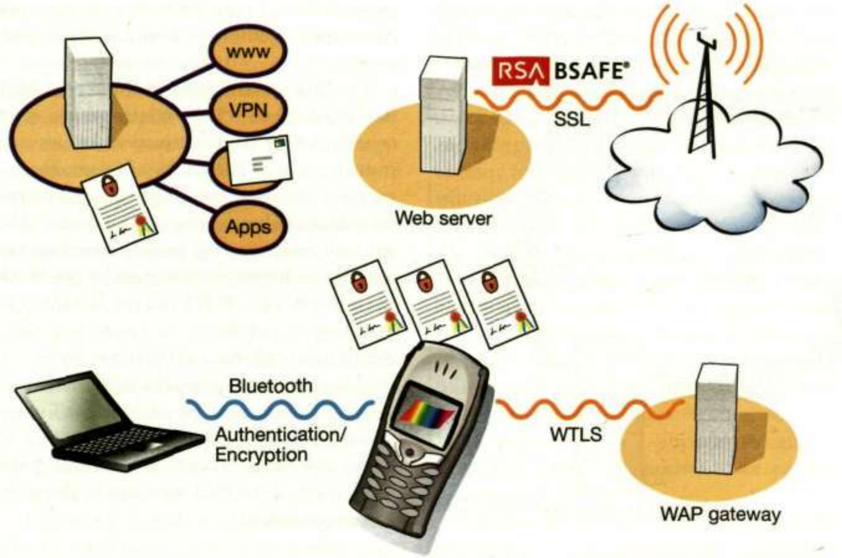


Figure 3
The WIM in a non-commerce application: secure access to an e-mail client.

clients—a PC was used for accessing services, and a mobile phone served as the authentication device.

Within the EU, activities are underway to regulate mobile e-commerce—for instance, in the eEurope smart card charter and the data protection act. Here, too, Ericsson is playing an active role to ensure that legislation takes into account the abilities and limitations of the mobile phone

Ericsson will also ensure that MeT compliance, including the level of security supported in its entire product portfolio, meets the requirements of the different stakeholders, since secure transactions are expected to be a mainstream application and not a feature exclusively targeted to advanced high-end users.

MeT

As discussed above, mobile phone manufacturers will face certain challenges as the use of mobile phones extends into new areas of application, such as secure transactions. To date, no single standardization body has provided a generic framework for regulating how business-to-consumer mobile electronic transactions (MeT) can be handled securely on mobile phones. To cope with this situation, Ericsson, Motorola and Nokia—the key players in the Mobile Internet market—jointly established the MeT initiative in April 2000. Panasonic, Siemens and Sony have since joined the initiative as sponsors, making MeT the most significant standardization body for mobile phone-centric electronic transactions. Apart from major handset manufacturers, network operators (France Telecom and Telia), security companies (for example, RSA Security) and the financial sector (including the Mobey Forum and BBVA) are also members of MeT, ensuring broad market and applications support.

To facilitate acceptance of all parties involved, and applicability in all relevant markets, the strategy for MeT has been to base the framework on existing standards, and to design it to be independent of underlying system standards. The core technologies for the MeT framework are

- WAP—the global, open application execution environment for the Mobile Internet;
- WIM—which provides the security elements required for electronic transactions;
- Bluetooth wireless technology—the core technology for local communication with good cost/performance ratio; and
- wPKI—the wireless public key infrastructure is the standard security framework for managing keys and certificates adapted to the wireless environment.

In short, the mission of the MeT initiative is to define how these core technologies will be used in mobile phones to enable elec-

BOX C, WIM IMPLEMENTATIONS

The WAP WIM specification, on which the MeT specifications are heavily dependent, defines the interface between the WAP client and the WIM module. The WAP WIM specification basically includes two parts with complementary objectives:

1. A functional, implementation-independent definition of WIM.
2. An example of WIM implemented as a smart card—based on the ISO7816 series of sotandards [ISO7816] and, where applicable, the related GSM specification [GSM_SIM].

The specification also clearly states that it has been written to enable alternative implementations based on secure token technology. The most likely implementation in the near future is called WIM on SIM, or SWIM, but other implementations can also be envisaged. For example, the financial sector is afraid of losing business opportunities when operators who own and issue SWIMs start to offer payment services. Consequently, to retain full control of their business, the financial sector proposes that WIMs should be physically separate from SIMs, a proposal referred to as dual-chip.

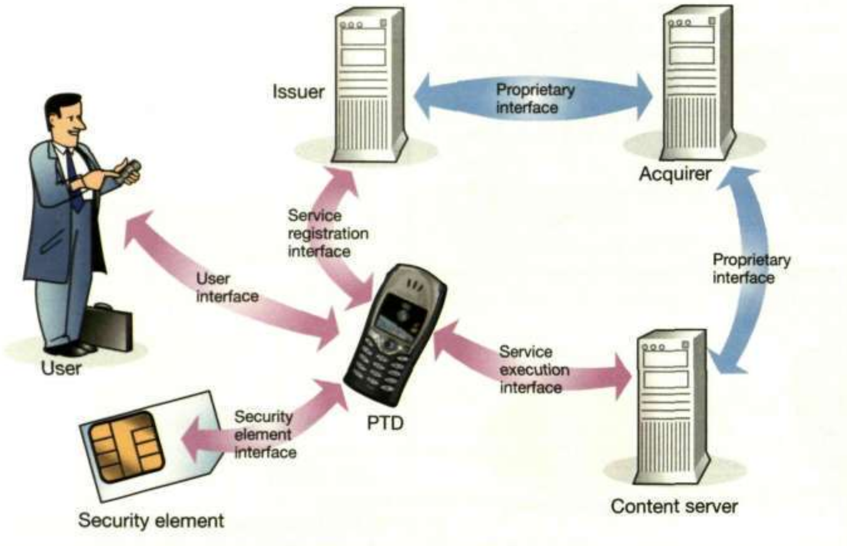


Figure 4
The MeT reference model.

BOX D, CORE MET FUNCTIONS

Five core functions have been defined for the MeT specifications:

Initialization
The mobile phone (PTD) is provided with public-private key pairs that are stored on a SWIM card or on a separate WIM card. The key pairs might also be created in the secure token.

Registration
The PTD is provided with service and root certificates that can be provided on the SWIM or WIM card at the time of initialization, or can be downloaded to the PTD by means of the WPKI framework.

Establishment of secure session
Transport layer security is used to establish a secure session. The technology employed might vary from environment to environment.

Authentication
The client is authenticated using the corresponding service certificate. Use of the certificate is controlled by a related access rule, which requires an access PIN.

Authorization by user
The PTD uses a signature key (different from the private key used for authentication) to create a digital signature. To access the signature key, the user must enter a signature PIN.

tronic transactions. The framework should thus prevent fragmentation and ensure global interoperability. The first comprehensive version of the MeT specifications, published in February 2001, includes

- the core specification, which defines a system reference model (Figure 4), applicable environments, a set of core functions, and security technologies;
- terms of reference;
- a definition of the personal-trusted-device concept;
- a core usage model for account-based payments; and
- usage scenarios, such as for remote banking, retail shopping and WAP shopping.

To complement these specifications, the MeT initiative provides guidelines on a consistent user experience for the entire security framework. A cornerstone of MeT, CUE defines the sequence of interactions in typical transaction phases, such as certificate download, authentication, and the process of entering a digital signature. In addition, CUE defines mechanisms for heightening user awareness of security and branding. The next revision of MeT specifications, due in early 2002, focuses on ticketing, receipts, the personal environment, and local payments.

To fulfill its strategy and objectives, MeT has established close relationships with other relevant standardization bodies and industry fora. The Mobey Forum, for example, provides the financial industry's view on mobile electronic transactions. The security framework is heavily based on the security technologies specified by the WAP Forum—WTLS, WIM and the WMLScript SignText. Similarly, in Bluetooth standardization efforts, an interest group is working to ensure that Bluetooth technology will be fine-tuned to support MeT usage scenarios in the local environment. And finally, the WAP Forum is addressing the adaptation of the PKI standard to the wireless environment.

Enabling technologies

As mentioned above, the strategy of the MeT initiative is to promote—and if necessary, extend—existing standards and technologies as the foundation for electronic transactions. Basically, three types of key enabler are needed:

- an open application execution environment that can support mobile phones when they
  - subscribe to a service; and
  - access and execute a service for electronic transactions;
- a generic but versatile security toolbox; and
- a range of access technologies that offer cost-effective solutions for the remote, local, and personal environments in which mobile devices are expected to operate.

In addition to embodying these key enablers, the mobile phone should be able to operate as a personal trusted device. That is, besides being a device for mobile telephony, it is also a device with the characteristics required for handling electronic transactions in a secure, trustworthy and consistent way.

Early on, the standardization activities of the WAP Forum were recognized as being key to the MeT initiative. The WAP browser environment, or WAP application environment and associated user agents, can efficiently handle service registration interfaces and the service execution interface. Furthermore, the WAP browser technology is expected to be available in virtually all mobile phones by the time the MeT specifications are deployed. This means that service designers have a generic application environment from which they can quickly and easily design, verify and provide new services. In the future, other open application environments, such as Java environments, will also probably be enabled to support electronic transactions. However, this is not presently the case for the Java 2 platform, micro-edition (J2ME)—before it can support electronic transactions, this version of Java must first be extended with class libraries that correspond to the security features found in WAP. This is expected in about two years (2004).

The WAP Forum is also driving activities to standardize other key technologies, such as

- transport layer security (WTLS and TLS)
- a generic security toolbox based on the WAP identity module; and
- wPKI.

Due to constraints in memory and processing power, the task of providing a security framework in a mobile phone is complex. Therefore, the security framework must be applicable to all applications that require security. It must also be applied consistently across all services and applications, so that end-users find it easy to use and can recognize the basic usage scenarios as they move from service to service.

The WAP protocols have been designed to be independent of system standards. As such, they can be run on practically any wireless access technology in the remote environment as well as on technologies for local and personal environments, such as Bluetooth. Enhancements to the Bluetooth and protocol technologies are being considered in relevant standardization groups to better adapt these technologies to the MeT usage scenarios. Examples include shorter discovery times and faster connection set-up in payment scenarios.

Conclusion

Although the market for secure transactions is still immature, there are clear indications that a multiservice mobile society is emerging. All major handset manufacturers, including Ericsson, have joined forces in order to define how security is to be implemented in the mobile phone and how it is to be used. Local markets around the world are maturing, which means that users are becoming increasingly willing to use their mobile phones for new kinds of services and applications. Banks and operators, currently the two strongest players in the mobile transactions arena, are setting their roadmaps. In mature markets, such as Scandinavia, we will see a roll-out of remote financial services in 2002; other markets will follow in 2003. Services that depend on Bluetooth-enabled infrastructure, such as local payments and physical access to premises, will need another three to four years of deployment.

Ericsson is convinced that future mobile phones will be equipped with unique digital credentials to be used for a multitude of services. The phone is thus becoming a personal trusted device, not only because of technical advances but also because consumers are increasingly willing to use their phones for non-traditional services, such as transaction-based services. Many new technologies and applications (Java, MExE, transaction-based services) will be or already are supported in the current portfolio. All of them demand security. Since the digital ID is essentially a WAP identity module (WIM), and WAP protocols have been designed to be independent of systems standards, the security elements can be used in this broad spectrum of applications.

The obvious placeholder for the WIM is on a card stored in the phone. Built-in solutions strongly discourage consumers from using multiple phones. In Ericsson's current product portfolio, many phones already support the combined WIM on SIM (SWIM). Moreover, beginning in 2002, Ericsson plans to implement this solution throughout its entire product portfolio. SWIM offers the level of security required by the major players, and does not put constraints on design or burden the handset manufacturer with liability. Of course, Ericsson acknowledges the need to grow the market for secure transactions, and is therefore also investigating alternative WIM implementations.

BOX E, STANDARDS BODIES

MeT (www.mobiletransaction.org)
MeT defines a technical framework for secure electronic transactions from a mobile phone, ensuring consistent user experience. Its members include all major handset manufacturers, security companies, the financial sector, and operators.

Mobey Forum (www.mobeyforum.org)
The mission of the Mobey Forum, which is driven by the financial industry, is to encourage the use of mobile technology in financial services, such as payments, and remote banking and stock trading. Its members include BNP Paribas, Ericsson, Nokia, Nordea, and VISA International.

GMCIG (www.gmcig.org)
GMCIC is an open and global organization that develops and submits secure wireless standards to existing standards organizations. Its members include Deutsche Postbank AG, Europay International, Motorola, and NTT DoCoMo.

Radicchio (www.radicchio.org)
Radicchio aims at establishing a common foundation for secure m-commerce by reaching a consensus on important interoperability issues. Its members include Ericsson, Gemplus, Siemens, Sonera Smarttrust, and Mastercard.

WAP ecomeg
WAP ecomeg is a working group of the WAP Forum. The focus of the group is on all issues relating to e-commerce.

BOX F, PRODUCT ROADMAP

By working with standardization and through an aggressive product roadmap, Ericsson has shown its commitment to the MeT initiative.

- With the R320, Ericsson introduced WAP technology with a basic level of security.
- With the R520, T39 and T65, Ericsson's product portfolio enables basic MeT usage scenarios by providing key technologies, such as
  - SWIM (WIM on SIM);
  - root and client certificates through combined initialization and registration;
  - transport layer security through WTLS class 3; and
  - application-level digital signature support (SignText).
- Ericsson's next generation of mobile phones is expected to be fully compliant with MeT—that is, in addition to supporting the features mentioned above, future phones will support WPKI and fulfill MeT requirements for consistent user experience (CUE).

On-demand mobile media—A rich service experience for mobile users

Erik Ekudden, Uwe Horn, Mats Melander and Jan Olin

Images, voice, audio and video content—indeed, all kinds of rich media—will soon reach mobile terminals as part of mobile multimedia services. The introduction of standardized mobile messaging services, such as EMS and MMS, will enrich person-to-person messaging and pave the way for content-push services. On-demand mobile media services will be delivered to users via media streaming and download techniques that enrich mobile browsing and general content access.

The authors describe the business opportunities, technology, and product strategies for the delivery of on-demand mobile media.

Introduction

Multimedia-enriched services are expected to drive usage, operator revenue and bandwidth consumption in mobile networks. At the center of these multimedia services are end-users with their wireless mobile devices. End-users are likely to have great expectations, since in the end, they are the ones who must pay for service. Therefore, the services must be compelling. Unlike the delivery of media on the fixed Internet, where service is free but is not backed by any guarantees, quality of service (QoS) is important in the wireless Internet.

On the Internet, news sites are being enriched with short audio and video clips, comic strips are being turned into animated graphics, and simple text pages are being enhanced with images and scrolling text. The technology with which to deliver on-demand media is currently in the process of being standardized and implemented in mobile devices. While the basic technology of the wireless Internet is similar to that of the fixed Internet, the fundamental challenges are not, and the associated services will probably also differ.

Mobile multimedia begins with messaging services. Enhancements to short message service (SMS) are already well on the way in the form of media-rich EMS and MMS.[1] In the case of WAP and access to Web content, two important enhancements brought by GPRS and WCDMA services are media download and media streaming.

Thanks to the increase in data-transmission rates in GPRS and WCDMA mobile networks—from 40 to 384 kbit/s for typical wide-area coverage—operators and content providers who are able to deliver attractive, high-quality media services have major business opportunities.

Mobile services and the merits of on-demand mobile media

Delivery of mobile media service

Over time, all existing services will probably be enhanced with rich media in order to meet growing user expectations. Rich media content also forms the basis of many new services. Figure 1 provides a selection of possible, near-term, on-demand services that are well suited to mobile use, including news and sports services, live cameras providing traffic information, and surveillance. Likewise, many entertainment services will be attractive and grow in popu-

BOX A, TERMS AND ABBREVIATIONS

3GPP	Third-generation Partnership Project	ROM	Read-only memory
API	Application program interface	RTCP	Real-time transport control protocol
CDR	Call detail record	RTP	Real-time transport protocol
DRM	Digital rights management	RTSP	Real-time session protocol
EMS	Enhanced messaging service	TCP	Transmission control protocol
GPRS	General packet radio service	SDP	Session description protocol
HTML	Hypertext markup language	SMIL	Synchronized media integration language
HTTP	Hypertext transport protocol	SMS	Short message service
IETF	Internet Engineering Task Force	UDP	User datagram protocol
IP	Internet protocol	UMTS	Universal mobile telecommunications system
MIPS	Millions of instructions per second	W3C	World Wide Web (WWW) Consortium
MMS	Multimedia messaging service	WAP	Wireless application protocol
O&M	Operation and maintenance	WCDMA	Wideband code-division multiple access
OEM	Original equipment manufacturer	WML	Wireless markup language
PDA	Personal digital assistant	XHTML	Extensible HTML
QCIF	Quarter-CIF (common interchange format)		
QoS	Quality of service		
RAM	Random access memory		

Figure 1
Examples of on-demand media services.
"Garfield" courtesy of Paws, Inc. All rights reserved.
Tombraider © 2001 Paramount Pictures Corp. Tombraider and Lara Croft are trademarks of Core Design Ltd. All rights reserved.
BBC courtesy of BBC News.

larity—the cartoon of the day, quiz of the day, movie previews in ticketing applications, and so on.

For some applications, such as news summaries, a bit rate of less than 20 kbit/s will provide adequate quality for voice and images. But for others, such as live sporting events, the bit rate must be at least 256 kbit/s (Figure 2). Many interesting lower bit rate services can efficiently make use of slide shows in which the audio is combined with a set of images. The elements of a presentation can also be tied to active links, offering interactivity. Truly powerful mobile multimedia services are based on a strict set of standardized capabilities, to ensure interoperability, but permit enough flexibility to be composed differently for different purposes.

The delivery of on-demand mobile media will enable powerful services. But perhaps even more important are the combinations of services and increased traffic that will result thanks to the wealth of content transported over the mobile network—when users run across content that is interesting or humorous, they often want to share it with friends. Thus, a benefit of on-demand services is that they can give rise to an increase in messaging. Similarly, persons who receive interesting media clips might feel inspired to test new on-demand services.

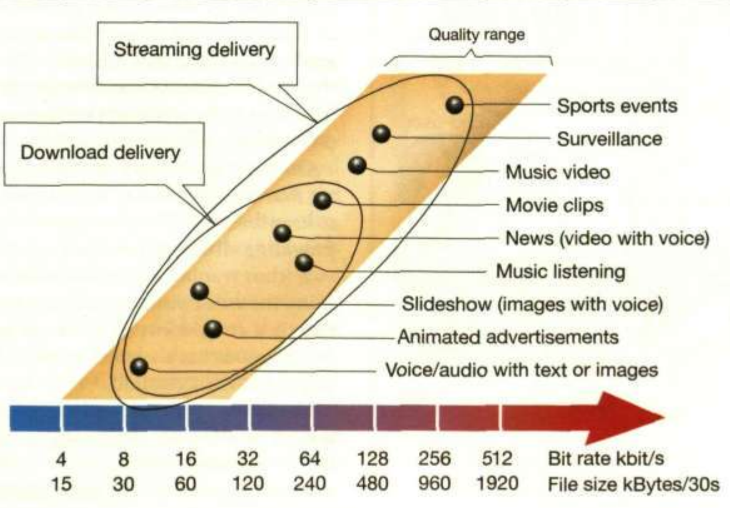


Figure 2
Bit rate demands and storage requirements for different on-demand mobile services. The storage requirements are expressed in kilobytes and computed for a clip duration of 30 seconds.

BOX B, A MOBILE TICKETING SERVICE

An operator might offer an advanced ticketing application through a WAP cinema portal. When a user visits the portal, a positioning service determines the user's geographical location. Based on this information, a list of nearby movie theaters and associated program schedules is obtained from a database on the Internet. The attributes of the films are matched against the user's preferences, which are stored on a profile server. In less than five seconds, the service generates a WAP page showing a selection of available movies and show times. If the user clicks on a preview button while browsing through the list of films, then a streaming service is invoked to play the corresponding movie trailer on the mobile device. If the user likes what he sees, he can purchase a ticket—the purchase amount is automatically charged to the user's phone bill.

User benefits of mobile media

In many respects, the delivery of on-demand mobile services differs from that of fixed Internet services. While the multimedia capabilities of mobile devices are more limited than PCs, the convenience of an always-on, always-with-me device for on-demand media delivery is distinguishing. Take ticketing services, for example (Box B)—clearly, mobile users enjoy significant benefits when components are properly integrated into advanced services:

- mobility (rich media anywhere, anytime);
- high quality of service;
- presentation that is suited to the mobile devices (voice/audio and images/video make a better interface than scrolling text)
- complete service packages; and
- location-based multimedia services.

A new channel of media consumption is opening up as the media, Internet and wireless industries converge to yield wireless multimedia. Given the forecasted number of subscribers of third-generation services, the media industry is anxious to provide content to users anywhere, any time.

Business opportunities and operator roles

Compelling multimedia services will certainly attract many users. This, in turn, will generate interest on the part of various companies in providing wireless multimedia-based services. Notwithstanding, it does not make sense to require every service provider to dimension its equipment to serve millions of users. At the same time, however, every potentially successful service must be dimensioned appropriately. A successful service is also dependent on adequate quality of delivery. Many services will cost (subscription fee or pay-per-view), which means that a transaction must be carried out for the service to take place. Some services might also benefit from knowing the identity of the user, so that individual preferences can be served, service access can be managed, and so on.

End-users expect service to be good and worthy of the price they pay. Media service providers expect a channel to consumers which provides adequate quality, enables them to get paid, and which protects against piracy. An operator that chooses to take on the role of media distributor can ensure that these expectations are fulfilled.

By functioning as distributor and enabling operator-powered services, operators have many opportunities to add value to Mobile Internet services—in particular, to mobile multimedia services.

The value chain in Figure 3 shows an overview of the different roles that are required to enable mobile multimedia services. The gray boxes represent roles that can solely be fulfilled by the operator. The green boxes represent roles that can be fulfilled by the operator or its partners. The blue boxes denote functions affected by new technologies. These are introduced as enablers of on-demand mobile media services.

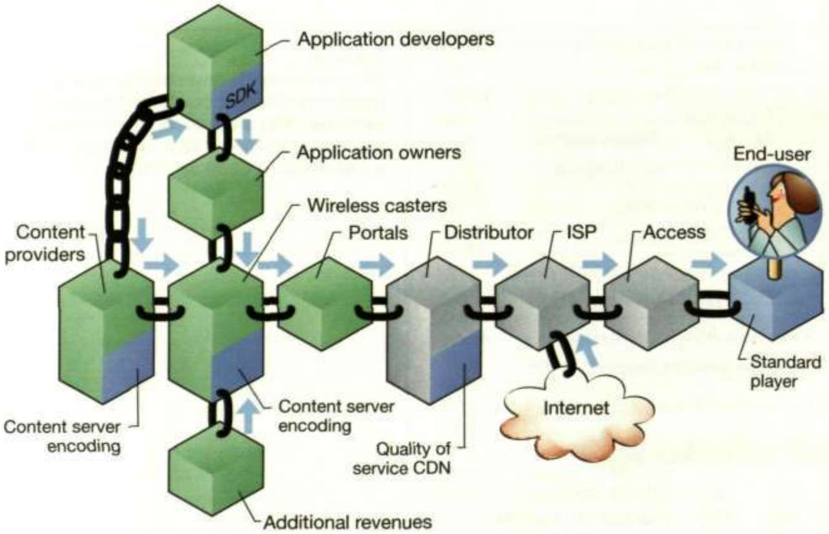
The content provider (Figure 3, left) owns the content. The wireless caster provides the actual service. The service is developed by the application developer. The owner of the application owns and probably also brands the service. Portal providers set up portals from which the end-user can find and access on-demand mobile media services.

On a service-by-service basis, the operator may decide which of these roles it wants to keep for itself. There is a trade-off between providing all the elements in a complete service (thus reaping all the revenues and carrying the entire risk) and offering a product that is a component in a service (reduced risk). I-mode has shown that only a fraction of services is successful.

Flexibility is crucial, and the operator has several options in deploying a content-delivery service:

- The operator can host a content-delivery

Figure 3
Roles in a simplified value chain of mobile media distribution.



service via servers inside the operator domain. The operator would thus function as a broker and content manager.
- The operator can provide operator power to enhance content from any service partner. The end-user will be able to browse any Internet site and access streaming content. Through agreements with service providers, the operator delivers operator-powered services to its subscriber base.
- The operator can combine these scenarios through its own portals.

Key factors of success for future operators and the Mobile Internet are quality multimedia content and portfolios of compelling applications (possibly provided by partners).

Content distribution chain

Figure 4 gives on overview of the media distribution chain, which describes the steps that an operator must take to deliver on-demand multimedia content to end-users. The five most important steps are:
- content creation;
- distribution of content to servers, from which it can be accessed on-demand;
- content management, which helps the end-user to find specific content of interest;
- delivery of content from servers to the end-user; and
- content presentation.

The process of creating content starts with raw content, which originates from cameras, VCRs, audio sources, CDs, and so on. If the raw content solely exists in analog format, it must first be digitized.

Most raw digital content needs editing, which typically produces a set of elementary video and audio clips and still images. Some text might also be generated at this stage to accompany or complement other content.

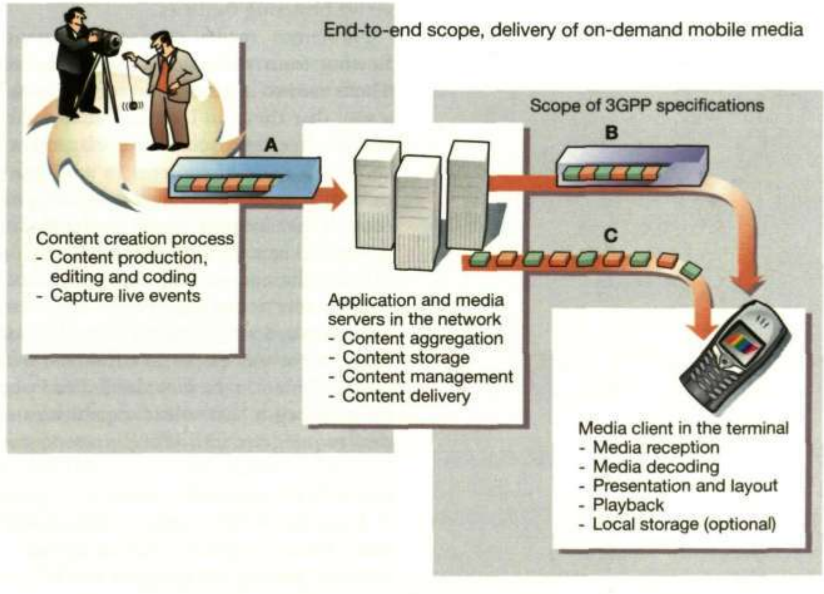
Before multimedia content can be delivered to end-users, it must be transformed into a format that is compatible with the end-user device (physical display size, processor power) and network (protocols and available bit rates).

For example, to transport video content over low bit rates (say, 64 kbit/s), the video resolution must first be decreased, typically using quarter-CIF (common interchange format) at 176 x 144 pixels and 10-to-15 frames per second. Moreover, the video must be compressed to further reduce bandwidth—quality decreases at lower bandwidths, but also depends on the complexity of the scenes. Greater motion and more complex textures require larger bandwidth. This is why sporting events are so demanding, whereas news services can get by with less bandwidth (Figure 2).

Different kinds of media require different codecs. The compression of a TV signal, for instance, results in two bit streams: one audio and one video. Since the two belong together, logically, they need to be processed and played back in sync. Therefore, they are usually stored together in the same media file according to a certain file format. Besides the actual bit stream, the media file contains
- information on the codecs—this allows synchronized playback of the bit streams; and
- information on the creator and owner of the content.

After the content has been edited and compressed it is put onto a media server. It is also usually put into the context of a specific application. In the simplest case, this context is provided by a Web or WAP page, which contains links to the media clips on the media server. However, HTML and WML are limited in their abilities to deal

Figure 4
Distribution of mobile media content. The content-creation process (left) results in distinct media files (A). The process of delivering content (right) contains two options: media file download (B), and media streaming (C).



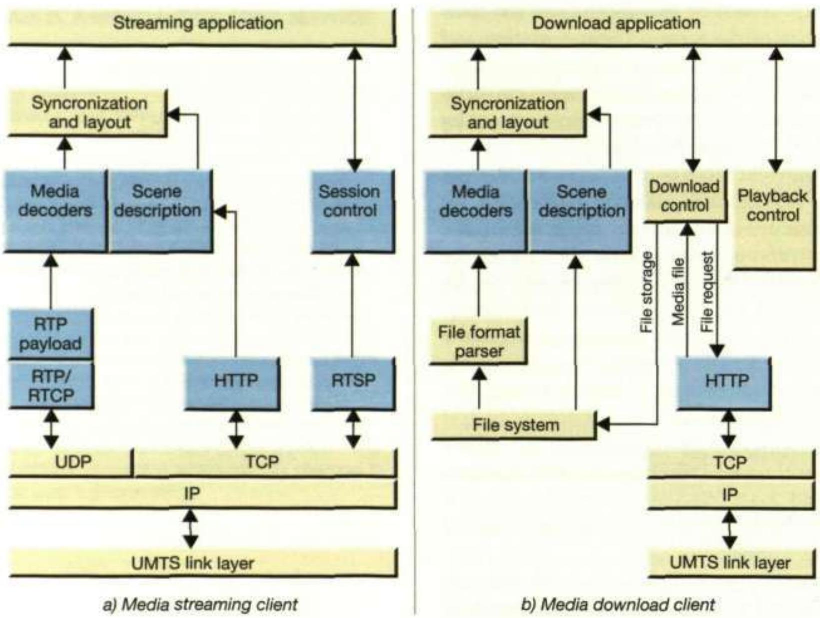


Figure 5
Mobile media player protocols and their applications. The protocols ensure simultaneous playback of voice, audio, images, video, and formatted text into mobile multimedia applications.

with continuous media types, such as audio and video. Therefore, specific scene-description languages (like SMIL) have been developed allowing application developers to express relationships, in space and time, between media elements, and to add interactive browsing facilities.

Numerous media clips are constantly flowing from professional media services. These need to be stored and tagged in such a way that they can later be found and retrieved. In essence, content management is about the reuse of content. In addition to providing links to content, more advanced content management can be used including integrated content search functions.

Once the content has been put onto a media server, it can be delivered on-demand via download or streaming. To download content, the user clicks on a link and waits for the content to be downloaded and playback to begin. Download capabilities are easy to integrate into Web browsers, since the hypertext transfer protocol (HTTP) can be used for downloading files.

To access streaming data, the user clicks on a link to start playback, which is almost immediate. Because streaming is a semi-real-time service that receives and plays back data at the same time, it puts greater demands on protocols and service implementation, especially when the service is to work over networks with little or no quality of service.

During playback of downloaded or streaming media, the compressed bit streams of the different media types are decoded and played back in sync according to the synchronization information in the downloaded file or in the streamed data.

Mobile media delivery

The greatest technical challenges associated with on-demand mobile media services have to do with wireless delivery and presentation. Because the last hop of the content distribution chain is over a wireless link to the mobile terminal, all content must be coded and compressed to make efficient use of spectrum. Thus, before the content can be played back, it must be uncompressed and decoded in the terminal. In general, however, most mobile terminals have limited memory, processing power, display size, and audio capabilities. The lack of memory limits storage capacity for multimedia clips; small displays limit the resolution of images and video. Finally, in contrast to PCs, mobile terminals usually cannot dynamically download and upgrade software components, such as new media codecs. Notwithstanding, service quality must be high to attract end-users, content owners and application developers. And spectrum must be used efficiently to keep pricing at a reasonable level.

The capabilities of mobile terminals and of the mobile network are limited compared to the fixed Internet, but the mobile application environment is much more heterogeneous. Today, a variety of mobile terminals (PDAs included) can play back multimedia content. In addition, mobile systems encompass a variety of different radio access networks, each of which has different physical characteristics that limit typical radio access link rates. Third-generation systems also contain a sophisticated UMTS bearer model that allows services to use bearers with different quality of service, ranging from best-effort bearers that demand fewer resources to more expensive bearers that demand considerable resources but offer excellent link quality at high bit rates.

A comparison of the two main delivery mechanisms, namely downloading and streaming, shows that it is much more chal-

BOX C, 3GPP ON-DEMAND MEDIA PROTOCOLS

The protocols used for on-demand mobile media consist of several layers and can be grouped into UMTS protocols and TCP/IP protocols. Access to Internet-based servers can be provided to third-generation terminals using TCP/IP over UMTS. The 3GPP has defined a UMTS bearer model that provides different traffic classes. The streaming traffic class in this model can guarantee a minimum bit rate over the UMTS network. The TCP/IP protocol stack consists of three layers (Figure 5):

1. In the network layer, the Internet protocol interconnects all networks between destination and source nodes;
2. Either UDP or TCP is used in the transport layer; and
3. Depending on the service, several protocols can be used in the session layer.

The hypertext transport protocol (HTTP), which is used for browsing applications on the Internet, provides the semantics of requesting and transferring information between servers and clients in a distributed and collaborative way. Most often, the information requested from the server consists of HTML objects. However, HTTP is not restricted to HTML—essentially any information can be requested and transferred between the client and the server, including media objects. HTTP is used on top of TCP.

The real-time transport protocol[2-3] (RTP) is a well-established transport protocol for real-time data. Designed to run on top of UDP, it is used by many applications including IP telephony. It is currently the preferred way of sending audio and video data over packet networks. Associated with RTP is the real-time transport control protocol (RTCP), which reports transmission quality to the server.

RTP adds a time stamp and a sequence number to each UDP packet in a special RTP header. The time stamp is related to the sampling or the presentation or composition time of the media carried in the payload of the RTP packet. It is used for playing back media at the correct speed, and together with RTCP, it is used for synchronizing the presentation of other streaming media. A payload specification defines the interpretation of the time stamp and other RTP fields. The recipient uses the sequence number to detect the loss of packets. Statistics on loss can be reported to the server by means of RTCP.

The real-time streaming protocol[4] (RTSP) is used as a session control protocol for media streaming applications. Several features and semantics in RTSP have been inherited from HTTP. Using RTSP, a client streaming application can establish a session with a media-streaming server. Using this session, the client can ask the server

- to start streaming media;
- to pause, back-up and replay, and fast-forward streaming media; or
- to stop streaming and disconnect the session.

RTSP is usually used on top of TCP but can also be used on top of UDP.

The W3C defined the synchronized media integration language (SMIL) for the synchronization and layout of media streams for on-demand media applications. Profiles can be created from this language. Moreover, several modules have been defined in the SMIL specification. A SMIL module is a collection of semantically related elements and attributes; a language profile is the collection of modules that pertain to a particular application domain. For streaming media, the 3GPP has defined a profile for SMIL using a subset of the SMIL 2.0 language profile.

lenging to integrate streaming services into mobile application environments than it is to download multimedia. To ensure interoperability between clients and servers from different vendors, streaming requires the standardization of media codecs and streaming protocols.

The IETF has worked quite some time on the standardization of streaming protocols, but as yet a common Internet streaming standard has not been established. Instead, all major vendors of streaming solutions have developed proprietary solutions. Furthermore, because the task of addressing the specific challenges of mobile streaming lies outside the scope of the IETF, Ericsson initiated a work item on packet-switched streaming in the 3GPP. The initiative, which is supported by all major telecom companies, was formally approved in June 2000. Figure 5 depicts the architectures of

- a client that is streaming media (shown left); and
- a client that is downloading media (shown right).

The grey boxes denote the components that are subject to the IETF and 3GPP standardization of on-demand mobile media. For details on the protocols used in 3GPP, see Box C; the media codecs are discussed in Box D. Note: the client on the right requires only standardized components for media decoders and the scene description. The transmission protocols associated with streaming components (RTP/RTCP, RTP payload, and RTSP) have been replaced by a download/playback control function that downloads a media file over HTTP and plays it back.

3GPP release 4 (March 2001) includes

two specifications that define the basic 3GPP mobile streaming standard.[5-6] The focus of further work in the 3GPP is on additional media types and further enhancements to transport over wireless links, capability exchange, and content protection.

Robust transport over wireless links

Users have little tolerance for "wireless" disturbances and the effects of outages. Hence, the quality of any pay service must live up to end-user expectations. In the context of an interactive streaming service, quality refers to the standard of received video or audio streams and to latency aspects, such as service response times. For instance, playback must start at most a few seconds after a stream has been requested. Furthermore, interruptions of any kind must be avoided during playback. This last requirement represents a major challenge, given that connections can break temporarily (100-1000 ms) during cell handover.

Capability exchange

As noted above, the capabilities of mobile terminals and of the radio access network vary. Notwithstanding, the task is to give the largest possible group of users access to content and to deliver it at a level of quality that matches the capabilities of the mobile terminal and radio access technology in use.

The 3GPP is addressing the issue of heterogeneity through appropriately designed capability exchange mechanisms which allow the terminal and media server to negotiate

- the capabilities of the mobile terminal;
- the capabilities of the mobile network; and
- user preferences.

This approach permits the server to send multimedia data that has been adapted to the capabilities of the user's mobile terminal and network. For example, a user who accesses a specific service via a WCDMA network can have content delivered at higher bit rates than someone who accesses the service via a GPRS network.

Content protection

Content must be protected to secure the value chain. That is, if the rights of the content owner cannot be protected, then valu-

BOX D, 3GPP MEDIA CODECS

Every media codec developed and standardized in, or adopted by, the 3GPP is suitable for use in wireless applications. Important factors affecting codecs are bit rates, complexity (in terms of millions of instructions per second, MIPS), memory (in terms of RAM and ROM), quality in error-free conditions and in typical wireless environments, and cost of implementation. The available, standardized media types for text, voice, audio, images and video are listed below:

- XHTML basic: XHTML Basic is designed for Web clients that do not support the full set of XHTML features; for example, Web clients such as mobile phones, PDAs, pagers, and set-top boxes. The document type is rich enough for content authoring.
- AMR: The AMR voice codec (3GPP TS 26.090) encodes narrow-band (4 kHz "telephony" bandwidth) voice using one out of 8 rates between 4.75 kbit/s and 12.2 kbit/s. The media type is well suited to all voice and news clips. AMR was standardized by ETSI in 1998.
- AMR wideband: The AMR wideband voice codec (3GPP TS 26.190) encodes wideband (8 kHz audio bandwidth) signals using one out of 9 rates between 6.60 kbit/s and 23.85 kbit/s. The media type is well suited to all kinds of voice and audio clips, and provides high-quality sound. AMR wideband was standardized by 3GPP in 2001.
- MPEG-4 AAC-LC: The low-complexity version of the Advanced Audio Codec of MPEG-4 provides a range of bit rates and quality levels. The media type is well suited to high quality audio at bit rates of 64 kbit/s and above. MPEG-4 AAC was standardized by ISO/IEC in 1999.
- JPEG: The JPEG compression provides lossy compression of typically photographic images with a compression ratio of 4:1 to 30:1 leading to typical mobile compressed file sizes of 5-20 kBytes. JPEG was standardized by ISO/IEC in 1992.
- GIF: A format for color raster images which is among the most widely used formats for graphics compression. While GIF87a supports still images only, GIF89a also supports simple animation sequences.
- H.263: The H.263 standard supports video compression for applications, such as video-conferencing, video-telephony, and video-streaming. H.263 aims particularly at video coding for low bit rates (typically 20-30 kbit/s and above). H.263 baseline was approved in 1995, while the latest version including the profile and level specifications is from 2000.
- MPEG-4 visual: The simple profile is suited for mobile application scenarios since it is optimized for low bit rate coding and includes coding tools for increased error-resilience. Level 0 of the simple profile supports video formats up to QCIF resolution at a maximum frame rate of 15 Hz.

able content will never find a legal way into on-demand mobile media applications. There are several levels of content protection ranging from simple solutions which simply disallow the storage of the consumed content up to complete digital rights management (DRM) solutions which use encryption and conditional access to protect and manage access to valuable multimedia content.

Ericsson strategy and products

Ericsson's stance on on-demand multimedia is to embrace and drive standards through the 3GPP. Ericsson is taking an end-to-end approach to the issue by collaborating with

- strategic partners who are addressing the needs of the media industry (including 3GPP multimedia-compatible authoring tools); and
- suppliers who will provide 3GPP multimedia-enabled handsets.

Terminals

By providing an integrated platform solution for OEM vendors, Ericsson ensures that the basic technologies required for on-demand mobile media are available in an efficient and optimized way.

Ericsson's mobile platform includes hardware and software components that have been optimized for small, built-in environments. Moreover, the platform contains the protocol and media decoder solutions for streaming and downloading media.

Vendors of terminals who develop products on Ericsson's mobile platform access the services through an open API. This solution, which permits vendors to develop applications and graphical user interfaces on top of the platform, enables vendors to differentiate their products from those of other vendors.

The platform uses open standards and is tested and verified through reference designs that prove the concept of interoperability. Specifically, the platform provides the following on-demand media components:

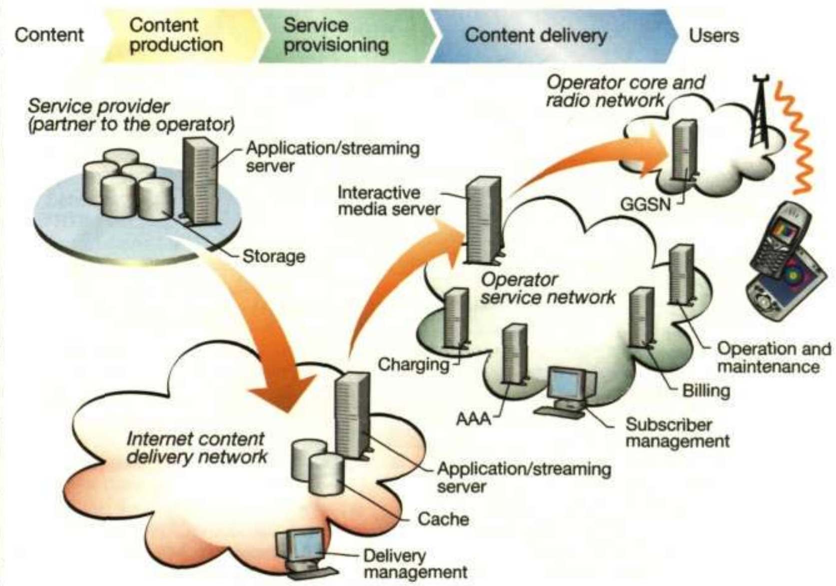
- a UMTS and GSM/GPRS protocol stack;
- a TCP/IP protocol stack and associated session protocols, such as RTP/RTCP, RTSP and HTTP;
- support for file formats of downloaded media files;
- media decoders for audio, video, still images, and voice; and
- an integrated environment that enables on-demand media applications to access the platform components through an open API.

The Interactive Media solution

The Ericsson Interactive Media suite consists of products that enable the operator to function as a distributor of on-demand multimedia services. The core product, the Interactive Media server, works with authoring tools from leading vendors and with 3GPP-compliant terminals from any vendor. The Interactive Media products are also compliant with the Ericsson service network framework.[7]

In a typical use case, the end-user finds a link to compelling streaming content while browsing. Since the service provider to this link has an agreement with the operator, the request is redirected to the Interactive Media server in the operator environment (Figure 6). If the content has not been stored there in advance, it is trans-

Figure 6
Typical use case of interactive media.



ferred (almost immediately) to the Interactive Media server.

The interfaces of the Interactive Media server (Figure 7) contain the following functions:

- network integration node (this contains Ericsson's value-adding functions);
- streaming server component (the component holds streaming data and functions as a "bit pump"); and
- optional Web server (the server holds static data, such as SMIL presentations and images).

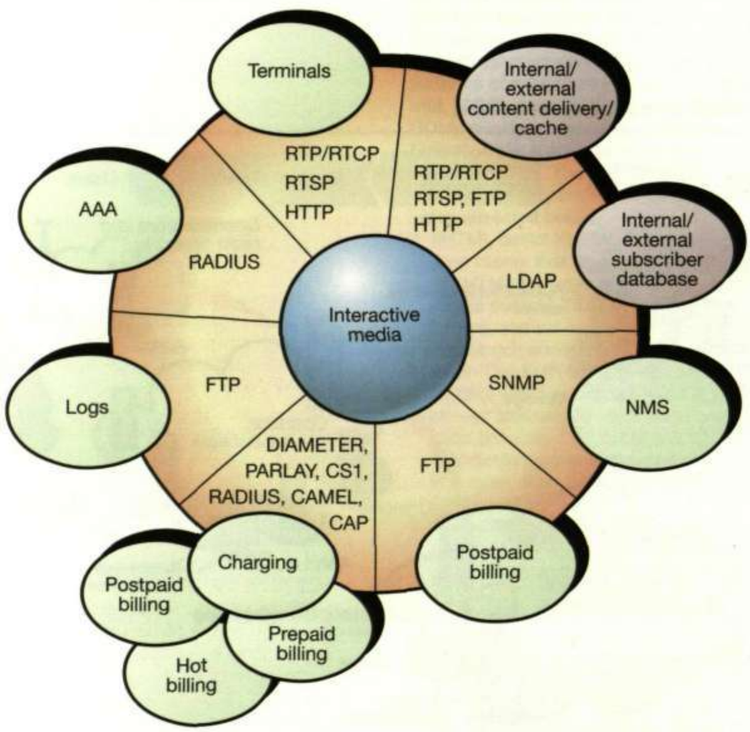
The network integration node, which is the heart of the Ericsson solution, enables the operator to take on the value-adding role of media distributor, thereby securing a sustainable future in providing services instead of merely serving as a transport provider.

The network integration node is designed to select media according to current network conditions. It does so by managing the complete set of ongoing sessions, instead of allowing sessions to be managed independently of one another as is the case in ordinary Internet solutions. Multiple sessions affect one another's performance, since all sessions must share the air interface. The network integration node makes optimized decisions in an aim to yield as many completed high-quality sessions as possible. Its decisions control each session in terms of the bit rate and codec that are enabled in the authoring process. The bit rate can be changed by changing the mode from, say, video streaming to audio with static elements, or by changing the actual bit rate of the stream.

The network integration node ensures that media delivery services are smoothly and seamlessly integrated into the operator network environment. The integration includes

- flexible service charging;
- authentication and the mapping of IP addresses to MSISDN;
- subscriber database connection—to manage user sessions; and
- operation and maintenance, including statistics.

Figure 7
External interfaces to the Interactive Media server.



Deployment scenarios

The Interactive Media solution has been designed to give flexible deployment. The product can first be deployed as a centralized solution—a high level of access security can be maintained when a centralized solution is employed as a gateway to external content or service providers located outside of the operator domain. Later, as the number of subscribers increases, the Interactive Media server can be deployed in a distributed configuration within the operator environment.

Conclusion

The building blocks for on-demand mobile services have been standardized in the 3GPP, providing a solid base for multimedia services. Key factors of the success of operators and the Mobile Internet are quality multimedia content and a portfolio of compelling applications that comply with these building blocks.

Operators have numerous opportunities to add value to Mobile Internet services. In particular, they can function as a distributor of multimedia services and assume multiple roles in the value chain. Indeed, operators will be able to add value to any multimedia service, regardless of who provides it.

With its Interactive Media suite, Ericsson is well-prepared to fulfill the needs of operators who want to be first in the market with new and compelling multimedia services. At the same time, Ericsson provides an integrated multimedia-enabled terminal platform for OEM vendors. This platform enables on-demand media applications to access platform components through an open API.

In the future, more services will be enhanced with rich media. The first steps in this direction have already been taken for messaging and on-demand services. More advanced and powerful media types, such as vector graphics and synthesized audio will also be supported. Improved transport mechanisms will further enhance the quality of on-demand mobile media services.

REFERENCES

1 Novak, L. and Svensson, M.: MMS—Building on the success of SMS. Ericsson Review Vol. 78(2001):3, pp. 102-109.
2 Schulzrinne, Casner, Frederick, Jacobson, "RTP: A Transport Protocol for Real-Time Applications," IETF RFC 1889, http://www.ietf.org/rfc/rfc1889.txt.
3 Schulzrinne, "RTP Profile for Audio and Video Conferences with Minimal Control," IETF RFC 1890, http://www.ietf.org/rfc/rfc1890.txt.
4 Schulzrinne, Rao, Lanphier, "Real Time Streaming Protocol," IETF RFC 2326, http://www.ietf.org/rfc/rfc2326.txt.
5 3GPP TS 26.233, Transparent End-to-End Packet Switched Streaming Services (PSS); General description (Release 4); ftp://ftp.3gpp.org/Specs/2001-03/Rel-4/26_series/26233-400.zip.
6 3GPP TS 26.234, Transparent End-to-End Packet Switched Streaming Services (PSS); Protocols and codecs (Release 4); ftp://ftp.3gpp.org/Specs/2001-03/Rel-4/26_series/26234-400.zip.
7 Boman, L.: Ericsson's Service Network: A 'Melting Pot' for Creating and Delivering Mobile Internet Service, Ericsson Review, Vol. 78 (2001):2, pp. 62-67.
8 Handley, Jacobson, "SDP: Session Description Protocol," IETF RFC 2327, http://www.ietf.org/rfc/rfc2327.txt.
9 3GPP TS 22.140, Technical Specification Group Services and System Aspects; Service aspects; Stage 1 Multimedia Messaging Service (Release 4); ftp://ftp.3gpp.org/Specs/2001-03/Rel-4/22_series/22140-410.zip.
10 3GPP TS 23.140, Technical Specification Group Terminals; Multimedia Messaging Service (MMS); Functional description; Stage 2 (Release 4); ftp://ftp.3gpp.org/Specs/2001-03/Rel-4/23_series/23140-420.zip.

WAP 2.x architecture—Features, services and functions

Peter Arnby, Johan Hjelm and Peter Stark

Web services adapted to the wireless environment have been a tremendous success, especially in Japan, where more than 25 million people use i-mode. However, the architecture of i-mode has some limitations. These will be overcome with WAP 2.x.

In this article, the authors describe WAP 2.x, the universal wireless application environment. They explain the similarities and differences between the WAP 1.x and WAP 2.x architectures, and depict conceptual components and features of the second-generation architecture. These components—the application environment, protocol framework, security services, and service discovery—are key to successful application and service development. Modularity, an important feature of the second-generation architecture, enables developers to select modules from different components and to create user services that fulfill market demands.

Introduction

Most Web services cannot be used directly in a mobile terminal since they have been designed for larger screens and for use that is incompatible with the mobile terminal. Despite these limitations, mobile terminals are positioned to become devices of ubiquitous information access. In Japan, for example, NTT DoCoMo's i-mode service is used by some 25 million subscribers. And this number is growing by approximately 70,000 users per day. Similarly, in Europe, the wireless application protocol (WAP) has been a huge success among developers. Indeed, in some countries, it is becoming difficult to find major websites that do not contain pages in the wireless markup language (WML), and new services are being launched daily.

Until now, WAP and i-mode have not been compatible. The markup in i-mode is based on an older version of HTML, whereas WAP is based on the new XML data format. The protocols used are totally different—for instance, i-mode uses optimized versions of the protocols used on the Web, and that does not fulfill the design constraints for WAP.

To make life easier for developers, NTT DoCoMo has been working with the WAP Forum to develop the new version of WAP, which will contain the best features from i-mode and WAP, while maintaining compatibility with future standards and the installed WAP base, and exploit features from the next generation of the World Wide Web (WWW). The result, which NTT DoCoMo claims they will use in coming versions of i-mode, is WAP 2.x (WAP Release 2001).

A goal of WAP 2.x initiatives has been to use the wireless application protocol to fully enable the mobile Internet while avoiding problems such as the "worldwide wait." The WAP Forum has become the most important source for collecting feedback on specifications for the mobile Internet and their implementation. As one of the founders, and chairing several important working groups, Ericsson is also playing a key role in the WAP Forum.

Note: Ericsson Review previously discussed hypertext transfer protocol (HTTP) and transport services in the WAP 1.x architecture. However, because the Internet Engineering Task Force (IETF) is still working with the wireless optimization of existing transport protocols, these topics are not discussed in this article.

BOX A, TERMS AND ABBREVIATIONS

CC/PP	Composite capabilities/ preferences profile	SyncML	Synchronization markup language
CSD	Circuit-switched data	TCP	Transport control protocol
CSS	Cascading style sheet	TLS	Transport layer security (formerly secure socket layer)
DNS	Domain name server	UAPROF	User agent profile
ECMAscript	Formerly Javascript	UDP	User datagram protocol
EFI	Extended functionality interface	vCal	Calendar interoperability standard by the Internet Mail Consortium, superseded by iCal
HTML	Hypertext markup language	vCard	Business card interoperability standard by the Internet Mail consortium
HTTP	Hypertext transfer protocol	W3C	World Wide Web Consortium
iCal	Calendar interoperability standard established by the Internet Mail Consortium	WAE	WAP application environment
IETF	Internet Engineering Task Force	WAP	Wireless application protocol
IP	Internet protocol	WIM	Wireless identity module
IPv4	IP version 4	WML	Wireless markup language
IPv6	IP version 6	WSP	Wireless session protocol
MMS	Multimedia messaging service	WTA	Wireless telephony application
OSI	Open systems interconnection	WTLS	Wireless TLS
OTA	Over the air	WTP	Wireless transaction protocol
PKI	Public key infrastructure	XHTML	Extensible HTML
RDF	Resource description framework	XML	Extensible markup language
RSVP	Resource reservation protocol		
RTP	Real-time transport protocol		
SMTP	Simple mail transfer protocol		

WAP 2.x

The WAP 1.x architecture consisted of the origin server, gateway, and user-terminal environment. The server could be a WAP or HTTP server; the gateway translated the protocol layer and application information. By contrast, the WAP 2.x architecture consists of four conceptual components, namely the

- application environment;
- protocol framework;
- security services; and
- service discovery.

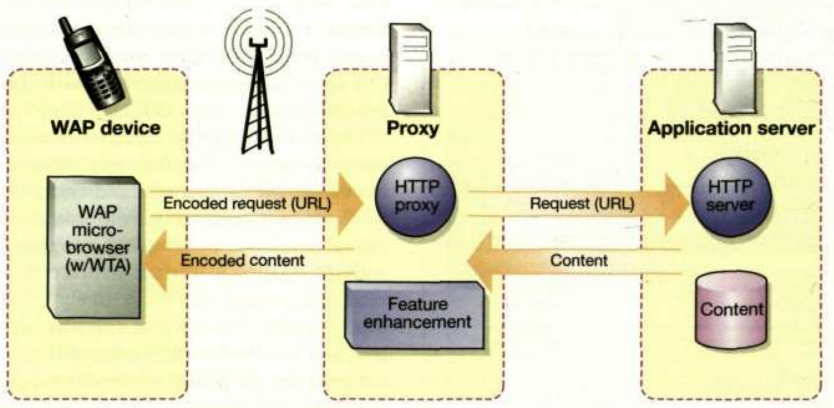


Figure 1
The WAP 2.x architecture supports the use of proxies which enhance the capabilities of the server.

The WAP 2.x architecture does not have strict divisions between the server, gateway, and user-terminal environment. And there is no longer any intermingling between transport and service. Instead, functions—which are accessed via the Internet—can be outsourced to capability servers in a WAP network that implements support for, say, a

- wireless telephony application (WTA);
- public-key infrastructure (PKI) portal;
- provisioning server; and
- user-agent profile (UAPROF) repository.

Communication from WAP clients can take place directly with the server, but it will most likely take place through a proxy. Proxies are being established as one of the main points of control (for example, through firewalls) and as central points for resource interconnection. WAP clients support a proxy-selection mechanism that allows them to choose the most appropriate proxy for a specific task. This extends the current Internet proxy model.

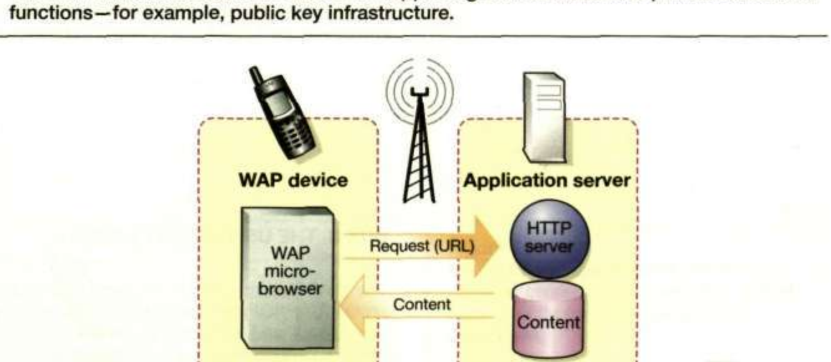
The WAP 2.x protocol is compatible with WAP 1.x, but it relies more extensively on standard Web protocols (such as HTTP) and formats (such as XHTML). WAP 2.x also clearly separates the

- bearer (CSD, GPRS, IMT-2000);
- transport (WDP, TCP);
- session layer (cookies, CC/PP); and
- applications.

Most protocol services in the WAP 1.x suite are also available in new Web protocols. But the WAP push service cannot be realized through existing Web protocols without significant changes to the current Web architecture. Both the WAP 1.x stack and Internet protocols (such as hypertext and multimedia transfer services) can provide some services, but only WAP is capable of providing others, such as the WAP push service.

From the start, WAP has been based on the browsing paradigm made popular by the Web: it adapts technologies from the Web, making them work better in wireless networks. In the WAP 2.x architecture, more of the Web technologies are adopted directly (as they are) rather than adapted, and then extended with WAP-specific functions. These functions enable service

Figure 2
The WAP 2.x architecture also allows for supporting servers which can provide additional functions—for example, public key infrastructure.



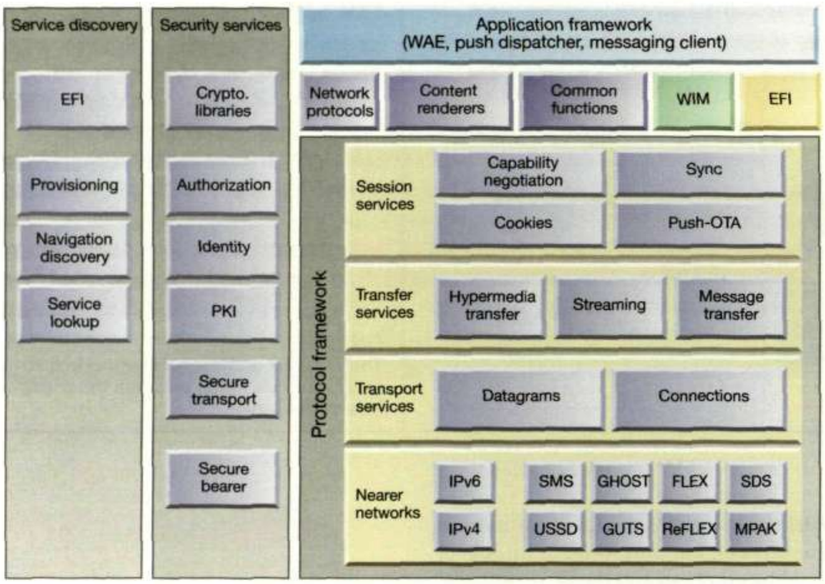


Figure 3
The WAP 2.x architecture is a layered architecture that includes different functions within its framework.

providers and developers of systems, content providers, and devices to provide users with greater added value.

Modularity is one of the main features of the second-generation architecture (modules interact through well-defined interfaces). The security-service and service-discovery components of the architecture span every layer of the open systems interconnection (OSI) model. The application environment component resides on top of OSI layer 7; the protocol framework comprises everything from OSI layer 2 to 7.

The architecture allows components to interact. Developers can thus select modules from different components and create new user services. Conceivably, a minimal device can be developed by selecting components with the smallest footprints. In practice, devices and proxies will most likely implement either a dual stack or only the Internet stack. Backward compatibility is achieved by providing continued support for WML 1 in the client. The WAP conformance profiles (Box B) determine the configuration of the devices as well as how they work together.

The application environment component

The application-environment component enables the following services:

- WAP application environment (WAE)—that is, the browser, calendar agent, and other user agents;
- user-agent profile;
- multimedia messaging and other data formats; and
- push service.

The application environment provides the user interface and other functions that display content. Because it is a flexible environment, modules can be added on an *ad hoc* basis (optional) or through the WAP Forum's specification development process.

BOX B, THE USER-AGENT PROFILE

The WAP Forum defined the term *user-agent profile* (UAPROF) for use within the composite capabilities and preferences profile (CC/PP), which is used to describe the capabilities of the user terminal application environment. The user-agent profile is a data format. A specific set of properties and values describes each terminal. The WAP Forum standardized the property names and values as part of the UAPROF vocabulary.

The CC/PP is defined in an XML framework—called the resource description framework (RDF)—which enables users to connect a property to an object (the CC/PP is an application of the RDF). The resource description framework can be used for annotations, metadata, and profiles that describe users or their terminals. By knowing the information display capabilities of a terminal, the server can create a display that is optimized for that terminal. Including profile information with the request minimizes the number of transactions needed to optimize the information, and it can be cached in a proxy or retrieved from a repository that the device manufacturer maintains. This minimizes the amount of information transmitted over the air and speeds up information access. Designers can create pages or page templates to be used with database servers (such as the Ericsson WAP application server), displaying them in formats that are adapted to user devices.

WAP application environment

The WAP application environment is in the mobile terminal. It contains a subset of XHTML (for display formatting) and a subset of the cascading style sheet (CSS) language (for content formatting). It also contains user agents for WTA and programming interfaces for use in mobile devices. WML and WMLscript execute in the WAP application environment.

Cascading style sheets

The WAP 2.x architecture contains a subset of the cascading style sheet language, which is the most widely used display language on the Web. Using cascading style sheets, an author can define how each element in a document is to be displayed. This gives authors greater control—compared to when the display is specified inside the markup. A style sheet need only be downloaded once from the network server. After that it can be retrieved from a local cache.

Cascading style sheets can adapt automatically to the capabilities declared by a device's user-agent profile. This is particularly important because display capabilities vary significantly among devices. A format that looks good on one device might be displayed differently on another. The user-agent profile ensures that the device gets the most appropriate style sheet. And because style sheets separate display from content, authors can use the same WML document for many different devices with significantly different display capabilities.

Contact and calendar information

WAP 1.x versions contained the vCal and vCard data types, which are not part of the browsing environment. The Internet Mail Consortium standardized vCal and vCard as structured data types for displaying contact and calendar information. iCal, which was developed from vCal, is used in products such as Microsoft Outlook and Lotus Organizer. It is also used in Ericsson's AirCalendar, which allows users to synchronize the electronic calendars they use in the fixed environment with the calendars on their mobile terminals. WAP also accommodates other data types, such as audio and video.

Multimedia messaging

Many of the functions of the new WAP environment are available in existing Internet architectures, but the WAP 2.x application environment contains two modules that were developed in WAP 1.x. These two modules contain functions that are not available in other systems: multimedia messaging service (MMS) was one of the highlights of this year's GSM-UMTS Forum, and push services are not possible on the Web using standard HTTP. MMS is an e-mail-like mechanism for the transmission of multimedia messages (electronic postcards with sound), which are expected to become very popular applications, especially in third-generation mobile systems.

Push services

Service providers use push services to send information to users (who need not initiate any action). As simple as it might sound, the push-service architecture has been a major item on the WAP Forum agenda. In the WAP 2.x architecture, the push service has been divided into the user-agent module and the session-layer module.

The protocol framework component

The WAP application environment relies on a protocol framework component that enables the functions needed to provide the services described above. The protocol framework consists of four modular layers, which can be combined:

- the session service layer;
- the transfer service layer;
- the transport service layer; and
- the bearer service layer.

In traditional Internet environments, the protocol framework solely provides transport services for applications, such as hypermedia transport (HTTP), streaming

BOX C, XHTML

XHTML is HTML reformulated as an XML application, defined by the W3C. For WAP 2.0, the WAP Forum has defined a strict subset of XHTML called the *XHTML mobile profile*. The profile is in turn a superset of the W3C *XHTML basic profile*, a W3C recommendation (December 2000) for XHTML in small devices. An XHTML page, written in XHTML mobile profile or XHTML basic, can be viewed in a WAP 2.0 browser or any standard Web browser. All basic XHTML features are supported, including text, images, links, checkboxes, radio buttons, headings, horizontal rules, and lists.

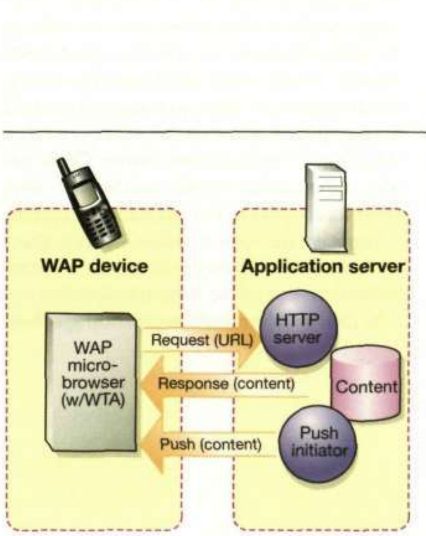


Figure 4
The WAP 2.x architecture allows the application server and the client to connect directly for push and content responses.

(RSVP and RTP Internet protocols), and message transport (standard Internet protocols, such as SMTP). In the WAP architecture, a logical layer has been added: the session service layer.

Session service layer

In the WAP 2.x architecture, the session service layer, which resides between the transport layer and the application environment, brings several new services to applications. Sessions do not exist in HTTP, but cookies can provide sessions. Cookies, which are database markers included in the request and looked up on the server side to identify the user, are part of the WAP 2.x architecture. They enable the reuse of mechanisms that already exist in the Internet and solely give an indication about the relationship between a single server and the user agent. Cookies cannot be used as a general source of information.

The session service layer also includes a technology for reporting to the server. The reports contain information on terminal capabilities and on the terminal application environment. This information is used to optimize the display format.

Data synchronization

Synchronization is another new service in WAP. The WAP Forum has been working with SyncML (another industry consortium) to create a language for data synchronization. Synchronization of data that has been updated in mobile and fixed environments can be a thorny issue. Users retrieve data from the network and store it on a mobile device, which they use to access and manipulate the local copy of the data. Periodically, users reconnect to the network to send changes to the networked data repository. Users also have the opportunity to learn about updates made to the networked data while their terminal was offline. Occasionally, users need to resolve conflicts between their local updates and the networked data. This reconciliation operation (during which updates are exchanged and conflicts are resolved) is known as data synchronization. The data synchronization protocol synchronizes networked data with that on many different devices, including handheld computers, mobile phones, automotive computers, and desktop PCs.

Push sessions

The push solution also contains a session component. The push over-the-air (OTA) session service enables the establishment of push sessions

- across communication links that might not be persistent; and
- in instances when addresses are dynamically assigned.

There is no binding between the transport of data and the session on the Web. The data transport is transparent to the session. Once a hypertext transport transaction is finished, the state it created disappears. In the WAP 1.x stack, the wireless session protocol (WSP) and wireless transaction protocol (WTP) can be used in combination to create and maintain a state, and through it, sessions. This has several advantages (for example, to enable push). By including HTTP as a transport method, the WAP Forum now enables both stateful and stateless transport. Session services provide a "memory" of previous transactions (this feature does not exist in HTTP, since it is a stateless protocol) that enable the retention of terminal characteristics and make for faster initialization of complex transports (such as data streaming).

Apart from the transport of text documents, the WAP architecture has also been prepared for the next generation of messaging. It contains a multimedia transport mechanism for asynchronous message transport (messages are encapsulated for transmission between multimedia and WAP servers in a WAP-specific protocol). The data-transport mechanisms also include IETF data-streaming formats.

Transport service layer

The transport services in WAP 2.x are either datagrams (connectionless service) or connections. Datagrams are more efficient for services that are not dependent on a persistent connection. The datagrams comply with either the user datagram protocol

BOX D, ACQUIRING WAP CERTIFICATION

While some modules are mandatory, a device designer can select and implement desired modules and still satisfy prerequisites for WAP certification. A set of conformance profiles for WAP 2.x determines what a WAP device is, and which components it should make use of. This also helps users to make their choice of devices, since they can easily match the conformance profile with the functions they need.

WAP certification will be increasingly important, since many implementations of technologies developed in the IETF or W3C do not follow the specification, and in practice, implement only a minimal set of functions. This will not happen in a WAP environment whose rigorous certification process determines conformity standards.

(UDP), which is used on the Internet, or the wireless datagram protocol, which was defined for the WAP 1.x architecture.

The connection-oriented aspects of the new architecture are handled by the transmission control protocol (TCP). TCP, however, does not work well over mobile networks, so the WAP Forum is discussing an optimized mobile profile, to enable the mobile terminal to function optimally over the mobile network with its special characteristics.

Bearer service layer

In WAP 2.x, the bearer services have been extended considerably. They now include the mobile radio bearers used to transport WAP (such as SMS, FLEX, USSD, and GUTS), as well as IP version 4 (IPv4) and IP version 6 (IPv6).

WAP can be transported over different networks, and mapping can be handled directly from the WAP stack to several bearer services. WAP 1.x contains several modules for bearer networks, some of which (broadcast networks, for example) could not be handled using TCP transport. In WAP 2.x, bearers can be managed by the IP stack or directly by the WAP datagram or connection service, which uses the Internet's transmission control protocol.

Security services component

The security services component is positioned orthogonally to data-transfer and data-use services within the protocol framework component.

Security on the Internet is currently a hot issue, and the WAP 1.x architecture has received a lot of criticism. The telecommunications industry has been a leader in the security area for a long time, and this experience has been transferred into the WAP 2.x architecture.

WAP security services span all layers of the WAP 2.x architecture, thus creating opportunities for users to set up extremely secure environments (in fact, much more so than what is currently possible on the Web). How? By combining application-layer, transfer-layer, transport-layer, and bearer-layer security—the possibilities are endless.

Security services include

- mechanisms for signing and encrypting data as a WMLScript crypto library;
- authentication services;
- an identification service that uses the wireless identity module (WIM);
- a PKI system;
- transport layer security (TLS, previously called SSL); and
- WTLS, the WAP 1.x-adapted version of TLS.

Service discovery component

The service discovery component is another orthogonal component in the WAP 2.x architecture that embraces what is available on the Internet and extends it by adding mobile-specific components.

One example, the service lookup protocol, uses the existing domain name server (DNS) from the Internet. Terminal functionality is extended through the extended functionality interface (EFI), which enables a WAP device to have external entities attached to it (thermometers, pressure gauges, and so on).

Provisioning, which is another telecommunications-specific protocol, is translated into WAP. Devices can thus be provided with all the parameters they need to function through the network.

Navigation discovery allows a client to discover services in the network—for example, a client might need to find a proxy in order to download data.

Conclusion

The WAP environment is a consistent architecture composed of standard components—taken from the Internet where available, or constructed separately. They enable a consistent application environment for the mobile system. Thanks to the certification process of the WAP Forum, interoperability is also ensured for all the different functions of the WAP 2.x architecture. Using WAP, the developer gets a well-known development environment, and does not have to create several different versions of the application for mutually incompatible environments.

RF multicarrier amplifier for third-generation systems

Bo Berglund, Thorsten Nygren and Karl-Gösta Sahlman

Multicarrier transmitters traditionally have one power amplifier per carrier and a combiner circuit at the output, or they combine the carriers at low power level and then utilize a common multicarrier power amplifier (MCPA) for power generation. This latter solution seems to be simpler, but the need for high linearity—to avoid intermodulation distortion—makes the multicarrier amplifier more complex.

The modulation scheme in third-generation (WCDMA) mobile systems involves a high degree of amplitude modulation, which requires a certain amount of linearization even for single-carrier amplifiers. Notwithstanding, the high capacity and flexibility of the multicarrier amplifier makes it a logical choice for third-generation systems.

The authors describe the technical challenges and Ericsson's solution for building the MCPA for WCDMA.

BOX A, TERMS AND ABBREVIATIONS

3GPP	Third-generation Partnership Project
ACLR	Adjacent channel level ratio
ACP	Adjacent channel power
ADC	Analog-to-digital signal converter
DAC	Digital-to-analog signal converter
DC	Direct current
DSP	Digital signal processor
EPA	Error power amplifier
FF	Feed-forward
GaAs	Gallium-arsenide
LD-MOS	Lateral-diffusion metal-oxide semiconductor
MCPA	Multicarrier power amplifier
MPA	Main power amplifier
MTBF	Mean time between failures
RBS	Radio base station
RF	Radio frequency
Si	Silicon
TDMA	Time-division multiple access
WCDMA	Wideband code-division multiple access

RF power amplifier techniques

The composite RF signal is amplified and fed to the antenna via a transmitter bandpass filter (Figure 1). The coverage and capacity of the base station define the required amount of output power. Several transistor stages connected in series constitute the gain, and the final stage delivers the output power. High-power amplifiers often have multiple transistors connected in parallel at the output. Consequently, the output stage consumes large amounts of power. Silicon (Si) or gallium-arsenide (GaAs) field-effect devices are adequate choices at an operating frequency of 2 GHz.

The low-power stages at the input side are biased in a linear mode called class A. The high-power transistors operate in a more efficient but less linear mode called class AB. As seen from the input-output characteristics in Figure 2, class A mode is more linear for small signals.

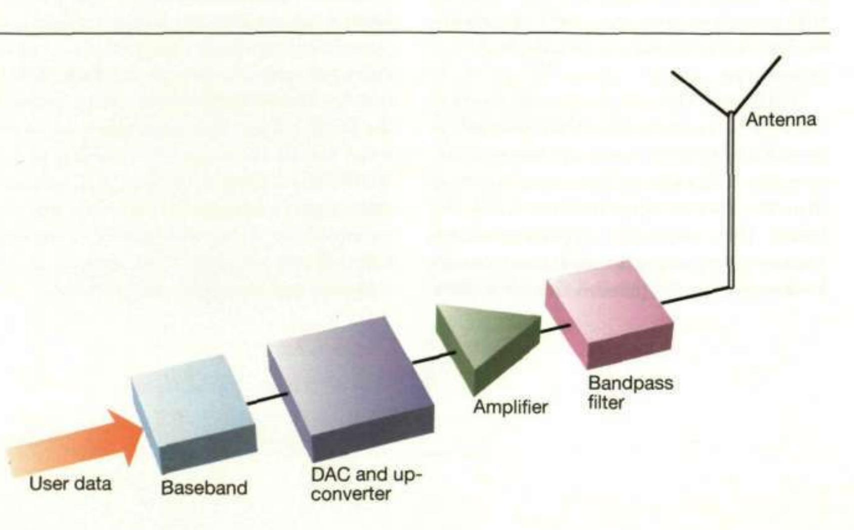
The complex modulation schemes in WCDMA systems have a high peak-to-average power relationship, which produces amplitude and phase distortion in non-linear amplifiers. This distortion is even more pronounced as the output power level of the amplifier approaches saturation. Some amplifiers also experience a memory effect—that is, the output signal at a given moment is affected not only by the instantaneous input signal, but also by the previous signal history. Distortion causes signal quality to deteriorate and adjacent channel power (ACP) to increase; it also gives rise to spurious emissions.

Amplifier efficiency is dependent on the characteristics of the power transistors and the bias scheme. Efficiency improves with higher output power, but distortion increases rapidly as the power level approaches saturation. The high peak-to-average power relationship is thus fraught with compromise.

Figure 3 shows a simulated example of how a two-carrier WCDMA signal is distorted by the gain transfer characteristic of the class AB amplifier. The figure shows the frequency spectrum before and after amplification. The output signal does not satisfy system requirements for spurious emissions.

The best solution for multicarrier amplifiers is to operate the power transistors in efficient but slightly non-linear class AB

Figure 1
Transmitter power amplifier in radio base station (RBS).



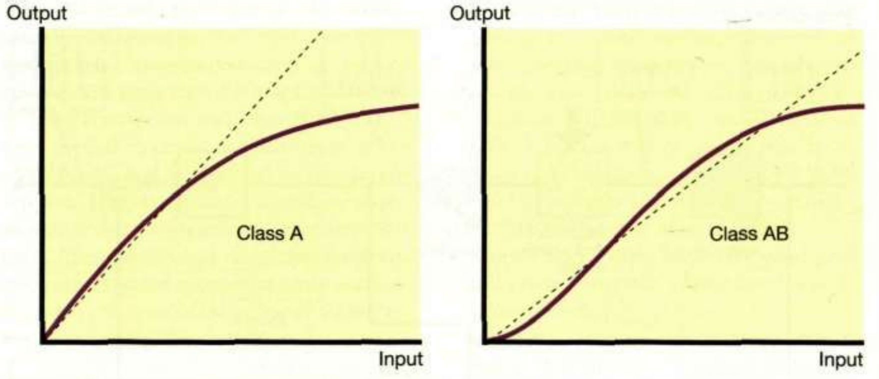


Figure 2
Input-output characteristics for transistor amplifiers.

mode, which will need a linearizing technique that complies with system requirements for spurious emissions and adjacent channel power.

First generation MCPA for WCDMA

Technical challenges

The MCPA is part of the RBS 3000 for WCDMA systems.[1] The primary technical goals in developing MCPAs for WCDMA have been to achieve

- high linearity—to satisfy requirements set by the Third-generation Partnership Project (3GPP); and
- optimum efficiency—to reduce the power consumption of the base station.

Another goal was automated, large-volume production of MCPAs at different facilities around the world.

Stringent 3GPP requirements for adjacent channel power, the spectrum mask, and spurious emissions make it necessary to linearize the power amplifier. A minimum requirement of four-carrier operation (WCDMA) stipulates a bandwidth of at least 20 MHz. For detailed WCDMA radio requirements, see 3GPP TS 25.104.[2]

MCPA technology for WCDMA

Selection of technology

Feed-forward technology was selected as the main linearization method for the first-generation MCPA for WCDMA, since it can yield the necessary linearity and bandwidth. It is a mature linearization method that can achieve good linearity over fairly large band-

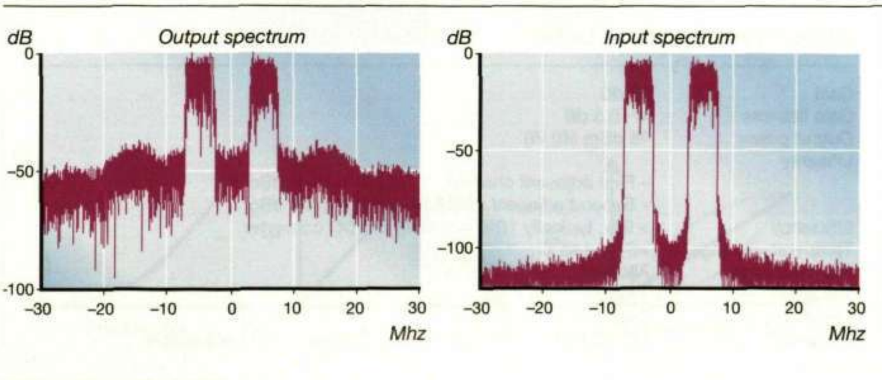


Figure 3
A simulated two-carrier WCDMA signal spectrum at output and input of a non-linear power amplifier.

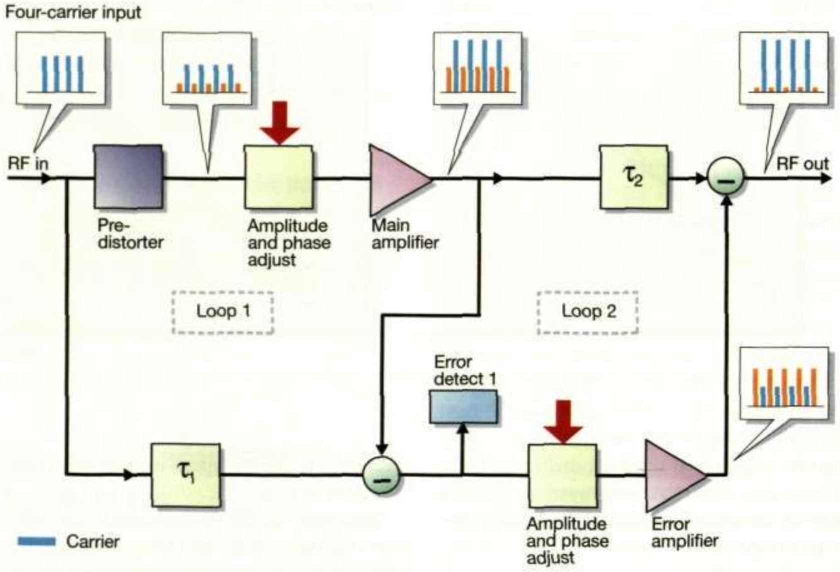


Figure 4
Block diagram of a feed-forward MCPA for WCDMA.

widths. Feed-forward MCPAs for TDMA systems, which have been on the market for some years, have bandwidths of up to 20 MHz and a spurious emission level of around –60 dBc. A drawback to the feed-forward technology has been low efficiency—approximately 6-7%. Therefore, to yield greater efficiency, and to be suitable for high-volume production, the technology had to be improved.

Feed-forward concept

Figure 4 shows a block diagram of the feed-forward architecture. The input signal is split into two paths. The signal in the top path is amplified by the main power amplifier (MPA), which operates in class AB. The non-linearities in the MPA result in intermodulation distortion, which adds to the original signal. A sample of the MPA output signal is fed to the subtracter—a directional coupler—where the signal is subtracted from a delayed portion of the original signal (present in the lower part). This results in an error signal that contains the distortion signal. Ideally none of the original signal energy should remain. But in reality it is possible to suppress the carriers

BOX B, MCPA TECHNICAL DATA

Gain	52 dB
Gain flatness	+/- 0.5 dB
Output power	46 dBm (40 W)
Linearity	
	– First adjacent channel ACLR 1 < -51 dBc
	– Second adjacent channel ACLR 2 < -60 dBc
Efficiency	> 9%, typically 10% including DC/DC converter
Temperature range	+5° to +45°C
Hot-swapping	Allowed
Volume	7 liters
Weight	7 kg
Bandwidth	Any 20 MHz band within 2110 – 2170 MHz, field-adjustable

with 25–30 dB. In Figure 4, the carrier-cancellation operation is marked Loop 1. The feed-forward cancellation loop is marked Loop 2. The error signal is amplified linearly in an error power amplifier (EPA)—to a level needed to cancel the distortion in the main path—and is then fed to the output coupler. The MPA output signal is delayed to match delay through the error amplifier path. The contributed distortion from the two paths is added in opposite phase and ideally only the amplified original signal remains at the MCPA output.

Analog RF predistortion

The main power amplifier can be improved by employing a predistorter whose transfer characteristic complements that of the main power amplifier. A configuration in which the predistorter and main power amplifier are cascaded ensures that the resulting system has low distortion (Figure 5). The nonlinear predistorting element operates at the final carrier frequency. This method has the advantage of linearizing the entire bandwidth of an amplifier.

Predistortion does not add losses at the carrier output, since distortion is compensated for at the input. Thus amplifier efficiency is not affected. In fact, the efficiency can be increased by driving the MPA closer to compression with the same intermodulation level.

Product technology

Several design challenges had to be overcome in order to achieve the function as described:

- Because the MPA must have inherently good linearity and bandwidth as well as corresponding efficiency, a class AB amplifier was designed using lateral-diffusion metal-oxide semiconductor (LD-MOS) transistors. With this technology, at adjacent channel level ratio (ACLR) values of approximately –40 dBc, the MPA efficiency is around 20%.
- Loops 1 and 2 must have very good gain and phase flatness to ensure good cancellation in each loop. Flatness is tuned electronically during production, which means that manual trimming is not required.
- Adaptive loop control ensures good loop stability. A digital signal processor (DSP) is used for this control.
- The delay (τ_2) in Loop 2 must have low loss to reduce the amount of output power lost as heat. Excessive loss reduces MCPA efficiency. To minimize loss, the delay element has been implemented as a cavity bandpass filter.
- To cancel distortion, the EPA must have a bandwidth three to five times greater than the MCPA signal bandwidth in order.

A combination of good linearity and good efficiency is needed for the EPA, which must amplify distortion from the MPA without introducing further distortion. Since good efficiency is one of the primary design goals, power consumption in the EPA must be minimized.

In classical MCPA designs, the EPA is a very linear class A amplifier (to achieve superior linearity). Nonetheless, advances in

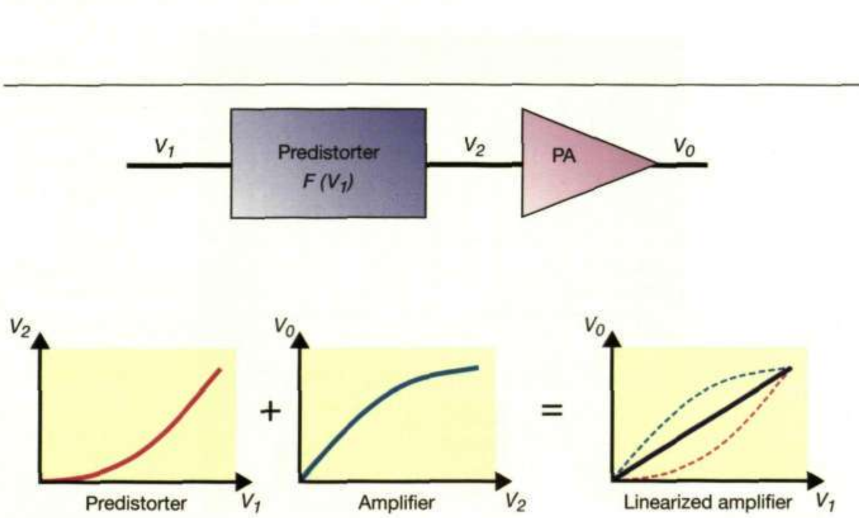


Figure 5
The predistortion concept.

Figure 6
Ericsson's MCPA for WCDMA.

LD-MOS transistor technology have made it possible to design a class AB amplifier with considerably lower power consumption.

For optimum efficiency, an analog predistorter has been added in front of the MPA. The predistorter improves MPA linearity by 5 to 10 dB. The linearity enhancement allows for more output power from the MPA with improved efficiency.

The final MCPA design (Figure 6)—including the DC/DC converter—yields an efficiency of 10%.

Measurements

Figure 7 shows a typical measurement with two WCDMA carriers at total average output power of 40W/46 dBm. The distance between the center frequencies of the two carriers is 10 MHz.

The adjacent channel level ratio and spurious emission levels are far below the maximum values allowed.

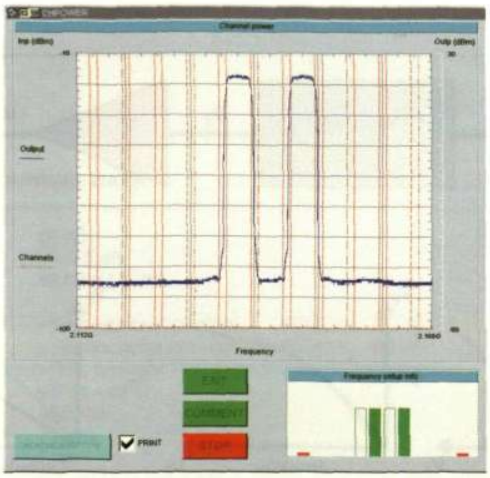
Future MCPA technologies

Present-day RF power linearization techniques employ the feed-forward technique and variants thereof. To further enhance performance, attempts are being made to improve amplifier linearity by predistorting the signal to a power amplifier. Predistorted signals can be generated in either analog or digital techniques.

Digital linearization techniques

Given that semiconductor technology has improved DSP, ADC and DAC techniques, steps have been taken to design an MCPA based solely on predistortion in the digital domain. This approach makes use of modern signal-processing techniques and promises to be cost-effective. A down-converted sample of the RF output signal is compared to the digital input signal. The

Figure 7
Measured output power spectrum of the MCPA.



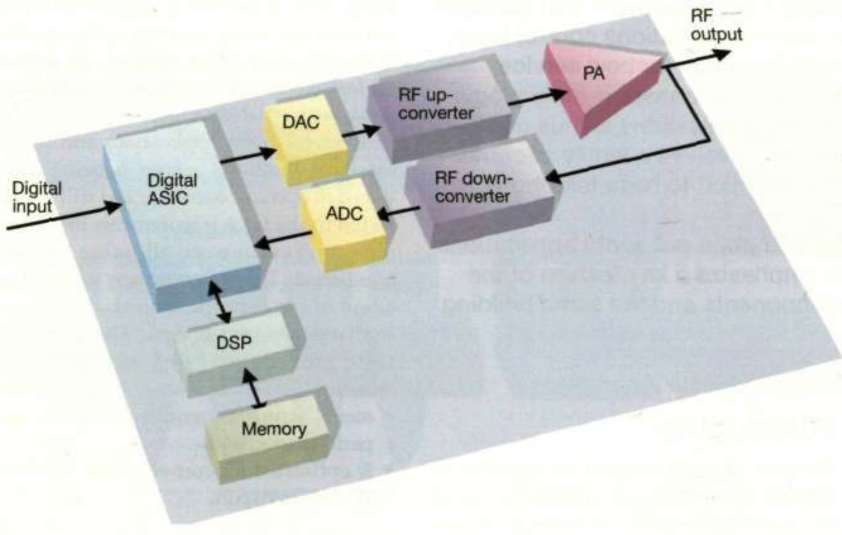


Figure 8
Digital predistortion MCPA for WCDMA.

difference is minimized by predistorting the input signal in a digital ASIC controlled by a DSP for adaptive update. Figure 8 shows the basics of digital predistortion. Greater efficiency is feasible, since distortion is compensated for at the input of the power amplifier.

Ericsson is conducting intensive research in the area of digital predistortion, and the application is expected to become a mature solution for future WCDMA systems.

Ericsson is also investigating power amplifier concepts (efficiency-enhancement technologies) that will yield greater efficiency in the power amplifier itself.

Conclusion

The growing market for third-generation systems requires a base station RF power amplifier designed for large-volume production. Ericsson has chosen the feed-forward technique with analog predistortion, which gives excellent spurious emission values and high production yield. A design for efficiency and cooling guarantees the best MTBF in a small-size amplifier.

Progress in semiconductor technology and improved power amplifier designs have set the stage for a next-generation MCPA with better efficiency, smaller size, and high MTBF.

REFERENCES

1 Zune, P.: Family of RBS 3000 products for WCDMA systems. Ericsson Review 3/2000, pp. 170-177.
2 Third-generation Partnership Project; Technical Specification Group Radio Access Networks; UTRA (BS) FDD; Radio Transmission and Reception.
3 GPP TS 25.104.

Ericsson's family of carrier-class technologies

Göran Ahlforn and Erik Örnulf

Ericsson is leading the way to a horizontally layered network that consists of a content and user applications layer, a communications control layer, and a connectivity layer. This network sets the stage for new service applications and for the introduction of a common transport technology.

Ericsson's way of developing systems and applications is thus undergoing a transformation: the company strategy involves a family of carrier-class technologies—AXE, TSP and CPP—on which to base telephony, server and access applications.

The authors describe the technologies and point out some applications and products based on them. They also emphasize a key feature of the strategy—the use of common system components and the same building practice for AXE, TSP and CPP.

Introduction

The ever-growing demand for rapid introduction of novel services and the pursuit of the most effective transport technology have led the telecommunications industry to shift to a horizontally layered architecture.[1] An underlying trend—the convergence of telephony, data communication and media technologies—is further accelerating the development of new products and services for introduction into the network. In response to this trend, Ericsson—besides evolving AXE—has introduced two carrier-class technologies: TSP for servers, and CPP for gateways (Figure 1). Each of these meets the fundamental requirements for high availability and has been optimized for its role in the network.

In the layered network architecture, telephony services are being aligned with what is called the server-gateway split (applications and control functionality are separated from connectivity and transport). In accordance with this architecture, services are introduced in the applications and control layers. A family of servers is being introduced in operator networks and will be subjected to the same requirements for carrier-class performance as all other network equipment. The servers occupy a central position in the network, storing and processing large amounts of data. The access-network products are based on a technology which

- meets stringent requirements for high performance; and
- is optimized for cost-effective ATM and IP data transport.

A family of carrier-class technologies

AXE has distinguished itself as the most successful switching system in the market. AXE equipment installed all over the world provides a base for extending networks and evolving them into the new-generation network, which is capable of circuit switching and packet data transport.

The new generation of AXE is designed to function as a complete switching node and as a server. It accommodates all software functionality for existing network solutions and new layered-network solutions.

TSP is Ericsson's choice of carrier-class server technology for all new multimedia applications and control functionality. Designed for server solutions where high availability is required, TSP has been introduced in second-generation networks. Networks built with TSP technology enable operators to configure their nodes by packaging different kinds of functionality on the same mechanical structure. In this way, operators can reduce the amount of site equipment and simplify network operation.

CPP is a packet-switching technology for ATM and IP, introduced in the market as Ericsson's technology for third-generation radio access. Ericsson uses the CPP technology to optimize the access transport solutions in its base station, radio network control and media gateway equipment.

Principal requirements

The principal requirements put on systems and applications based on carrier-class technology are

BOX A, ABBREVIATIONS

AAA	Authentication, authorization and accounting	IP	Internet protocol
ASIC	Application-specific integrated circuit	MGCF	Media gateway control function
ATM	Asynchronous transfer mode	MGW	Media gateway
BSC	Base station controller	MP	Main processor
BTS	Base transceiver station	MSC	Mobile switching center
CDMA	Code-division multiple access	PDSN	Packet data service node
CPP	Cello packet platform	QoS	Quality of service
EM	Element management	RBS	Radio base station
GEM	Generic Ericsson magazine	RNC	Radio network controller
GGSN	Gateway GSN	SGSN	Serving GSN
GMSC	Gateway MSC	STM	Synchronous transport module
GPRS	General packet radio service	TDM	Time-division multiplex
GSM	Global system for mobile communication	TSP	Telecom server platform
GSN	GPRS support node	UMTS	Universal mobile telecommunications system
HLR	Home location register	WCDMA	Wideband CDMA

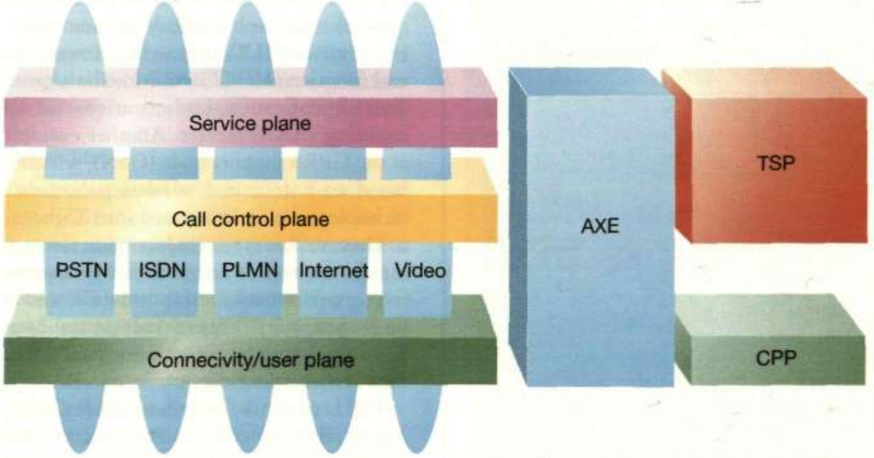


Figure 1
Carrier class technologies in the layered network.

- high availability;
- real-time performance;
- scalable capacity; and
- openness.

AXE

AXE 810 represents the present stage of AXE development with a dramatically reduced footprint (Figure 2). Its innovative design, which incorporates commercially available, standardized components, excels at voice traffic and packet switching.

AXE 810 is being deployed in GSM and UMTS core networks in which AXE switches continue to serve as
- mobile switching centers (MSC);
- base station controllers (BSC); and
- MSC servers connected to media gateways in layered network solutions.[2]

Likewise, in Ericsson's ENGINE solutions for multiservice networks, AXE is employed as a telephony server connected to a gateway AXD 301 ATM switch for packet data. To a large extent, the open architecture of the AXE system is achieved through the use of
- commercially available hardware components;
- standard hardware building practices and buses; and
- commercially available software components and interfaces.

Increased capacity and enhanced in-service performance have been the main focus of attention in the continuing development of the central processor. The newest version, APZ 212 40, is a milepost in hardware im-

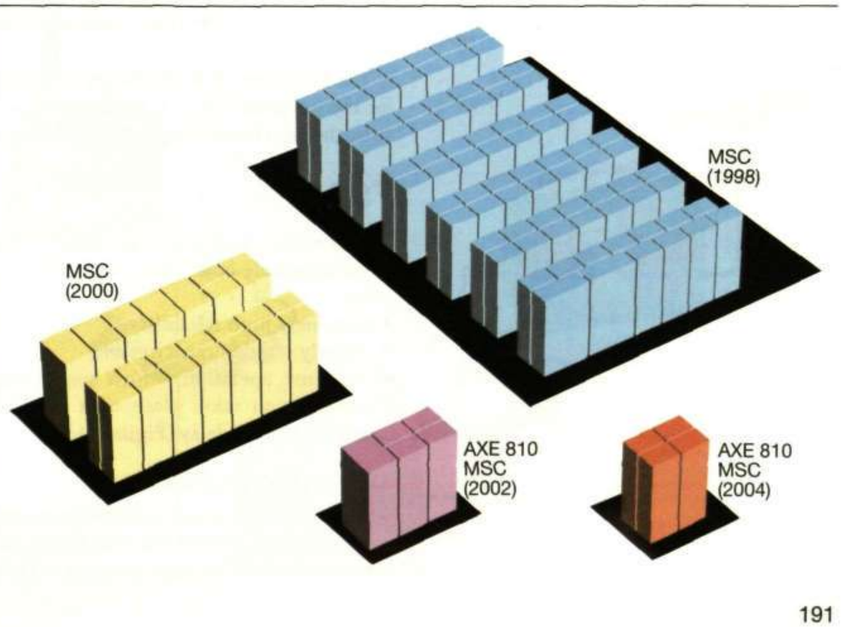
BOX B, CARRIER-CLASS TECHNOLOGY

In Ericsson parlance, *telecom platform* denotes a specific technology and architecture.

A telecom platform, generic in nature and composed of hardware, an operating system, and applications on top, serves as a technological and architectural base for the development and evolution of systems and applications.

In the past few years, the term *platform* has been used somewhat indiscriminately to denote a technology, an architecture, a system, or an application. The term as such is also considered to have too obvious connotations of mechanical engineering and "squareness." That is why, for the purposes of this article, the authors have opted for the synonymous term "carrier-class technology," hoping that it will gain general acceptance when used in its proper context.

Figure 2
Volume reductions for a typical AXE MSC site.



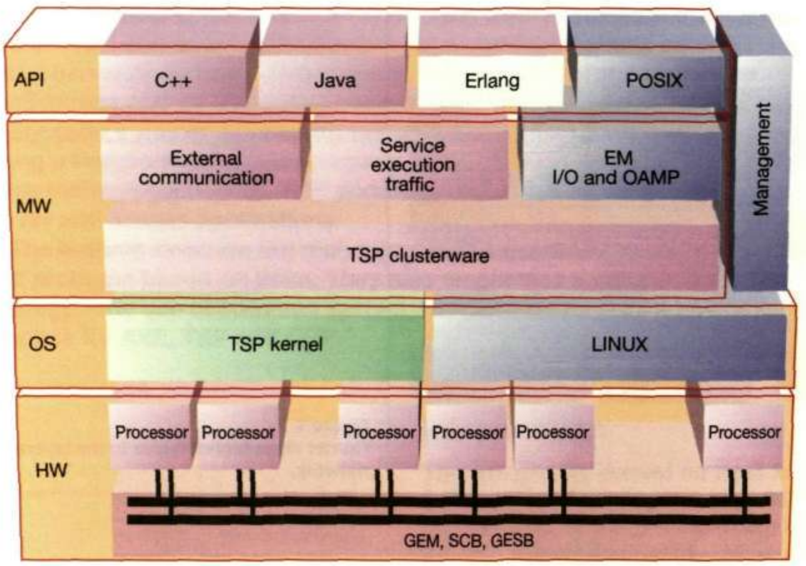


Figure 3
The TSP architecture.

plementation—its processor core is based on industry-standard, high-end microprocessors. The implementation technique reduces time to market, in particular for future upgrades, since they can be based on mainstream microprocessor development. Even so, seen in terms of the application, the APZ 212 40 is fully backward compatible, which means that operator investments made in AXE applications software are secure.

The evolution of AXE enables it to grow for the foreseeable future while continuing to be the basis for all telephony applications.

TSP

TSP is more robust and fault-tolerant than any comparable open server technology. It offers

- extremely high reliability;
- linearly scalable capacity; and
- real-time operation, which means that transmission takes place with minimal and controlled delay (Figure 3).

TSP—a strategic choice

TSP is a strategic choice for server and control nodes that provide the availability and scalability needed by large systems in existing networks and in future IP multimedia networks. Examples include the home location register (HLR), the media gateway control function (MGCF), multimedia servers, and authentication, authorization and accounting (AAA) servers. Another example is the GPRS support node (GSN), which is based on a dedicated wireless packet-data technology being integrated into TSP.

These nodes, typically located in the center of a network, must provide superior in-service performance and must be allowed to grow in capacity. They include central databases that contain essential user data, traffic data and charging data.

TSP is particularly well-suited for building small and medium-sized systems from the same components and functional units. And being highly scalable, it enables operators to expand their networks at the desired rate. Key characteristics of TSP are extremely efficient internal signaling and a unique distribution algorithm.

TSP employs a component-based architecture that consists of several functional units within a framework of open interfaces. This allows the same software to be run on hardware from different suppliers. In principle, TSP uses commercially available components. Only the most essential components for adding value in terms of robustness and scalability are developed in-house. The TelORB software package with TSP clusterware and kernel facilitates these special characteristics.[3] It is capable of executing on different processor architectures and with different operating systems, which makes for easy adaptation to general technical developments.

Linearly scalable capacity

TSP also contains basic functions that are common to several applications, such as a complete and standardized management interface and protocol for network signaling. A cluster of loosely coupled central processors in TSP supports scaling from very small systems to nodes with several dozen processors. This design makes for the creation of extremely powerful server nodes whose capacity can be increased linearly through the addition of new processors (Figure 4).

Development plans include the introduction of Linux in the TSP processor cluster. Together with TelORB, Linux components will guarantee TSP scalability and availability.

The cluster software design guarantees availability. All board components (power and Ethernet ports, for example) are duplicated, and all software and data on each processor board are copied to another board using the TelORB distribution algorithm. In this way, entire nodes can be duplicated.

Reliability

The database used by TelORB is distributed among all the processors in the system. If one processor crashes, any other processor in the system can take over. The same mechanisms are also used for identifying and handling software faults. Thus, there is no risk of a software fault hanging the system: if a fault occurs, the system is automatically restarted in real time.

CPP

Ericsson's CPP carrier-class technology can be used for

- small base stations that contain few radio channels; and
- large control units that employ hundreds of processors and fill dozens of cabinets.[4]

CPP is based on a duplicated packet switch that supports the switching of TDM, ATM and IP traffic with selectable quality of service (QoS) levels. The switch handles user data and all control data in the node as well as to and from other nodes. CPP supports various types of network interface (including E1/J1/T1, STM-1 and Ethernet) for connecting user data, signaling information, and operation and maintenance channels.

The first CPP applications solely switch ATM traffic, but IP is being introduced in CPP to enable access-network products to switch ATM and IP traffic (Figure 5).

Physically, CPP employs several magazines with different types of circuit board including

- switch boards, which use a proprietary ATM switch-core circuit;
- processor boards;
- interface boards;
- transcoders;
- echo cancellers; and
- application boards for multi-stream functions, for example.

Each board is fitted with a device board module that includes communication circuits for the processors and memory. All inter-board communication is handled by the magazine backplane and switches.

Extremely scalable

The traffic load on CPP is distributed among several processors, including main proces-

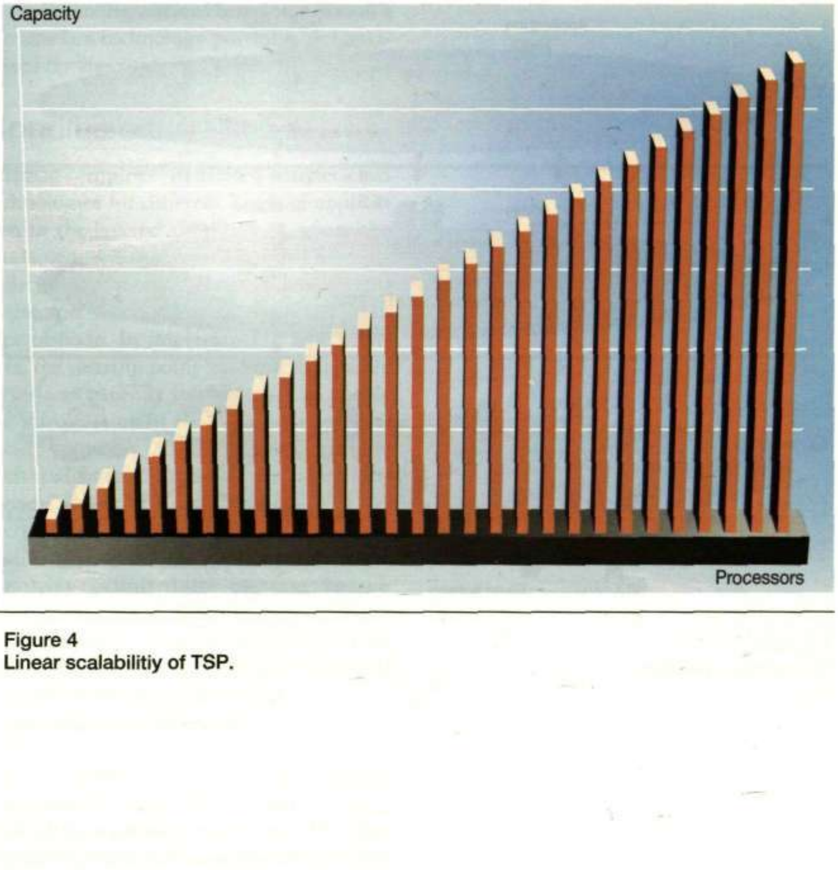
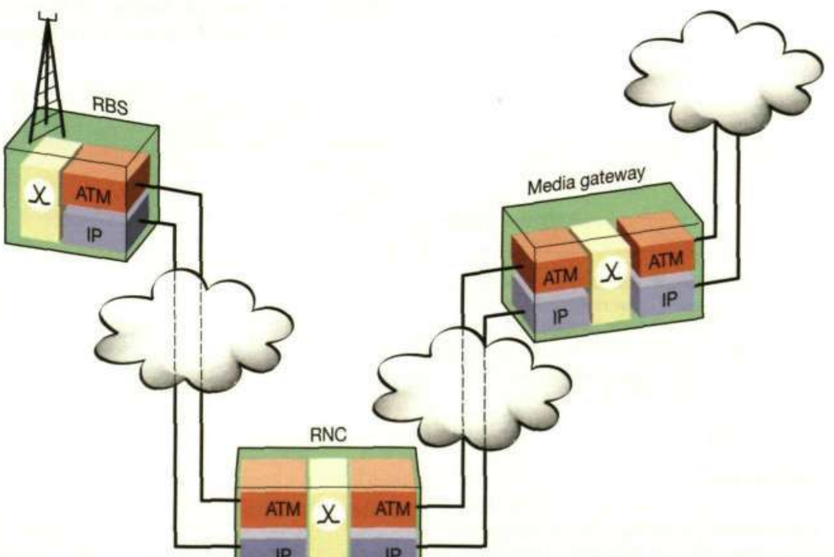


Figure 4
Linear scalabilitiy of TSP.

Figure 5
CPP network nodes.



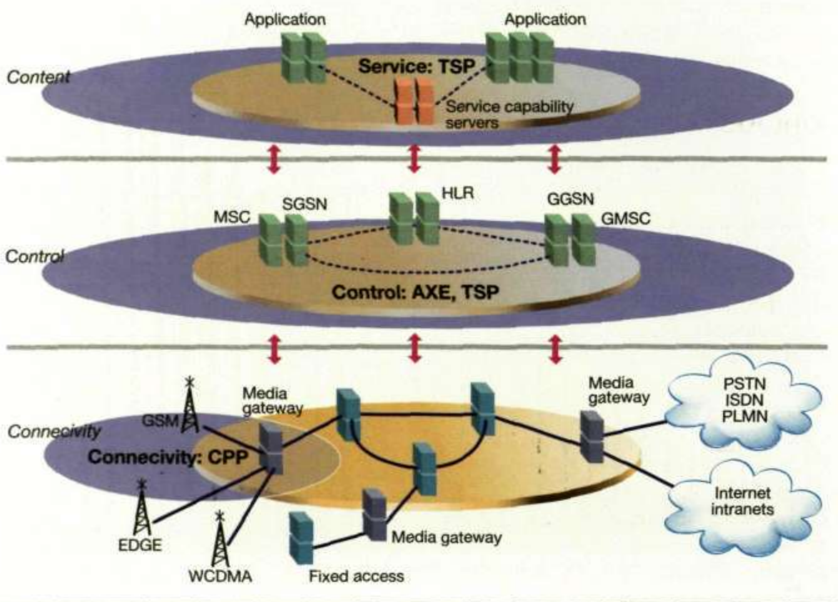


Figure 6
The network solution.

sors (MP) and board processors (BP). Each node must contain at least one MP and one BP on each board. As is the case with TSP, robustness and fault-tolerance are achieved through processors working together in a cluster. The architecture of CPP has been designed to guarantee that essential functions survive hardware faults.

CPP is extremely scalable, both upward, where 30 or more magazines can be combined, and downward, where the limit is one radio channel.

Already in the market

CPP applications include radio base stations (RBS) and radio network controllers (RNC) for WCDMA and media gateway products; likewise, CDMA2000 will use CPP for its radio access and packet data service node (PDSN) solutions. CPP is also used for Ericsson's RXI real-time routers.

Like TSP, CPP is based on an open and dynamic operating system. Apart from its application-specific integrated circuits (ASIC) and switching components, which have been developed in-house, CPP uses commercially available processors and programming languages.

With the introduction of CPP, Ericsson can offer effective ATM and IP transport solutions for access networks.

Common components

As described above, Ericsson has introduced two new carrier-class technologies, TSP and CPP, and is continuing to evolve AXE for future applications in the layered network architecture. Each technology is optimized to support different kinds of application or network node, since each type of node must be adapted to its specific requirements.

A far-reaching improvement is the generic Ericsson magazine (GEM), a high-capacity, flexible and scalable magazine in which several functions can be combined. This means that considerably fewer magazine and board types are needed in each node.

TRADEMARKS

CDMA2000 is a trademark of the Telecommunications Industry Association (TIA).

By moving interfaces to the GEM backplane, designers were able to eliminate many cables, which resulted in an even more compact system.

The GEM concept can be used to build anything from an extremely small node to a very large one, using the same types of board and magazine. As capacity requirements grow, operators can add one or more GEMs with a suitable combination of devices, thereby smoothly extending the node without interruption of traffic.

The use of standard switch boards, such as Ethernet, makes the GEM an open magazine prepared for future developments.

By basing all the technologies on the same building practice, Ericsson can use common mechanical components and common control processors, Ethernet switches and interface boards. At a site installation, this gives operators the advantage of having one cooling system, one power supply system, and uniform alarm system handling.

The protocols and middleware for Ericsson's family of carrier-class technologies are also being based on the same software solutions.

Because all nodes must fulfill the same basic requirements for carrier-class performance, and meet customers' needs of reducing complexity and cost of network operation, AXE, TSP and CPP are based on

- the same building practice;
- the generic Ericsson magazine (GEM); and
- the same element management (EM) solutions.

All hardware and software are being aligned, which means that AXE, TSP and CPP can be seen as configurations based on Ericsson's carrier-class technology portfolio and optimized for their specific purposes.

Conclusion

Ericsson employs dedicated carrier-class technologies for different kinds of application in the layered architecture, where the fundamental requirement for high availability is the same for all network nodes but where specific requirements apply to each type of node. In implementing new products, the starting point has been to separate server and gateway functionality to enable the introduction of new services as well as packet-transport ATM and IP in parallel with traditional circuit-switched telephony services. AXE, TSP and CPP

- comply with carrier-class requirements in all network nodes;
- provide optimized site solutions through the packaging of different types of functionality on the same mechanical structure, which results in fewer nodes and simplified network operation;
- are uniquely scalable; and
- offer support for all transport technologies—TDM, ATM and IP—separate from the applications and control layers of the network.

Ericsson's carrier-class technology facilitates smooth migration all the way from the present-day circuit-switched network, where wideband packet-data services in the form of GPRS, WCDMA and ENGINE solutions are now being introduced, to the interactive multimedia network of the future.

REFERENCES

1 Dahlin, S. and Ornulf, E.: Network evolution the Ericsson way. Ericsson Review Vol. 76(1999):4, pp. 174-181.
2 Enderin, M.; LeCorney, D.: Lindberg, M. and Lundqvist, T.: AXE 10—The evolution continues. Ericsson Review Vol. 78(2001):1, pp. 10-23.
3 Hennert, L. and Larruy, A.: TelORB—The distributed communications operating system. Ericsson Review Vol. 76(1999):3, pp. 156-157.
4 Reinius, J.: Cello—An ATM transport and control platform. Ericsson Review Vol. 76(1999):2, pp. 48-55.

Building a reliable, cost-effective and future-proof fiber optical access network

Per O. Andersson, Ingvar Fröroth and Stefan Nilsson-Gistvik

The uptake of broadband services among residential and small- to medium-sized businesses is being helped by developments designed to relieve capacity bottlenecks in existing copper-based access networks. Likewise, the falling costs of laser equipment, fiber and cable installation are positioning fiber-optic solutions to overcome this capacity hurdle. The functionality, scalability and widespread understanding of Ethernet technology make Ethernet-over-fiber a strong candidate to become the preferred solution for extending optical techniques straight out to the wall socket. By offering symmetric transmission of 100 Mbit/s or better, Ethernet-over-fiber access networks will enable the delivery of advanced services including IP-based television, IP telephony, and robust, secure, LAN-interconnect services. Given current and emerging Ethernet standards, which address issues, such as priority traffic handling, LAN security, and circuit-oriented provisioning, it is likely that flexible, multipurpose access networks will soon be built entirely on Ethernet technology.

In many countries, more than one-third of the population regularly uses the Internet. Besides tremendous residential usage, the market for small and medium-sized enterprises is growing. Notwithstanding, the development of broadband service markets worldwide is dependent on having access networks that live up to the vision of the *Terabit society*. A true end-to-end infrastructure, including first-mile connectivity from the home and office, is needed to experience the full potential of next-generation services.

BOX A, TERMS AND ABBREVIATIONS

ATM	Asynchronous transfer mode
CAT5	Category 5 unshielded twisted-pair cable
CATV	Cable TV
DRG	Ericsson's residential gateway
DSCP	Differentiated services code point
DWDM	Dense wavelength division multiplexing
FSAN	Full-service access network
FTTH	Fiber to the home
IEEE	Institute of Electrical and Electronics Engineers
IP	Internet protocol
IT	Information technology
LAN	Local area network
LCS	Local community system
MAN	Metropolitan area network
OSI	Open system interconnection
PON	Passive optical network
PTT	Postal telegraph and telephone
RSVP	Resource reservation protocol
SDH	Synchronous digital hierarchy
SONET	Synchronous optical network
TOS	Type of service
VoIP	Voice over IP
WAN	Wide area network
xDSL	Digital subscriber line (of any type)

Many large companies already use high-capacity fiber-optic access networks, but small businesses also want to benefit from services, such as e-business, Web-hosting, high-quality video and audio conferencing, and multimedia virtual customer contact centers. Whereas capacity in backbone and metropolitan networks is virtually unlimited, thanks to the widespread penetration of dense wavelength division multiplexing (DWDM) technology, the transmission bottlenecks in the first mile of connectivity are hampering the uptake of these kinds of services.

Fiber is the most effective medium for bridging this gap. Direct fiber connections are an increasingly popular way of delivering broadband services to high-density clusters of users, for example, in business parks and apartment blocks. Fiber has the advantage of almost limitless capacity as fiber-optic networking technology advances. But as yet, availability is limited, and the industry is still awaiting agreement on a common standard.

Although fiber offers perhaps the greatest potential as an access medium, it must be emphasized that all current high-capacity "broadband" access solutions (including xDSL, microwave and cable) have important roles to play, especially in managing a smooth migration from legacy networks to the all-fiber-optic network of the future. The combined force of these broadband technologies will drive market demand and application development. As the services grow and mature, the full strength of fiber-optic solutions (symmetric, secure, and reliable high-capacity connections) will become self-evident.

Sweden—a broadband pioneer

Historically speaking, it has been prohibitively expensive to build a new, single-service, physical network to replace a network already in operation. Often, the high cost has been further exacerbated by a lack of long-term planning that would have helped mitigate short-term commercial considerations.

Sweden is a notable exception, and one of the leading nations in developing broadband access for all. A theoretical model (Figure 1) of a ubiquitous access network set out in a paper from the Swedish ICT-Commission[1], shares much common ground with Ericsson's working example in Hudiksvall, Sweden. The ICT report states:

"The communications architecture used for the Internet (IP architecture) is one of the factors driving the integration of telecommunication, data communication and video communication. All services, whatever their bandwidth requirement, will be carried by IP-based networks.

"The present IT infrastructure has been optimized for telephone services and will for the most part have to be replaced by a new infrastructure that has been adapted to the IP communications architecture and to the estimated annual doubling of traffic which will follow the development of new services and the addition of new users."

The ICT-Commission envisions a future-proof infrastructure in Sweden:

- In 2005, every household in Sweden will have a fixed Internet connection of at least 5 Mbit/s real throughput capacity.
- By 2005, Sweden should have constructed a fine-meshed fiber-optical network available to all households.
- The network is to be technically and competitively neutral and open to all operators. The aim is to give everyone, through free competition, access to high transmission capacity at low cost.

The new national IT infrastructure is to consist of a fine-meshed fiber network that offers redundancy and the ability to distribute traffic load. Several operators should be able to use the IT infrastructure simultaneously—that is, operators will have access to their own fiber pairs, so as to be able to build their own network structures (for example, star or ring networks). Among other things, this means that parallel network structures will be established using the same fiber cable.

The national fiber structure will thus have to be built in the form of a grid between main nodes that constitute centers of local and regional fiber networks. Communication will be based on IP, which means that data can find the best path in a network, assuming that alternative paths are offered. Redundancy is therefore needed between the main nodes in a municipality and between municipalities. A layered network infrastructure (Figure 2) offers a useful model for realizing this vision.

A guiding principle is that the infrastructure must be planned as a general network, and not as a specialized network based on rigid technical solutions or built for a limited number of operators. A general conduit network, containing redundant fiber, can be planned as the basis for carrying competing operators' active networks.

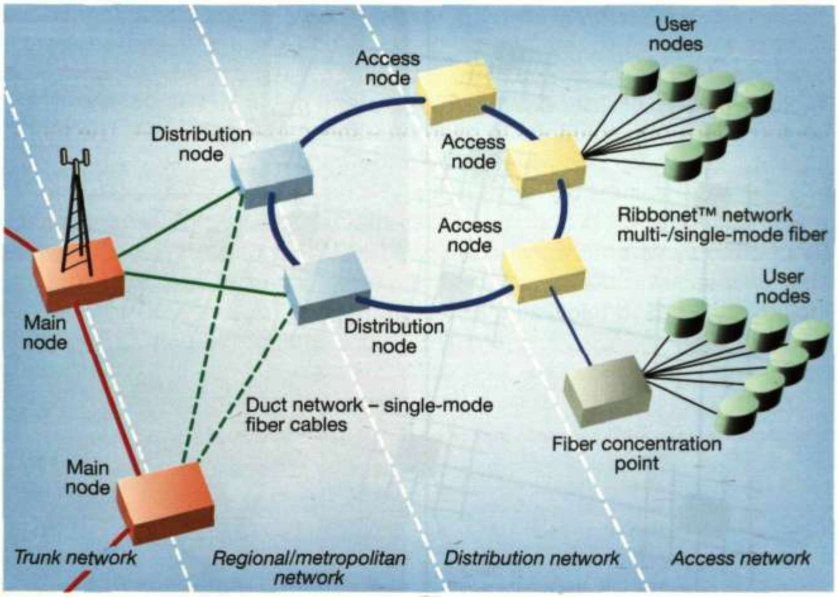
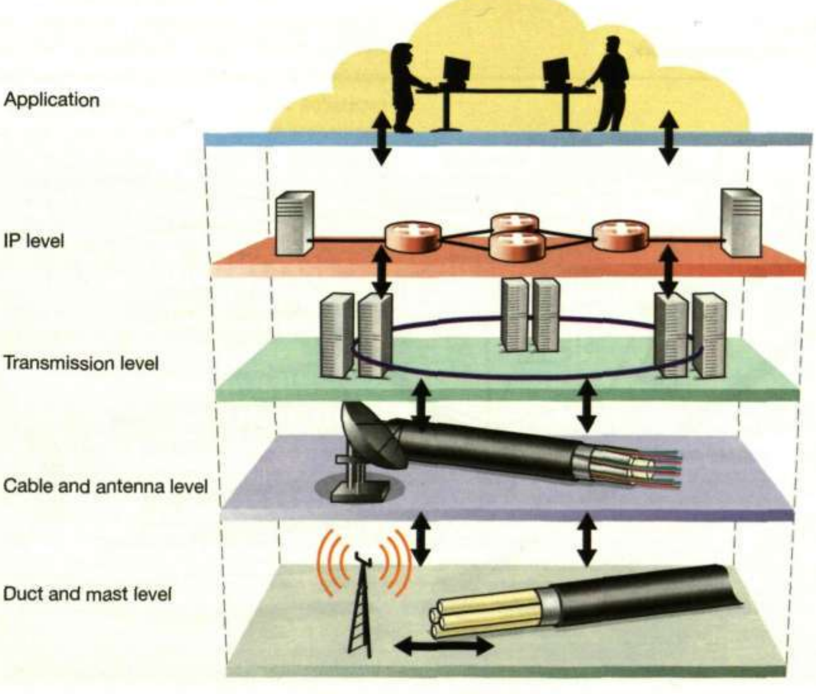


Fig 1
The proposed IT infrastructure.

Figure 2
The levels of the IT infrastructure.



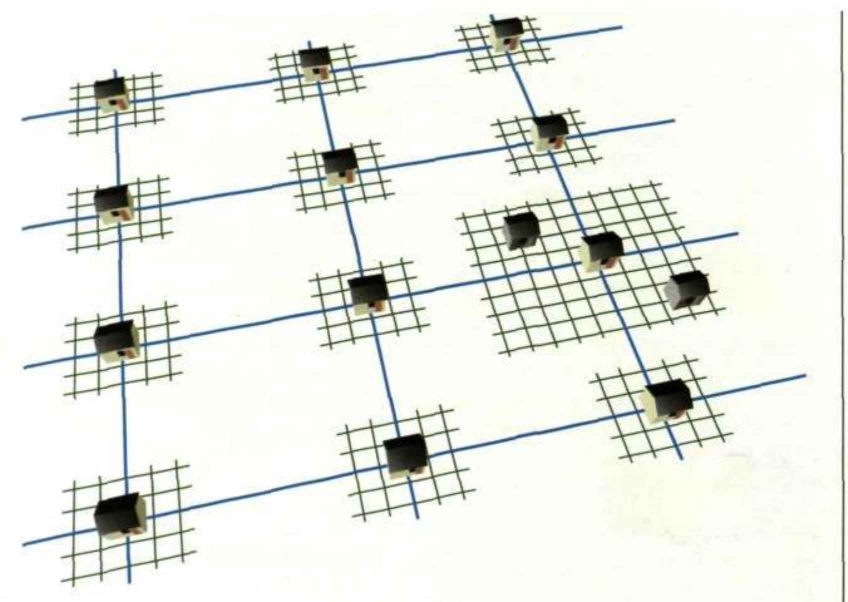
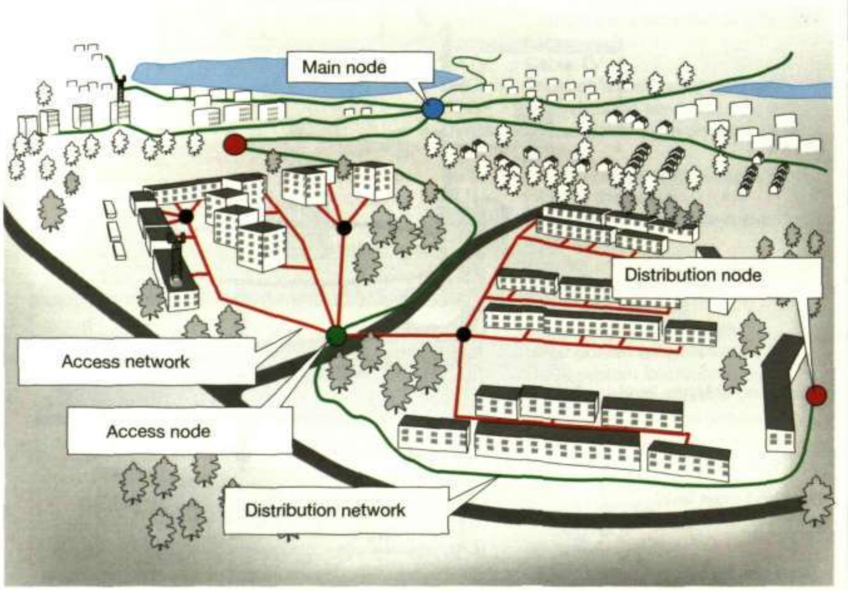


Figure 3
The fine-meshed network with main nodes.

The trunk network (Figure 3) links the main nodes. The requirement of redundant connections must be taken into account at an early stage of development. The main node is the central place for fiber connection. Distribution nodes are connected to main nodes through the regional/metropolitan network. Each main node serves roughly 20,000 - 30,000 households, enterprises or other organizations within the area. Large cities need several main nodes, whereas smaller municipalities can share a main node with a neighboring community.

Figure 4
The physical network.



Full redundancy is planned between main nodes: one main node is connected to several other main nodes to allow alternative transmission paths. The trunk network and the main nodes are subject to stringent security criteria. The main node might also include the service providers' active equipment.

The regional/metropolitan network links main nodes and distribution nodes in a locality. In municipalities that lack sufficient demand for a main node of their own, at least one distribution node will be connected to the national main network through a main node.

The distribution node is primarily a passive cross-connection point at the fiber level within the area. Even so, the distribution node must be designed in consensus with all operators to take into consideration the placement of active equipment. To achieve redundancy, the distribution node should be connected through different links to two different main nodes. The distance between the distribution nodes will vary in proportion to population density.

The distribution network, which comprises the connection between distribution nodes and access nodes, consists of a loop that links access nodes with two distribution nodes. The access node and distribution node can be co-located depending on the design of the distribution network.

A network built within a building is called an access network and is connected to the access node. An access network can comprise anything from a network within a building to networks in and between several buildings—for example, campuses, hospitals and industry parks.

From theory to practice—the Hudiksvall broadband island

The vision of nationwide broadband availability is already becoming a reality, as illustrated by a pilot project in Hudiksvall, Sweden conducted in a joint venture between Ericsson and Hudiksvallsbostäder AB (Figure 4). The project is based on the con-

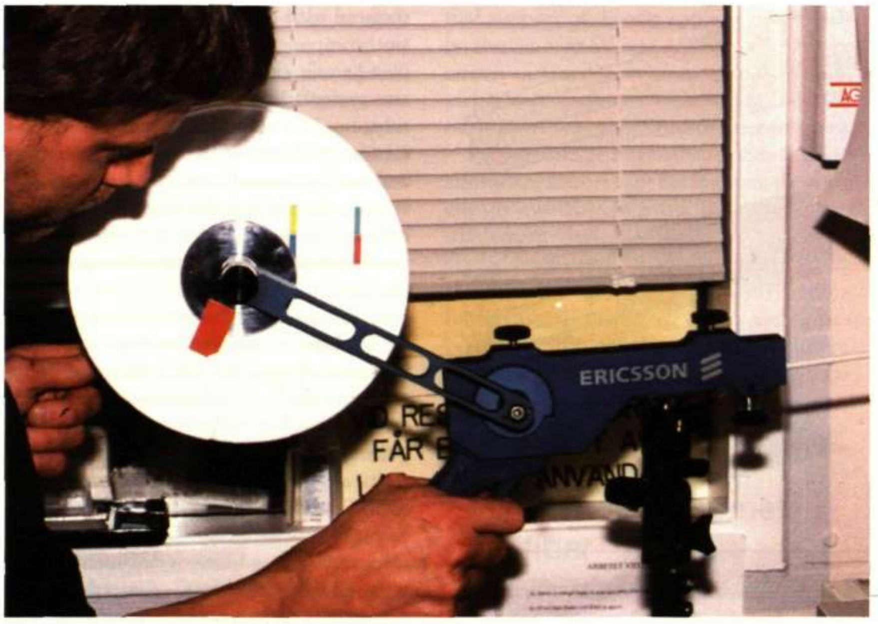


Figure 5
Fiber blown into the Ribbonet microduct using patented blowing tool.

cept of broadband islands—a network that delivers services from a number of providers with national and local presence. The aim of the project has been to build a high-capacity broadband network that offers fixed, symmetrical, high-speed Internet access suitable for home entertainment, small businesses, home-based working, schools and hospitals. To be prepared for emerging services, such as IP-based TV distribution, IP telephony, and other, yet undefined services, the access network delivers a capacity of 100 Mbit/s to each apartment.

The access network is based on an active star topology (a fiber-pair to each end-user whether actively connected or not) that radiates from a centrally located access node. Fibers are blown through an installed microduct system from the apartment to the access node, which is directly connected to the distribution network with a capacity of 1 Gbit/s. This capacity is shared by all users within the network area. Several areas are connected to the main node, which has a switching capacity of more than 100 Gbit/s. All apartments are connected to the access node by blown-in multimode fibers (Box B). Tenants that subscribe to broadband services are connected by a switched Ethernet network that operates over the multimode fiber connection. Initially, the network gives symmetrical access at 100 Mbit/s. Later, the network can be upgraded to Gigabit capacity by changing end-user and access node electronics. The installation described here is based on the Ericsson Ribbonet system.

The transmission layer – Ethernet-over-fiber

The choice of transmission technology is every bit as important as having a resilient and future-proof physical network. Carriers and service providers want stable solutions that are also flexible enough to accommodate a rapidly changing landscape. Ethernet-over-fiber meets these criteria and

BOX B, THE ERICSSON RIBBONET SYSTEM

A basic requirement throughout the project in Hudiksvall has been to minimize the cost of fiber installation, so that the passive parts of the network do not cost more than a CAT5 copper network. Innovative techniques have reduced costs and sped up the installation process. Ribbonet, for example, is a complete system developed by Ericsson that comprises a lightweight, compressed air-driven blowing device (Figure 5), microducts (Figure 6), fiber-distribution fields, fiber-optical cables and pre-terminated fiber-pairs. The system prevents potential damage caused by pulling cables, and saves time since no intermediate fiber splicing is required for point-to-point connections. One intermediate splice is needed if fiber concentration points are used (Figure 7).

Other benefits of the blown-fiber approach are that damaged fibers can be replaced quickly and multimode fibers can be replaced with single-mode fibers for certain routes. In cases where service must be deployed in stages, network builders can initially install a conduit network (empty microducts) to user premises. Later, fiber can be blown in more or less on demand.

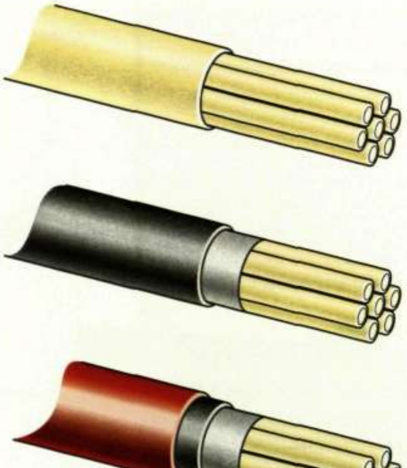


Figure 6
Ribonet multi-microducts (indoor use, top; outdoor use, center; direct burial outdoor use, bottom). Available in combinations of up to 24 single ducts.

is poised to become the preferred access solution in part thanks to its functional flexibility and low cost, and in part to widespread knowledge and acceptance of the underlying technology. As a consequence, the IEEE is working on a standard for Ethernet in the first mile: the standard, IEEE 802.3ah, is scheduled for completion by 2003.

Likewise, a group of equipment manufacturers is currently developing a solution it calls Ethernet passive optical network (PON). The solution is largely based on previous work within full-service access networks (FSAN), which standardized ATM PON. PONs use point-to-multipoint topology with passive optical splitters, to eliminate active electronics (regenerators, amplifiers, and lasers) in the access network. This can also be achieved with long-reach active optics, but the corresponding PON topology can reduce the number of outgoing fibers and lasers needed at the active distribution point (for example, at the central office).

PONs are effective at delivering services that require large bandwidth, such as analog video distribution and fast Internet access, but they do so in a technically complex manner with little or no flexibility for future upgrades and network expansion. Also, due to their asymmetric nature, PONs do not adequately serve the needs of users who request equal upstream and downstream connectivity.

Ethernet-over-fiber active access networks overcome most of the limitations associated with PONs. The falling cost of optical components has largely eliminated the cost benefit of deploying asymmetric passive systems in favor of symmetrical active solutions. An active access system—with a symmetrical infrastructure of transmitters and receivers at each end of the connection—is a proven platform for distinct and differentiated services, and a cost-effective and flexible basis for next-generation access solutions.

Ericsson's Ethernet-over-fiber access solution sets out to capitalize on all these benefits, building a switched Ethernet infrastructure that is based on low-cost standard components but takes the perspective of a public access network rather than the traditional enterprise networking view of Ethernet. One important aspect is circuit-oriented provisioning: In earlier solutions, subscribers would have a dedicated pair of wires or some other form of leased line that connected them directly to the service provider. In Ericsson's current solution, this objective is moved up to the data link layer. Each subscriber has a dedicated virtual LAN for basic connectivity services, and service providers have dedicated virtual LANs for providing services to subscribers.

Looking above and beyond the access network, Gigabit Ethernet has already established itself as the main alternative for new operators who are building metropolitan area networks (MAN or Metro), since the cost of Gigabit Ethernet is substantially less than that of synchronous digital hierarchy (SDH). Within the next two years, Gigabit Ethernet is expected to become commonplace for enterprise access applications, although not all businesses will require the full capacity supported by this standard.

Over the next few years, 10 Gigabit Ethernet (10GE) will probably emerge as the preferred solution for both MAN and wide area network (WAN) applications, since it provides adequate capacity and supports greater reach than all previous versions of Ethernet. Incidentally, unlike its predecessors, 10GE is also the first purely optical Ethernet standard.

Maturing Ethernet standards

Innovations in two main areas are driving the development of optical access networks:

- structured optical cabling systems, with new and efficient means for installation; and

- advanced, low-cost optics, such as optimized components from the Ethernet world—these yield cost-effective optical solutions that are suitable for deployment in the access network.

Regarding the transmission (or physical) and data link layers, maturing Ethernet standards comprise the single most important enabler of advanced new access networks that offer high capacity at low cost. In terms of functionality, recent and emerging standards are introducing features such as priority traffic handling, virtual LAN-based security, and improved network-management mechanisms, which means it may soon be possible to design flexible, multipurpose access networks using only Ethernet standards.

Full-duplex switched Ethernet (IEEE 802.1D and 802.3x) gives full-speed, two-way, point-to-point links with no collisions, and theoretically, no distance limitations. The addition of the appropriate implementation of switch nodes creates a flexible and powerful network with large throughput, limited overhead, and minimal end-to-end latency.

Likewise, the addition of priority traffic handling (defined by IEEE 802.1p) and switch hardware port queues, supports fine-tuned, differentiated traffic classes for optimal support of diverse applications. Finally, the use of IEEE 802.1Q virtual LANs for circuit-oriented provisioning and privacy, yields a fully functioning access solution built purely on standard Ethernet features.

For full-fledged broadband access, a few additional features are needed, but these are already available in many Ethernet switch implementations and will probably be included in the future standards suite. Among these features are bandwidth-provisioning mechanisms, managed delay and jitter bounds, traffic volume and flow counters, and enhanced virtual LAN management capabilities. Most of these features are currently included in Ericsson's local community system (LCS).

Ericsson and Ethernet-over-fiber

The Ericsson LCS, which is a consistent and integrated Ethernet access network optimized for dedicated fiber infrastructure and Ribbonet, enables large, symmetric bandwidth to each connected user and supports differentiated service level agreements and quality of service. Originating from the broadband island concept, LCS builds access connectivity almost entirely on data link layer (Layer 2, L2) switched Ethernet with

TABLE 1, COMPARISON OF ETHERNET STANDARDS AND COMMON EXTENSIONS

TYPE OF ETHERNET	STANDARD	REACH	DESCRIPTION
Fast Ethernet, 100Mbit/s			
MMF 100Base-SX	EIA/TIA standard	300 m	Formerly used in Ericsson local community system (LCS, see below)
MMF 100Base-FX	IEEE standard	Up to 2 km	Used for final drop in Ericsson LCS
SMF 100Base-FX	Not an IEEE standard, but vendor-specific implementations are available	Standards proposals (for example, 10 km) made to IEEE 802.3ah	Not currently included in Ericsson LCS, but feasible
Gigabit Ethernet, 1Gbit/s			
MMF 1000Base-SX MMF 1000Base-LX	IEEE standards for short range	<550 m	Option: available with Ericsson LCS
SMF 1000Base-LX	IEEE standard specifies only	5 km (10 km *de facto* standard) Less common extensions are 40 km, 70 km, 100 km	5-10 km and 70 km offered with Ericsson LCS

Figure 7
Ribbonet duct-/fiber-jointing cabinet for outdoor use, for example, in residential areas.

100 Mbit/s full-duplex user gross data rate. The system has been designed primarily for providing IP-based services with a guaranteed quality of service. However, LCS is also a transparent access network for enterprise users who need native Ethernet LAN connectivity for private data networking.

LCS is modular, flexible and scalable, with a network topology that can be adapted to almost any physical, fiber-based access infrastructure. It uses a dedicated fiber pair (it can also use copper CAT5 connections) from customer premises to an initial aggregation point that houses an Ethernet switch node. At present, these connections are multimode, but single-mode connections are foreseen in the future, enabling much longer fiber runs from (or to) the home.

The use of switching and virtual LAN technology makes LCS a circuit-oriented Ethernet access network and a versatile platform for providing IP-based services. Virtual LANs are employed to separate and segregate users and private networks, differentiated services, and access network elements. In this way, a robust foundation for user integrity and network security is established without making use of routing in the access network.

As mentioned above, individual service providers have one or more dedicated virtual LANs for providing differentiated services to their subscribers. For service providers connected to LCS on L2 Ethernet—that is, in cases where no intermediate routing is involved, end-to-end provisioning is completely transparent. This kind of transparency is considered a key property of next-generation access networks, since they are completely neutral and open to connection from any kind of service from any service provider. However, because not all kinds of service are currently suited for pure L2 end-to-end provisioning, the Ericsson LCS can also use higher-layer protocols. At the point where the L2 traffic domain is terminated, LCS L2 provisioning needs to be mapped to higher-layer provisioning mechanisms. A typical example is voice over IP (VoIP), since the media gateway might be situated far away on a routed WAN. In this case, IEEE 802.1p priorities are transferred to and from IP TOS/DSCP, and possibly RSVP.

In its default configuration, LCS also touches Layer 3 (OSI network layer) and higher-layer protocols for augmented IP multicast performance, and to block out spoofing and denial-of-service attacks. With regard to functionality that can be implemented on either L2 or L3, Ericsson's LCS solution adheres to the following guidelines:

- Although services can be provided more or less entirely on L2, such a scheme requires capable access termination equipment at customer premises. In its basic configuration, the Ericsson LCS includes a less capable device with no advanced L2 features, rendering lower cost per customer connection. In this configuration, most services are provided on L3 and L4, and consequently, most of them terminate in users' devices.
- When the Ericsson LCS is configured for VoIP dial tone, more sophisticated access termination equipment is needed—that is, additional L2 functionality is required in the user premises equipment (the Ericsson DRG). In this case, most of the service provisioning is moved down to L2 and controlled in the L2 access network, efficiently enabling specific quality of service.

Toward the all-optical network

With these solutions available today, an optical network should be the obvious choice

TRADEMARKS

for greenfield installations or refurbishing. Interestingly, with an annual refurbishing rate of 5-10% of a network, all-optical technologies will also dominate the access network in just a few years. However, the physical infrastructure in fixed public access networks tends to last longer than any other fixed communication infrastructure. This is primarily due to the immense quantities of cable and labor involved in building the networks.

Given the large base of legacy access networks, and that much of the deployment cost of fiber networks is determined by the cable installation, efficient migration strategies are needed. At some point, a decision to move from a coaxial access network to an optical fiber network will inevitably involve ripping up old plant and installing new fiber. However, the associated costs can be mitigated by adopting a mix-and-match approach. All-copper networks can be updated using a hybrid solution, where, for example, fiber is taken to the roadside, and individual customer connections remain on coaxial cable. Commercial considerations will determine how far an operator will take the fiber installation, but when demand makes it feasible, the ultimate goal—of replacing remaining copper and of having an all-optical access network—can be achieved.

As in all effective migration strategies, key factors are to balance investments in new equipment with the capabilities of existing infrastructure, and to stay one step ahead of market demand while building a foundation for generating future revenues. For example, a PTT migrating from narrowband over a twisted-pair connection might implement an Ethernet-over-DWDM solution in the metropolitan network in combination with Ethernet-over-VDSL in the access network. By the same token, a CATV operator moving from analog broadband to digital broadband-over-fiber might first implement Ethernet-over-cable before having to invest in a new optical fiber access network. Note that both of these examples build on early inclusion of Ethernet access technology, and pave the way toward a future all-optical Ethernet infrastructure.

The introduction of optical edge multiplexers will be an important factor in managing a smooth migration strategy. Optical edge multiplexers provide optical gateways that perform a mixture of electrical and optical transmission (for example, IP carried over Ethernet, ATM and SDH/SONET) and convert and combine traffic into suitable optical formats for onward transportation. Optical edge multiplexers will enable legacy technology to be reused as part of the new generation of converged, multiservice networking infrastructure. They will be built on thin SDH/SONET technology, which is to say that bandwidth will be managed in small chunks (50–150Mbit/s). With scalability a prime concern, the edge multiplexer will be designed to handle virtually all expected growth.

Gigabit Ethernet is becoming increasingly popular for metro and access solutions. It is particularly suitable for enterprise LAN and FTTH applications thanks to simplicity and cost benefits compared to alternative solutions based on ATM and SDH.

Most probably, in the end we will see a variety of complementary solutions for the access network. Operators focused on IP/LAN services may opt mainly for 10 Gigabit Ethernet or Gigabit Ethernet over DWDM, whereas telecom and multiservice operators might base their access networks mainly on TDM or TDM over DWDM.

Conclusion

All-optical access networks are starting to materialize before our eyes. After spending some two decades in feasibility studies, lab and field trials, public access networks based on fiber optics have eventually found their way into product portfolios, and are now becoming ready for deployment at scale.

Advanced optical backbone networks are continuously helping to prove the viability and potential of optical technologies, and providing innovations leading to improved components. The cost-sensitive public-access application benefits from this development, as optical transmission technologies mature.

The synthesis of cost-effective fiber optics on the one hand, and simple and efficient networking technology from the Enterprise LAN segment on the other, has resulted in Ethernet-over-fiber access solutions with unprecedented capacity and cost-effectiveness. The amendments required to optimize optical Ethernet networks for large-scale public access are largely in place today, and the rest is underway. The result is a ubiquitous infrastructure for IP-based communication, reaching out to connect any and every end-user, and carrying any kind of traffic.

REFERENCES

1 Swedish ICT-Commission, Report 37, 2001 (available at www.itkommissionen.se).

Previous issues

No. 3, 2001

MMS—Building on the success of SMS
SyncML—Getting the mobile Internet in sync
Further evolution of the GSM/EDGE radio access network
Evolving WCDMA
Telecom management for ENGINE
Approach to E2E service assurance on the mobile Internet
CDMA2000—A world view

No. 2, 2001

Wireless image applications and next-generation imaging
Ericsson's Service Network: a "melting pot" for creating and delivering mobile Internet service
The WISE Portal 2.0 solution—Timely delivery of tailored mobile Internet services
Supply as an enabler in the new telecoms world
The CDMA2000 packet core network

No. 1, 2001

Wireless-Trench technology for portable wireless applications
AXE 810—The evolution continues
Open architecture in the core of AXE
Cable modems—Broadband highway to the home
The R380s—The first smartphone from the Ericsson–Symbian partnership

No. 4, 2000

Experiences of the live WCDMA network in Stockholm, Sweden
Media Gateway for mobile networks
Ericsson's IP-based BSS and radio network server
Control servers in the core network
Ronja—A Java application platform
Wireless *ad hoc* networking—The art of networking without a network
Ericsson's WebOnAir Information Server

No. 3, 2000

ENGINE network server
ENGINE Access Ramp—The next-generation access architecture
Speech recognition technology for mobile phones
GPRS support nodes
Family of RBS 3000 products for WCDMA systems
Ericsson's network-based IP-VPN solutions
Fixed wireless access for broadband services

Telefonaktiebolaget L M Ericsson
SE-126 25 Stockholm, Sweden
Phone: +46 8 7190000
Fax: +46 8 6812710

ISSN 0014-0171
Ljungföretagen, Örebro 2001