Ericsson

REVIEW

1/2002

THE TELECOMMUNICATIONS TECHNOLOGY JOURNAL

Self-tuning networks

Automated network management

IP in WCDMA/GSM core networks

Communication using SIP

Ericsson Mobile Operator WLAN solution

ERICSSON

Cover: As cellular networks grow in size and complexity, the process of managing them becomes increasingly costly. The answer is automated network management—self-tuning networks.

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 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:
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: 79, 2002

Contents

Contributors

In this issue

Tomas Boström joined Ericsson Wireless LAN Systems in 1999 to work as system manager, defining system and network requirements for wireless LAN systems. In 2000, he became program manager at the same unit. He holds a master of science degree in electrical engineering from the Royal Institute of Technology, Stockholm, Sweden.
tomas.bostrom@ebc.ericsson.se

Tomas Goldbeck-Löwe is employed at Ericsson Research, IP Networks, where he works with mobility issues. Most recently he has been involved in several projects whose focus is on connecting wireless LAN to a cellular system. He joined Ericsson in 1995, working in development and system management. He has experience of most of the different wireless packet-data systems developed by Ericsson, including Mobitex, CDPD, PPDC GPRS, UMTS, CDMA2000 and wireless LAN. Tomas holds an M.S. degree from the Lund Institute of Technology.
tomas.golbeck-lowe@era.ericsson.se

Peter Granström, who has been with Ericsson for 17 years, is currently strategic product manager at Ericsson's IP Multimedia system development unit. He has held various positions in the areas of IP telephony, intelligent networks, international gateway switching, global VPN networks, and AXE support processor development. He holds a master of science degree in engineering physics.
peter.granstrom@etx.ericsson.se

Per Gustås has worked as a consultant to Ericsson testing and designing transport networks, AXE platform software, test automation products, and several GSM and PDC OSS products. He joined Ericsson in 1999 and is currently system manager of the R-PMO. He holds an M.S. in computer science from the Linköping Institute of Technology.
per.gustas@era.ericsson.se

Heino Hameleers is a senior product manager and master system engineer at Ericsson Eurolab Deutschland. He joined Ericsson in 1995 and has worked at Core Network Mobile Systems with the design, system management and product management issues of data communication, multimedia and core network architecture. Heino recently started working with multimedia strategies in the product management unit of the WCDMA/GSM mobile systems business unit. He holds an M.S. in computer science from the Universiteit van Amsterdam.
Heino.hameleers@eed.ericsson.se

Christer Johansson joined Ericsson in 1970. He is a system expert of network and system architecture in charge of driving system management activities for the connectivity layer in third-generation mobile systems. Christer has previously held several line manager and expert positions within Ericsson development organizations, including AXE development, and tools and methods development.
Christer.johansson@era.ericsson.se

Ralf Keller joined Ericsson Eurolab in 1996, working on wireless ATM systems and on mobile electronic commerce. He is currently a master researcher and team leader. His research examines multiaccess system integration in terms of service, network and business. Multiaccess integration includes the integration of wireless LANs and large user group support in third-generation systems. The main focus of his research is on the evolution of third-generation and future-generation systems. Ralf holds a Diploma degree in applied economics and computer science in 1991 and a Ph.D. in computer science from the University of Mannheim.
ralf.keller@eed.ericsson.se

Per Magnusson joined Ericsson in 1993 and has worked as system architect developing solutions for testing the conformance and performance of Ericsson's products for GSM, TDMA, PDC and WCDMA. In 1998, Per joined Ericsson Research as a senior specialist. His special area of focus is radio resource management. He holds an M.S. in applied physics and electrical engineering from Linköping Institute of Technology.
per.magnusson@era.ericsson.se

Sean Olson is one of the foremost experts on SIP in the Internet community. Until quite recently, he served as system manager at Ericsson's IP Multimedia product development unit, working directly with the CSCF. He holds a bachelor of science degree in computer science from Texas A&M University.

Jan Oom worked many years as a consultant to Ericsson developing computer hardware and communications software products for the operation and management of data and telecommunication systems. He joined Ericsson Research in 1995 as a system architect working on the design of Ericsson's WCDMA O&M system. He holds an M.S. degree in electrical engineering from the Royal Institute of Technology in Stockholm.
jan.oom@era.ericsson.se

Mark Peck was previously the product manager responsible for the CSCF at Ericsson's IP Multimedia product development unit. He holds a bachelor of science degree in electrical engineering from Rice University, and a master of computer science degree from the University of Texas at Dallas.

Niclas Storm has worked as a software designer of various telecommunication applications for fixed and mobile networks. He joined Ericsson in 1998 and currently works with WCDMA RNC product development, primarily RNC system design of O&M mechanisms. He holds an M.S. degree in applied physics and electrical engineering from Linköping Institute of Technology.
niclas.storm@era.ericsson.se

Editorial

Eric Peterson

Last year at this time I predicted that instead of just reading about the Mobile Internet, in 2001/2002 a lot of ordinary folks would actually begin experiencing it first-hand. Since then I have read about the launch of GPRS in Europe, WDCMA in Japan, and how many of the heavy players in the Americas are transitioning to GSM/GPRS. At the same time, however, the economy has stalled after the IT bubble burst and due to the 9-11 terrorist attacks. And consequently many operators have slowed their pace of investment in new networks and are instead concentrating on improving capacity in existing networks.

With this in mind, I began to wonder to what extent had my prediction held true? Have I actually seen anyone using the Mobile Internet outside of work? If I were to hit the streets of Stockholm for one day, would I see anyone using the Mobile Internet? Hmm. Good question. And so I determined that I would spend one day riding Stockholm's commuter trains, subways and buses during peak morning and evening rush hours, visiting Arlanda International Airport, Stockholm Central Station, speaking with day managers at hotels, and talking to salespeople at shops that sell mobile phones and accessories.

While riding the commuter trains, subways and buses I saw a lot of mobile phones, but very few people were conversing on them. Of course, this comes as no surprise—who wants to talk on the phone when surrounded at close quarters by total strangers? On the other hand, I saw several people playing games on their phones or reading and sending SMS. They might also have been reading news, etc. (WAP)—I couldn't always get close enough to tell. I also observed a good number of commuters passing the time listening to music via portable CD, MD or MP3 players. On one occasion, while I was riding the bus, a fellow sat down next to me and proceeded to surf the Web (on a Compaq iPAQ).

On entering the terminal for international departures at Arlanda International Airport a huge banner advertising wireless LAN immediately grabbed my attention. I inquired about this at a shop that sells mobile phones and was told that the service is quite popular. Those persons who want to use the wireless LAN but who lack a subscription can purchase a start packet that includes a PC card, software, and four 24-hour vouchers (with one-time password). Repeat users can purchase individual 24-hour vouchers. Although I did not see anyone using the wireless LAN while I toured the airport, I did observe numerous travelers running laptop computers. One fellow appeared to have a connection for mobile data via the IrDA ports on his mobile phone and laptop. While at the airport I also observed a very high percentage of people talking on their phones—in fact, I don't believe I have ever seen so many people using their phones in one place at the same time.

Back in Stockholm, I watched the crowd as it pulsed through Central Station. There is a lot of fast-moving activity there, so I did not expect to see anyone using mobile data services. But I did have my first sighting (apart from demos) of a Bluetooth headset. I also saw several people using their phones for traditional voice services. At the business lounge, however, I was shown a wireless LAN base station and told that the service is very popular with traveling businesspeople.

After leaving Central Station, I visited six hotels: three of Stockholm's most prominent hotels, two recently opened business hotels, and one very small business hotel. I spoke with the day manager on duty at each hotel. To my dismay, none of the "big-name" hotels offer or have plans to offer wireless LAN services. But the two newly opened business hotels have already installed wireless LAN—in one of them, on every floor!

At shops that sell mobile phones and accessories I chatted with the personnel about GPRS, Bluetooth, the Mobile Internet, wireless LAN, and different kinds of subscription plans. I learned that many people are buying GPRS phones, but few of them are subscribing to GPRS service. That's odd, I thought. I wonder why? And so they told me: "GPRS is currently a supplementary service that cannot be combined with prepaid subscriptions." Ah yes, prepaid. Fortunately, this situation might soon change. AT CeBIT, for example, Ericsson demonstrated a prepaid system that allows operators to charge for MMS in real time.

I must concede that my prediction was overly optimistic. The early adopters are experiencing the Mobile Internet, but not the majority of subscribers. On the other hand, a lot of people now own GPRS phones, so I am left wondering why they do not use the Mobile Internet—might this be due to a lack of subscription options? After all, a lot of ordinary folks prefer prepaid subscriptions. Another factor might be the availability of Mobile Internet services, but given that players like Sony and RealNetworks have announced that they will be providing content this will hardly remain an issue for very long.

For what it is worth, this exercise—while neither exhaustive nor scientific—has been entertaining, and based on my personal observations, I feel more confident than ever about the Mobile Internet—the fun is about to begin. Truly!

Eric Peterson
Editor

Real-time performance monitoring and optimization of cellular systems

Per Gustås, Per Magnusson, Jan Oom and Niclas Storm

As cellular networks grow in size and complexity, the process of managing them becomes increasingly costly for mobile operators. For this reason, automated network management is very appealing.

The authors describe an architecture that uses event-based statistics for real-time performance monitoring and optimization of cellular networks. They present the results of a field trial together with SmarTone in Hong Kong, and conclude with an overview of current product developments in this field.

Introduction

Cellular networks are becoming increasingly complex. In coming years the introduction of third-generation networks will compound this situation. As a consequence, there is a need for simple or automated operation and maintenance (O&M), to decrease operators' costs and to use hardware investments in the most effective way. Likewise, operators need customized functionality—a requirement that puts great demands on system flexibility. This kind of flexibility can be achieved by allowing the network to produce real-time events instead of counter-based statistics. But doing so requires an effective and well-structured network architecture.

This article describes an architecture that employs event-based statistics for building effective and flexible performance monitoring and optimization applications. It also points to Ericsson system properties, activities, and experiments, showing that Ericsson is well on its way toward creating a common platform for the monitoring and optimization of second- (GSM) and third-generation (WCDMA) mobile systems.

BOX A, TERMS AND ABBREVIATIONS

3GPP	Third-generation Partnership Project
BSC	Base station controller
BTS	Base transceiver station
CORBA	Common object request broker architecture
CT	Control task
CTR	Cell traffic recording
FAS	Frequency allocation support
FOX	Frequency optimization expert
GPEH	General performance event handling
GSM	Global system for mobile communication
GUI	Graphical user interface
HCS	Hierarchical cell structure
IP	Internet protocol
IRP	Integration reference point
MS	Mobile station
MT	Monitoring task
MTR	Mobile traffic recording
NCS	Neighboring cell support
NOX	Neighboring optimization expert
O&M	Operation and maintenance
OSS	Operations support system
PM	Performance monitoring
RANOS	Radio network operation support
RNC	Radio network controller
R-PMO	Real-time performance monitoring
TCH	Traffic channel
TCP	Transport control protocol
UETR	User equipment traffic recording
WCDMA	Wideband code-division multiple access

Optimization in a radio access network

The radio access network provides connection links between mobile stations (MS) and the core telecommunications network.

In GSM systems, the radio communication with each mobile station is maintained by base station controllers (BSC), which control several base transceiver stations (BTS). A BTS covers a limited geographical area known as a cell. When a mobile station moves between cells, it is up to the BSC to move the signaling to the corresponding BTS. This process is called handover. In WCDMA systems the principle is the same, but the terminology is somewhat different.

The operator uses a management system to manage the network. In Figure 1, the emphasis is on the management signaling within the radio network. Along with a variety of other applications—for example, fault management and customer services—the typical management system has tools for helping the operator to optimize, or tune, the radio network.

Networks have traditionally been optimized by a radio engineer who, using planning tools, performance analysis data, and experience, sends configuration commands to the network.

The term optimization algorithm should be interpreted to mean software that can assist the radio engineer and handle a wide range of optimization situations, from slow variations in the radio network, such as deploying new cells, to fast variations, such as changing traffic load. Ideally, the optimization algorithms optimize system performance without continuous operator intervention. However, in complicated situations it is important to include the radio engineer in decisions.

An algorithm could be employed to decrease the coverage of a cell due to temporary traffic overload. An optimization application of this kind continuously configures the network in accordance with measurements emitted from it. If this application were removed or ceased to function, the network could continue to maintain its traffic services, albeit in a less optimized way.

The architecture

Distributed agents

Traditionally, management applications execute in a central management node. Predefined statistics are periodically sent from network elements to the central management node. By contrast, real-time performance monitoring and optimization applications require a more flexible aggregation of measurement data. However, it is not feasible to send raw measurement data, since in large networks (having, for example, as many as 40,000 BTSs) this would put extreme performance requirements on the central management system and on the data capacity of the management network (Figure 2).

To create an effective environment for the performance monitoring and optimization functions that rely on large amounts of event-based measurement data, an architecture is needed that permits the distribution of functions. Raw measurement data must be handled efficiently and should be collected and processed near the source before it is forwarded to the next-higher level in the hierarchy (Figure 2).

The topic of distributed management systems has been dealt with extensively throughout the years. The challenge here has been to apply the theories to create an architecture that is suited for the real-time performance monitoring and optimization of cellular networks.

Accordingly, the applications have been split up into parts and subdivided throughout the management nodes and traffic nodes of the network. In this article, we call a module that constitutes a distributable part of the management application an agent. Note: The word agent is used simply to refer to distribution and an open interface. It does not try to quantify any degree of intelligence. Agents interact via open interfaces at different levels of the system. The concept of integration reference points (IRP) is being used within the Third-generation Partnership Project (3GPP) as the means of achieving a management standard for interoperable third-generation cellular systems.[1] At present, standards exist within the fault- and configuration management areas and Ericsson has submitted a proposal for performance management.

The use of interacting, yet autonomous, measurement and control functions calls for robust and scalable management systems that can be distributed across the network. Ideally, the distribution should be independent of the borders constituted by physical nodes. This is achieved by using protocols that expose the same interface to the individual agents regardless of whether the data

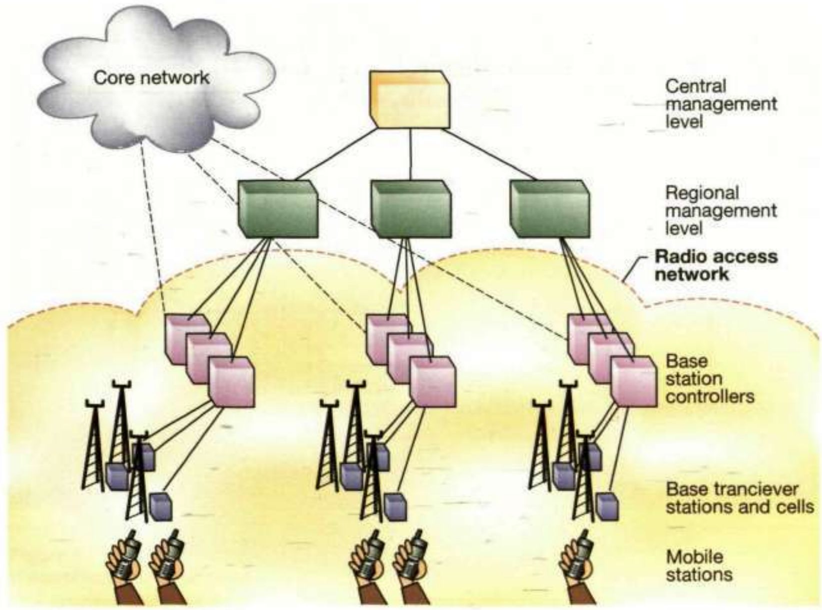
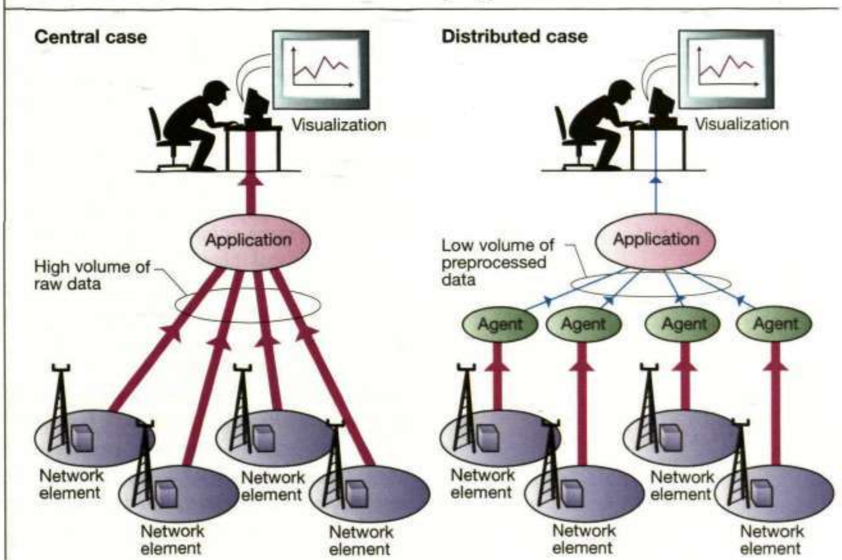


Figure 1
A radio access network.

Figure 2
Distributed functions are a solution for handling large amounts of data.



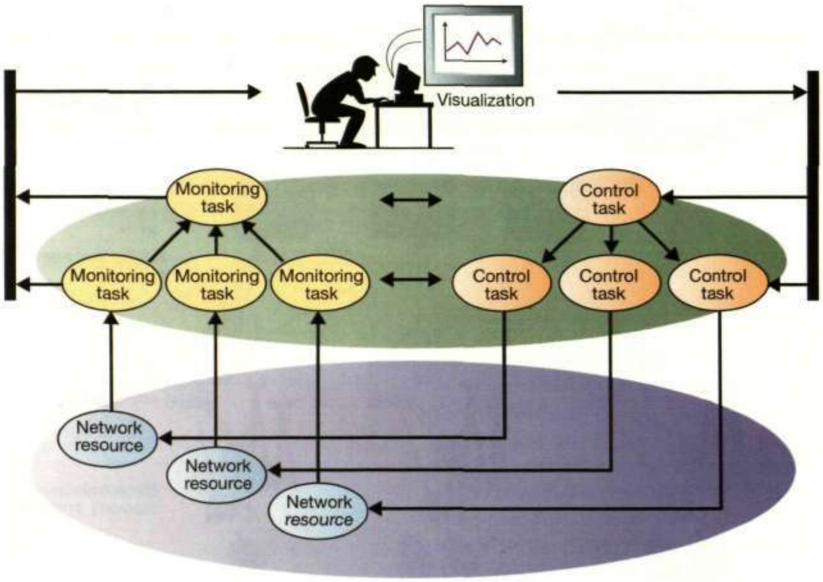


Figure 3
Schematic view of task allocation in a system (not considering deployment). The number of layers can be extended.

is transported locally or via a connecting network.

Monitoring and control tasks

A real-time performance monitoring or optimization application is functionally built up by means of monitoring tasks (MT) and control tasks (CT). Figure 3 shows a schematic and ideal distribution of tasks throughout a complete system. Note, however, that consideration has not been given to the physical deployment into different nodes or agents.

Measurement information flows upward in the system, whereas control information flows downward. Control tasks typically act on measurement information in the same layer and control information from the next-higher (superordinate) layer. Monitoring tasks are usually configured by control tasks at the superordinate layer.

Monitoring tasks monitor the performance of related resources (ordinarily subordinate agents or real network resources) by subscribing to real-time performance data. Once a subscription has been set up, events will be delivered to the subscriber until it cancels the subscription.

Although a monitoring task can be connected to other monitoring tasks in any way, a strict hierarchy (Figure 3) is most often suitable for cellular systems. The lowest level monitoring tasks collect low-level information, such as bit error rate, carrier-to-interference ratio, handover events, call drop events, and so on. It uses this information to calculate statistics, such as call drop rate. Higher-level agents subscribe to lower-level information in order to compute more abstract quantities, such as capacity and quality for a region or for the entire system. These quantities convey more meaningful information to an operator or a control system.

Control tasks also operate in real-time and are used for controlling network resources, either directly or via other (typically lower-layer) control tasks. A control action can be based on performance data from one or more monitoring tasks, external control actions received from other control tasks, or both. The control rules can be of almost any type suggested in the control literature. The control tasks can also be interconnected in any way. As with monitoring tasks, however, a strict hierarchy is often preferable for cellular systems. This arrangement solves the control problems at the lowest possible level, where the problems can be handled with maximum speed. But if this is not adequate, higher-level and more cautious controllers can be invoked to resolve the control matters.

Monitoring agents are built up from one or more monitoring tasks. Optimization agents are built up from a combination of monitoring and control tasks.

Tasks versus agents

Agents provide a means of building applications that can be scaled and distributed. If different tasks should execute on different physical nodes, then they must reside within different agents. Different tasks should also reside within different agents in order to achieve full flexibility and scalability. However, to achieve high performance, the control and monitoring tasks of the same physical node could be configured to constitute an agent:

- Tasks can communicate more efficiently than agents, since they do not require an open communication interface that supports distribution.
- If several tasks within an agent subscribe to the same information, the information needs only be sent once to the agent, which can then distribute it internally to the appropriate tasks.

- Different scheduling mechanisms can be used at the agent and task levels—that is, operating system-based scheduling can be employed at the agent level, whereas a simpler scheduling mechanism can be employed at the task level.

A monitoring and optimization prototype

A prototype project was launched in 1999 with the primary goal of demonstrating that real-time performance monitoring and optimization is feasible using this architecture in a real environment.

The optimization problem

Hierarchical cell structures (HCS) are used to push traffic down to a lower layer (smaller cells). For instance, by pushing the traffic down to layer-1 cells, operators make better use of the hardware in these cells and reduce the load in surrounding layer-2 and layer-3 cells. The result is an increase in the total network capacity.

The mobile station uses the received signal strength from a BTS to make handover decisions. The HCS feature defines a signal strength threshold (LEVTHR) for each layer-1 cell. If a mobile station that is using a layer-2 or layer-3 cell measures a layer-1 cell signal strength that exceeds this threshold it will attempt to hand the call over to the layer-1 cell. Likewise, if the signal strength falls below this threshold, the mobile station will abandon the layer-1 cell. Accordingly, there is a close relationship between the LEVTHR threshold and the geographical size of the cell. Notwithstanding, there were some drawbacks to this feature that needed to be resolved through the introduction of an optimization algorithm:

- The use of a threshold setting that is too aggressive can cause congestion in the layer-1 cell. When this is the case, it is difficult to provide service to the mobile stations that are close to the micro base station.
- A setting that is too aggressive can push traffic down causing it to suffer from interference due to frequency reuse in the network.

Note: In recent releases of GSM from Ericsson these drawbacks have been resolved.

Optimizing algorithm

An algorithm was developed to resolve these drawbacks. The goal was to improve capacity while maintaining quality in a network with hierarchical cell structures. The algorithm continuously measures the capacity and quality of all regulated cells, and periodically modifies the size (LEVTHR threshold) of layer-1 cells according to the following guidelines:

- To improve the capacity (grade of service), the traffic load should be balanced between layer-1 cells and upper layer neighbors. This is done by adjusting the size of the layer-1 cell.
- If the quality is inadequate in the layer-1 (micro) cells, the size of the layer-1 cell should be decreased. To determine quality, the algorithm measures
  - drop rate (percentage of abnormally terminated calls) in the layer-1 cells; and
  - failed handovers from neighboring layer-2 and layer-3 cells to layer-1 cells.

The criteria mentioned above are periodically combined into a new cell size proposal for each layer-1 cell, which is sent to the BSC.

Figure 4
Hierarchical cell structures.

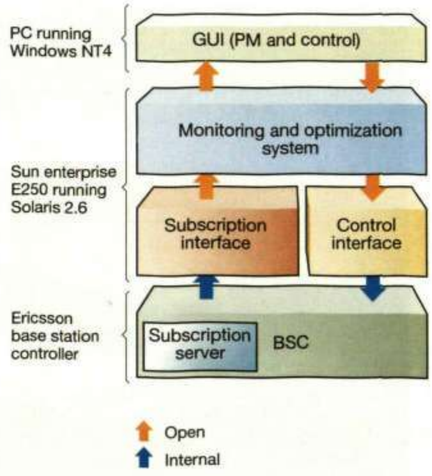


Figure 5
Architecture of the prototype system.

Prototype implementation

The prototype (Figure 5) consisted of

- a graphical user interface (GUI) implemented in Java—to present real-time performance statistics and to control the adaptive tuning algorithms;
- a monitoring and optimization system (also implemented in Java);
- a subscription and control interface that converts from a proprietary Ericsson format to an open agent interface (using CORBA); and
- an Ericsson BSC (Release 7) which had been extended with a subscription server that can extract real-time performance data.

The monitoring and optimization system (Figure 6) included the drop rate measurement, failed handover rate measurement, traffic load measurement, and HCS control tasks, which implement the optimizing algorithm.

The drop rate measurement task subscribes to the call disconnect and handover performed events from the BSC, and calculates the drop rate of each cell.

The failed handover rate measurement task subscribes to the handover attempt and handover performed events from the BSC, and calculates the handover failure rate from the layer-2 and layer-3 neighbors of a specific layer-1 cell.

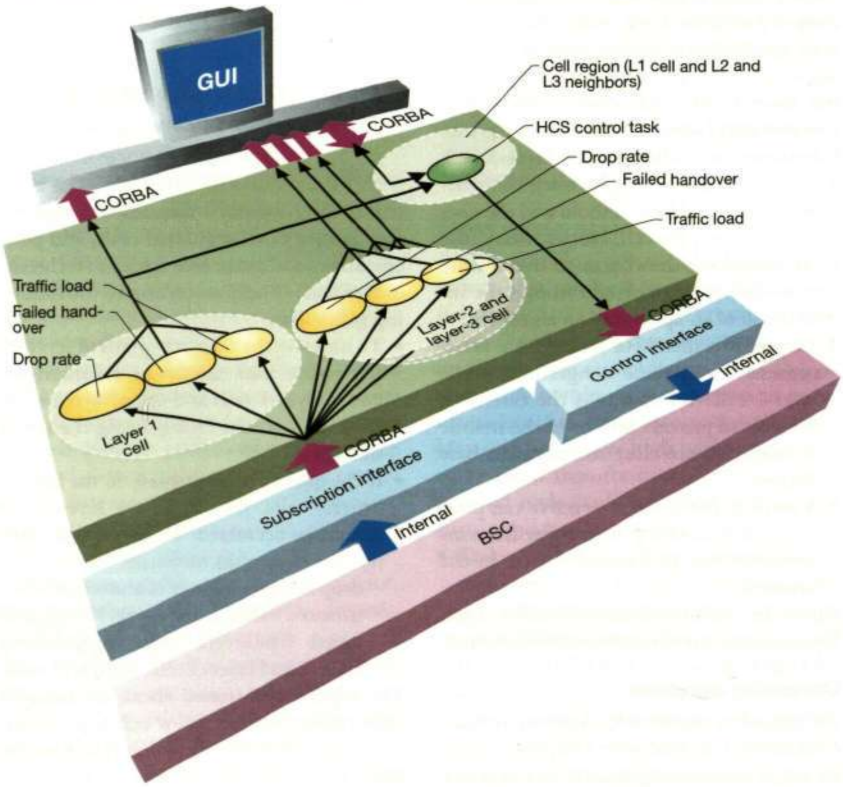


Figure 6
Task implementation showing the control of one layer-1 cell.

The traffic load measurement task subscribes to the traffic load event from the BSC and calculates traffic load for each cell. Traffic load is received every 10 seconds from the BSC.

The HCS control task subscribes to events from the drop rate, failed handover and traffic load measurement tasks of its own layer-1 cell and to traffic load measurement tasks of all neighboring layer-2 and layer-3 cells. It also regularly calculates a new LEVTHR threshold value.

A single host (Sun E250 workgroup server) was able to run the entire adaptive tuning system for the limited field trial area, which included a total of 105 cells. All the tasks were contained within one agent. Open interfaces (CORBA) were used between the GUI, optimization system, and the BSC, but not internally between tasks.

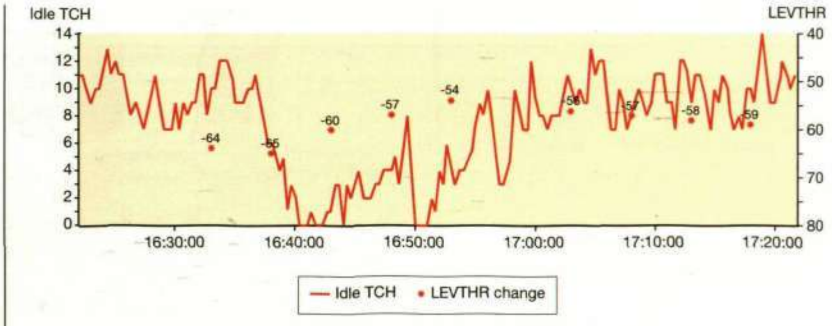


Figure 7
The threshold (LEVTHR) is increased when the number of idle traffic channels decreases, forcing mobile stations to hand over to other cells. This, in turn, further increases the number of idle traffic channels.

Field trial

The field trial took place in Hong Kong in cooperation with SmarTone. Five cells were chosen as layer-1 test cells. All were GSM 900 cells that could be categorized as street cells rather than microcells. Their neighbors, a total of 100 cells, were also included as test cells, giving a total of 105 cells.

Drop rate monitoring, traffic load monitoring, and HCS control tasks were instantiated for all 105 cells. These tasks subscribed to BSC events and periodically generated the following results:

- Events from subscribed monitoring and control tasks were sent every 10 seconds to the GUI for real-time display.
- A new LEVTHR value from each HCS control task was sent every five minutes to the BSC.

Algorithm evaluation

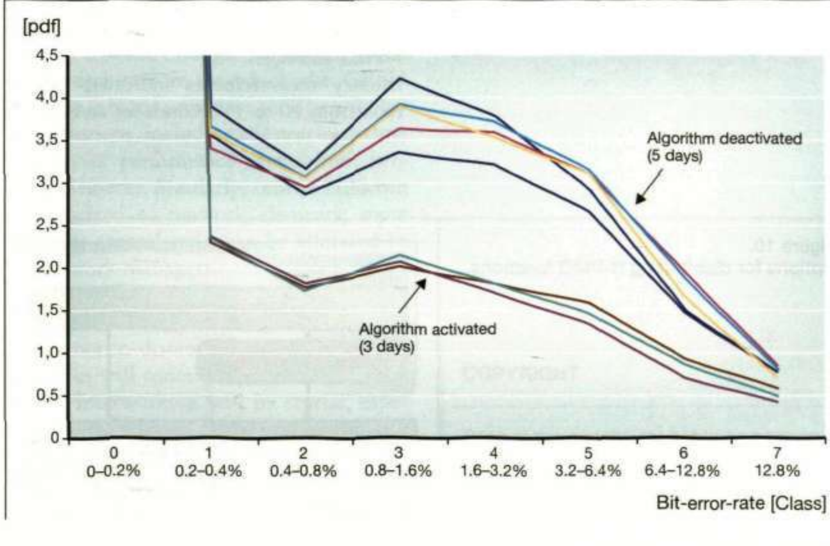
The adaptive tuning algorithm worked as expected according to simulations. Figure 7 shows how the number of idle channels in layer-1 cells can be increased due to changes in cell size (LEVTHR).

Ordinarily, the network worked fine. Therefore, we reduced the available capacity in the field trial area to evaluate the algorithm. When we activated the algorithm, we were able to measure enhancements to quality and capacity in the test area. Capacity improved in all five test cells—that is, congestion was reduced from 5-10% to 1%. We could also measure a significant improvement of the bit-error-rate probability distribution for two of the five test cells.[2] Figure 8 shows the improvement of one of the two cells (eight consecutive 24-hour measurements).

Architecture evaluation

During the field trial, the prototype system was able to monitor 75 of the 105 cells included in the area. However, there was not time enough to optimize performance.

Figure 8
Distribution of bit-error-rate probability for one of five test cells. Activation of the algorithm reduces the bit error rate.



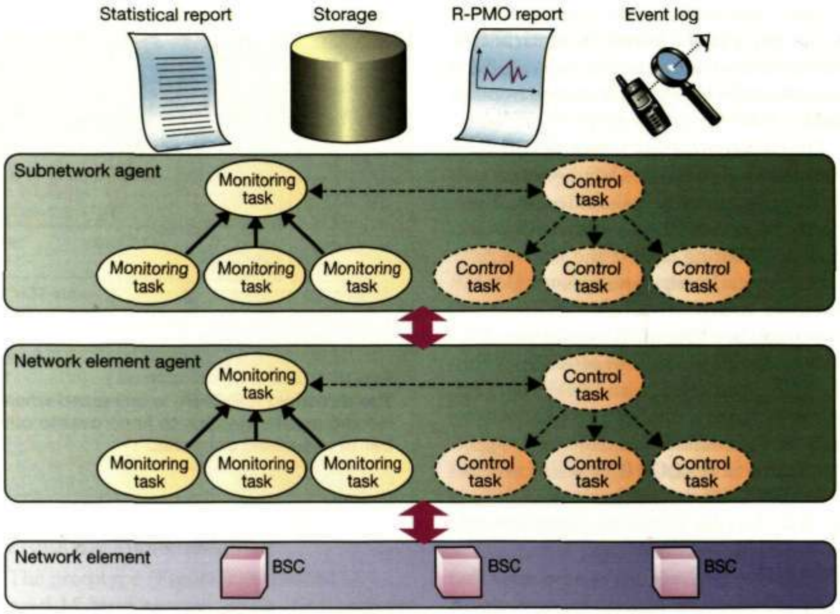


Figure 9
An overview of the structure of the R-PMO.

The field trial has given valuable input that can be applied to future products within this area. For instance we have learned that

- performance-critical parts should be implemented in C++ instead of in Java;
- the need for interprocess communication should be kept to a minimum; and
- when events are transmitted using CORBA calls they should be sent in batches (several events transmitted in the same message).

Latency measurements indicated that it took from 10 to 35 seconds for an event in the network (for example, a dropped call) to be measured (percentage change in dropped calls) and displayed in the GUI. In most cases a latency of less than one minute is acceptable.

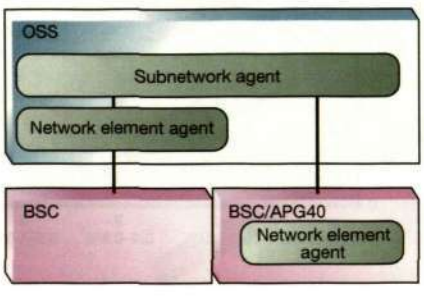


Figure 10.
Options for distributing R-PMO functions.

From ideas to products

Based on input from previous product experience as well as on the results from the prototype and field test, Ericsson is now introducing event-based reporting mechanisms into GSM (Release 9.1) and WCDMA (Phase 2).

Event mechanism in GSM

An event mechanism is being introduced into the BSC in Ericsson GSM BSS R9.1. The mechanism permits the operations and support system (OSS) to subscribe to real-time event data. The events are sent to the OSS over a TCP/IP connection.

In addition, a real-time performance monitoring (R-PMO) application is being introduced into the OSS to process real-time event data. The R-PMO provides real-time presentation of several basic monitors related to the performance of the radio network. Examples of data presented per cell are traffic level, TCH utilization, voice quality, and basic handover information. End-user reports provide an overview and detailed analysis of individual values.

The implementation of the R-PMO application has been greatly facilitated by the monitoring tasks (Figure 9), which perform simple and well-defined functions. The R-PMO implements a monitoring task framework that supports the administration of tasks. The framework also provides several basic services, such as giving access to configuration information, subscription mechanisms that allow monitoring tasks to subscribe to information from lower-level monitoring tasks or raw performance events. New monitoring tasks can be added to the OSS with little extra effort.

Control tasks are not currently supported, but the framework has been designed with them in mind.

The R-PMO server software has two main components—a subnetwork agent, which handles the overview of the network at the OSS level, and a network element agent, which is responsible for processing the data from a single network element.

The amount of data that is sent from the BSC to the OSS is relative to the traffic volume. For large BSCs this can amount to several hundred kbit/s during peak hours. The data format used on the internal BSC-OSS link is proprietary to Ericsson. However, external applications can access the event data and also monitor values via an open PM IRP interface in the OSS.

The APG40, a new Windows NT-based input-output processor for the BSC, can be used to move part of the R-PMO functionality to the BSC itself (Figure 10). The distribution of functions to the BSC makes it possible to reduce the load on the OSS-BSC link significantly. The APG40 deployment is not part of the R9.1 version of R-PMO.

Event mechanism in WCDMA

In Ericsson's WCDMA products (RNC, Node B and RANOS), general performance event handling (GPEH) is considered to be a cornerstone for providing real-time performance monitoring capabilities in future WCDMA releases (Figure 11).

The collection and recording of performance events is instrumented by monitoring tasks which are accessed through management interfaces (*Mun*, *Mur* and *Mub*) that support the CORBA solution set of the performance and notification integration reference points (IRP).

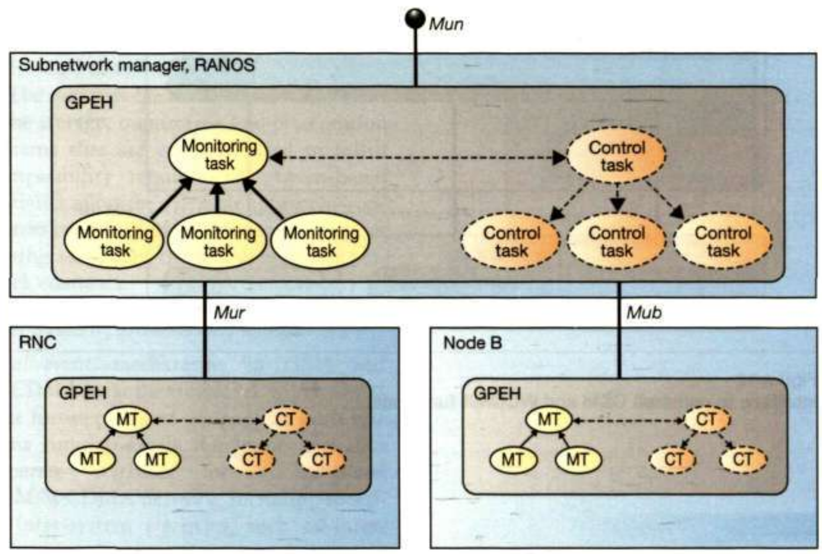


Figure 11
Overview of the GPEH.

Thanks to the GPEH function, operators can create and modify monitoring tasks in a flexible way. Individual monitoring tasks can be defined to collect an arbitrary set of performance events for an arbitrary part of the radio network. Different monitoring tasks can be controlled as separate entities. In the current implementation, the performance events are available as periodic recordings written to a file. Subscribers are alerted to the availability of a new performance event file through the *Mu* interfaces.

The architecture on which the GPEH function is based provides a solid base for future solutions, including the near-real-time delivery of performance events from network elements. Similarly, control tasks can be allocated to network elements; more complex control tasks can be allocated to subnetwork managers.

Combining GSM and WCDMA

In the not-so-distant future, many mobile networks will contain GSM and WCDMA nodes. Interworking will be crucial, especially when the WCDMA part is new and has limited coverage. The management system will need to monitor both types of node. Special emphasis will be put on providing

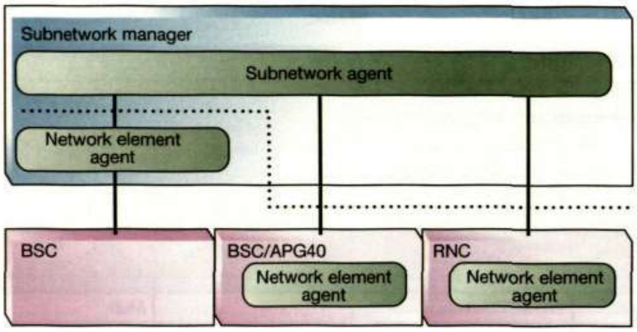


Figure 12
Interface to common GSM and WCDMA functions.

tools for monitoring and tuning intersystem behavior, such as handover.

In terms of type of available event and architecture, GSM and WCDMA event mechanisms share many common principles. This means that the existing mechanisms provide a good base on which to build applications that support GSM and WCDMA. The event-reporting mechanisms in GSM and WCDMA can monitor

- control functions (connection handling);
- mobility functions (handover, cell update);
- capacity functions (admission, congestion);
- bearer functions (data throughput);
- measurement reports;
- configuration management functions; and
- inter-node communication.

GSM and WCDMA each use subscription-based principles for reporting events. The applications can thus specify which events are to be activated. Moreover, the mechanisms support the IRP concept.

Applications

The event mechanisms in GSM and WCDMA provide a solid base that can be used for creating a wide range of applications.

Performance monitoring

The R-PMO application provides immediate feedback on the performance of the network. Monitors, such as counter values and gauges, can be shown with low resolution and very short delay. Operators can thus react quickly to problems in the network and see the immediate effect of configuration updates. This is useful for troubleshooting and can also be used when new features are being introduced and tuned, and when parameters are being changed. The R-PMO application includes a range of predefined reports on traffic load, quality, and so on. Real-time performance events could also be considered as enablers of a wide range of performance monitoring services, such as

- alarms that supervise traffic conditions in the network;
- the identification of specified situations in the network that depend on the correlation between events. This gives almost unlimited possibilities of finding specific behaviors in the network. Examples are identifying cells with a large amount of fast moving mobile stations or ping-pong handover situations; and
- the monitoring of end-user performance to determine whether or not the customer gets what he pays for.

Troubleshooting

Event data is a powerful tool for troubleshooting. The mobile traffic recording (MTR) application for GSM and the user equipment traffic recording (UETR) application for WCDMA allow operators to record the traffic behavior of select mobile stations, including measurement data relating to the air interface. This data can be used to check parameter settings in the network (for example, handover parameters).

The cell traffic recording (CTR) application for GSM and the corresponding application for WCDMA allow operators to record the traffic behavior of mobile stations in a specific recording area (for instance, a given set of cells).

Optimizing applications

The frequency allocation support (FAS) and frequency optimization expert (FOX) applications in GSM help operators to choose

TRADEMARKS

which frequencies should be allocated to the cells in their GSM networks. FAS presents the results in reports, so that the operator can decide which frequency changes to implement, whereas FOX automatically finds, suggests and implements frequency changes that will improve network quality.

The neighboring cell support (NCS) and neighboring optimization expert (NOX) applications give operators a way of optimizing the neighboring cell lists for all cells in their GSM network. NCS and NOX help operators to add and remove cells from the neighboring cell lists that every GSM cell needs to have in order for handovers to work between cells. NCS presents the recorded measurements and statistics in several reports, which the operator can study to see which neighboring cell relations should be removed and added. NOX has the added functionality of suggesting which neighboring cells should be removed or added, and of implementing these changes automatically. Real-time performance events could also be an enabler for a wide range of optimization services, such as

- neighbor cell list optimization for combined GSM and WCDMA networks;
- optimization of power settings; and
- loadsharing between cells, frequencies and systems.

Event-based statistics

Ideally, to preserve flexibility, the software that manages the radio network should not also have to calculate statistics. As the radio network is enhanced (has more features added to it) it becomes more complex. The addition of a counter-based interface to the switch requires a massive addition of new counters—to keep network performance visible. The same results can be achieved with a small set of performance events. Events contain detailed information about individual mobile stations and can be correlated with network elements, cells, parts of cells, and even individual calls or phone types.

Extracting data from the traffic machine also gives operators the opportunity to gather and correlate data from more than one network element. This gives performance data on relations between cells served by different network elements and different systems, such as GSM or WCDMA. New statistics can quickly be introduced and tailored to individual customers' needs.

The statistics can then be processed by the same storage, monitoring and presentation systems that are currently used to fulfill compatibility requirements. Event-based statistics allow for a flexible approach—operators can easily add new counters without having to update the software in the network elements.

Inter-system performance management

The event mechanisms in GSM and WCDMA are at the same level, which means that future products can support both systems simultaneously. Operators can thus generate statistics for a combined GSM/WCDMA network, including specific inter-system statistics, such as inter-system handover.

Moreover, operators will be able to adaptively tune inter-system aspects, such as inter-system neighbor cell lists.

External applications

The use of event data is not limited to the applications described above, but can also be used to create new functions that help operators to increase their revenue. The event data contains detailed information about the type and distribution of traffic in the network. When combined with other external data this can serve as the basis for services provided by the operator or a third party. One such example is road traffic information.

Conclusion

This article describes an architecture and prototype based on real-time performance events for monitoring and optimizing cellular networks. The events-based architecture admits fast and flexible development of performance monitoring and optimization applications. It also provides open interfaces that enable customized performance monitoring.

Existing mechanisms in GSM and WCDMA products adhere to many of the architectural principles described, which in turn allow efficient monitoring of combined systems.

REFERENCES

1 HYPERLINK3GPP TS 32.102 V4.2.0 (2001-09) 3G Telecom Management Architecture ftp://ftp.3gpp.org/specs/2001-09/Rel-4/32_series/32102-420.zip
2 ETSI GSM Technical Specification 05.08

IP Technology in WCDMA/GSM core networks

Heino Hameleers and Christer Johansson

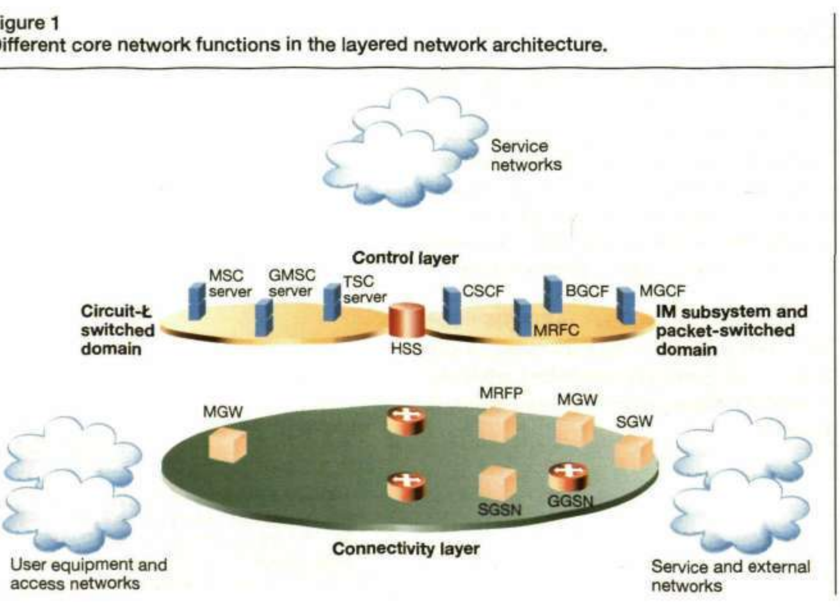
Mobility and the Internet, the two most dynamic forces in communications today, meet in the design and implementation of the mobile core network. Support for new end-user services and a common transport technology are the main drivers of integration of IP technology into our systems. The IP multimedia application plays a special role in providing these new end-user services. IP transport technology addresses the vision of multi-service backbone networks, based on a single network layer technology. Ericsson provides complete solutions and products to support deployment of new IP-based services and transport networks. Moreover, Ericsson's flexible core network architecture allows operators to address these drivers in an independent way.

The main parts of this article describe how the requirements for the two main drivers for IP technology are met in the mobile core network. Paying special attention to support for the IP multimedia application, the authors describe how support for IP applications is implemented. They then describe how IP transport technology can be supported, including site configurations and specific issues like quality-of-service and network redundancy.

Introduction

The marriage of mobile communication and the Internet has the potential to produce a revolution whose scope far outpaces that associated with the advent of the personal computer (PC). In parallel with the fantastic worldwide growth of mobile subscriptions, the fixed Internet and its service offerings have grown at a rate far exceeding all expectations. And this is only the beginning—the future will be even more spectacular given that the number of people connected to the Internet will continue to increase and that GPRS and WCDMA mobile networks will enable connectivity virtually everywhere and at any time with any device. A new form of interactive communication behavior is emerging from the combination of different media with applications that are randomly invoked by multiple users and end-systems.

Figure 1
Different core network functions in the layered network architecture.



Drivers of IP technology

There are two main arguments that drive the integration of IP technology into mobile core networks:
- support for (new) IP applications to generate (new) revenues; and
- a common transport technology to reduce costs.

The entire mobile telecommunications industry is funded out of the end-user's pocket. Therefore, to ensure future growth in the industry, end-user value needs to be enhanced. Service and application offerings are the prime drivers of the entire network and terminal evolution. End-to-end IP solutions must have access to the dynamic IP applications industry. This is the main prerequisite for success in the world of third-generation mobile networks. In this world, visualization and real-time behavior are fundamental components for bridging distance and for giving the end-user experience of "presence" a broader spectrum of human senses.

IP is gradually becoming a dominating transport technology thanks to recent advances in optics and routing technology and the impact that these have had on price/performance. When combined with other key technologies, such as IP-based virtual private networks (VPN), IP enables a new generation of advanced multiservice networks. The use of a common infrastructure based on a single technology simplifies network implementation and operation and helps reduce costs. Ericsson's core network architecture and product solutions allow operators to address these two main drivers of IP technology individually.

Core network overview

The mobile core network is at the heart of present-day mobile communication networks. It provides support for network features and telecommunications services, including essential functions such as session and call control, charging, mobility, and security. Because of its central role in the overall architecture, the mobile core network interfaces with and coordinates other network elements, including user equipment, radio access networks and service networks.

Core network architecture

Ericsson's core network solution is based on the separation of functionality into a control layer, a connectivity layer, and an application layer. The control layer hosts network control servers that are in charge of call or session set-up, modification, and release. The control servers might also handle mobility management, security, charging and interworking functions that relate to external networks at the control plane level.

The connectivity layer hosts routers, switches, signaling gateways, media gateways and other user-plane functions. Its routers and switches provide transport capabilities for traffic on the control and user planes. The media gateways facilitate interworking on the user plane. This includes interworking between different transmission technologies and media formats.

The interface between the control layer and the connectivity layer mainly consists of gateway control protocols. The network control servers use these interfaces to manipulate media gateway resources in the connectivity layer.

The application layer, which is implemented as part of the service network, hosts application and content servers. There are two interfaces between the core network and the service network: a horizontal interface and a vertical interface.

The horizontal interface between the core network and the service network refers to regular peer-to-peer or client/server mode of operation for typical end-user applications, such as Web browsing, e-mail and audio/video services. These applications are normally invoked by an end-user but might also be invoked by an application server.

The vertical interface allows applications that reside on specific application servers to complement or modify the normal procedures for setting up calls or sessions through the core network. These applications interwork with the core network through a set of standardized application program interfaces (API).

The layered architecture allows each layer to evolve independently and in pace with the evolution of the market and technology. It also supports the migration to new transport technologies, since the upper layers are independent of the transport technology deployed in the connectivity layer. The layered architecture also allows different, optimized technologies to be deployed in the connectivity layer (which is payload-processing-intensive) instead of in the control layer (which is transaction-oriented). Figure 1 shows how the different core network functions fit into the layered architecture.

In the circuit-switched domain, the mobile services center (MSC), gateway MSC (GMSC) and transit services switching center (TSC) servers are part of the control layer.

BOX A, ABBREVIATIONS

3GPP	Third-generation Partnership Program
AAA	Authentication, authorization and accounting
AF	Assured forwarding
AMR	Adaptive multirate
ATM	Asynchronous transfer mode
AUC	Authentication center
BE	Best effort
BGCF	Breakout gateway control function
BGP	Border gateway protocol
BICC	Bearer independent call control
BSC	Base station controller
CN	Core network
CPP	Cello packet platform
CS	Circuit-switched
CSCF	Call session control function
CSE	Customized applications for mobile network-enhanced logic (CAMEL) service environment
DNS	Domain name server
DS	Differentiated services
DSCP	DS code point
ECMP	Equal cost multipath routing
EF	Expedited forwarding
FTP	File transfer protocol
GGSN	Gateway GPRS support node
GMSC	Gateway MSC
GPRS	General packet radio service
GSM	Global system for mobile communication
GSN	GPRS support node
GSTN	General switched telephone network
GTP	GPRS tunneling protocol
HLR	Home location register
HSS	Home subscriber server
HTTP	Hypertext transfer protocol
I-CSCF	Interrogating CSCF
IETF	Internet Engineering Task Force
IMS	IP multimedia subsystem
IP	Internet protocol
IPSec	IP security
ISC	IP multimedia service control
ISDN	Integrated services digital network
ISP	Internet service provider
ISUP	ISDN user part
LAN	Local area network
LER	Label edge router
LSP	Label switched path
MAP	Mobile application part
MGCF	Media gateway control function
MGW	Media gateway
MPLS	Multiprotocol label switching
MRF	Media resource function
MRFC	MRF control part
MRFP	MRF processing part
MSC	Mobile services center
O&M	Operation and maintenance
OSA	Open service architecture
OSPF	Open shortest path first
PBN	Public backbone network
PCM	Pulse code modulation
P-CSCF	Proxy CSCF
PHB	Per-hop behavior
PLMN	Public land mobile network
PS	Packet-switched
QoS	Quality of service
RAN	Radio access network
RFC	Request for comment
RNC	Radio network controller
RTSP	Real-time streaming protocol
SCS	Service capability server
S-CSCF	Serving CSCF
SDH	Synchronous digital hierarchy
SDP	Session description protocol
SGSN	Serving GPRS support node
SGW	Signaling gateway
SIP	Session initiation protocol
SLF	Subscriber location function
SSF	Service switching function
STM	Synchronous transfer mode
TDM	Time-division multiplexing
TSC	Transit services switching center
UE	User equipment
UMTS	Universal mobile telecommunications system
VLAN	Virtual LAN
VPN	Virtual private network
WAN	Wide area network
WAP	Wireless application protocol
WDM	Wavelength division multiplexing

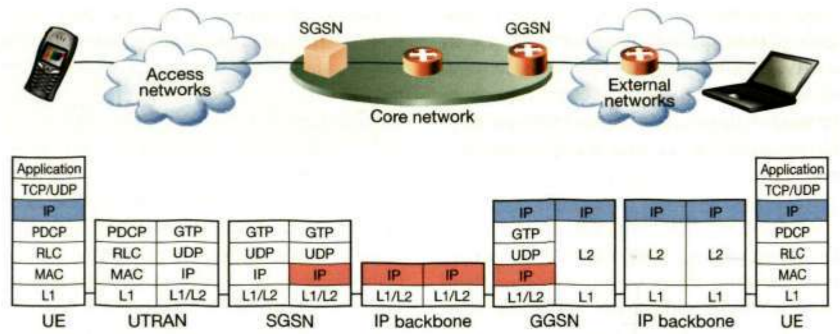


Figure 2
End-to-end protocol stack for an IP application that runs on top of a WCDMA/GSM packet-switched network.

The corresponding media gateway belongs to the connectivity layer.

In the packet-switched domain, both the serving GPRS support node (SGSN) and the gateway GPRS support node (GGSN) are considered to be part of the connectivity layer—they contain some control functionality, but the dominant functionality lies in providing IP connectivity.

With 3GPP Release 5, one more "domain" is being added to the mobile core network: the IP multimedia subsystem (IMS). The principal network entities of the IMS are the

- call/session control function (CSCF);
- media gateway control function (MGCF);
- breakout gateway control function (BGCF);
- media resource function (MRF) control part, or MRFC;
- MRF processing part (MRFP);
- media gateway (MG); and
- signaling gateway (SG).

The master subscriber database, called the home subscriber server (HSS), is common to the circuit-switched domain, the packet-switched domain, and the IP multimedia subsystem.

The two roles of IP

The two roles that IP plays in the mobile core network are also expressed in the core network protocol stacks. Figure 2 shows an end-to-end protocol stack for an IP application running on top of a WCDMA/GSM packet-switched network. The traffic leaves the mobile core network at the GGSN. There are also other scenarios in which traffic from the GGSN remains in the core network. However, for the sake of describing the roles of IP technology in the mobile core network a simplified scenario is used.

Two separate IP layers can be identified. The upper layer (drawn in blue) denotes the IP application layer, which runs between the user equipment (UE) and an external entity with which the UE is communicating. This would typically be an IP application server or another UE.

The lower layer (drawn in red) denotes the IP transport layer, which has only local significance to the public land mobile network (PLMN). The IP transport layer is needed to transport (control and user plane) traffic within the mobile network. In this particular case, the layer is terminated at the GGSN, where the traffic leaves the PLMN; routing is performed directly on the IP application layer. In other cases, IP application-layer traffic might continue to be carried over an IP transport layer even after it leaves the GGSN.

Support for IP applications

Role of the core network in providing IP applications

With respect to the IP application layer, the role of the mobile core network has traditionally been limited to providing a tunnel that allows the UE to communicate with another IP host. This support is implemented in the GPRS support nodes (GSN).[3]

In general, IP applications are transparent to the core network. This is true for all IP applications but one: the session initiation protocol (SIP) application. The main difference compared with other IP applications is the communication model. Most IP applications target a client-server model. The file transfer protocol (FTP) allows a client to download files from a server, the hypertext transfer protocol (HTTP) and

BOX B, HISTORY OF IP MULTIMEDIA STANDARDIZATION

Support for IP multimedia applications in mobile core networks was first discussed in 1999 in the 3G.IP forum. The 3G.IP forum was an industry consortium initially consisting of eight of the main operators and vendors, including Ericsson. It had set itself the goal of defining an IP technology based architecture for the next generation of mobile networks that would support voice, data and multimedia services.

This network architecture proposal was brought into the 3GPP forum. 3GPP has accepted the proposal and has since spent considerable effort to define a complete end-to-end architecture, including solutions for essential functions such as security, charging and QoS. 3GPP selected SIP as the session control protocol and has also mandated IPv6 for IP multimedia applications.

wireless application protocol (WAP) allow clients to download electronic pages with content from a server, and the real-time streaming protocol (RTSP) allows clients to "stream" content from a server. By contrast, SIP primarily targets a client-to-client communication model.

How is SIP supported in the mobile core network? The main parts are defined as the IP multimedia subsystem (IMS). Together with the circuit-switched domain and the packet-switched domain, the IMS builds the 3GPP mobile core network.

IMS: providing IP session control

To understand the IMS architecture, one must first understand the basic concepts on which the architecture has been built: the home-visited-interworking architecture and functional entities, and services.

Home-visited-interworking

In second-generation mobile systems, such as GSM, services are provided by the PLMN in which the subscriber is roaming. Personalized service information is transferred from the home PLMN to the visited PLMN. This approach requires that both the home PLMN and the visited PLMN support the service to be provided. That is, the services that can be provided to the end-user represent the lowest common denominator of what the home PLMN and the visited PLMN can support.

In the future, the differentiation between network operators will be made at the application and service levels, instead of at the access and network levels. Roaming subscribers will no longer be the exception, but will have become the norm. This implies that the provision of seamless services for roaming subscribers is increasingly important.

Ericsson has proposed the concepts of *home* and *visited* to describe the architecture of modern communication networks. *Home* denotes user data and services, whereas *visited* denotes connectivity and mobility. This implies that the main task for the visited network is to provide a subscriber with (mobile) connectivity to the home network (Figure 3). The home network hosts user data, session control and services.

Acceptance of these concepts implies that subscribers are always roaming in a visited network. However, the services are controlled from the home network, regardless of which visited network the subscriber is roaming in.

This approach limits the functional and protocol dependency between the home and visited networks, thereby

- minimizing the restrictions put on the services that can be deployed in the home network; and
- increasing the rate at which services can be deployed.

In addition to the personalized services provided by the home network, the home and the visited networks can each provide local services—however, these services are not tied to the user profile in the home network.

Although this implies that control signaling must always go through the home domain, the actual payload must not. Payload is routed independently of control signaling and can follow the optimal (shortest) path for efficient transmission and optimal quality of service (QoS).

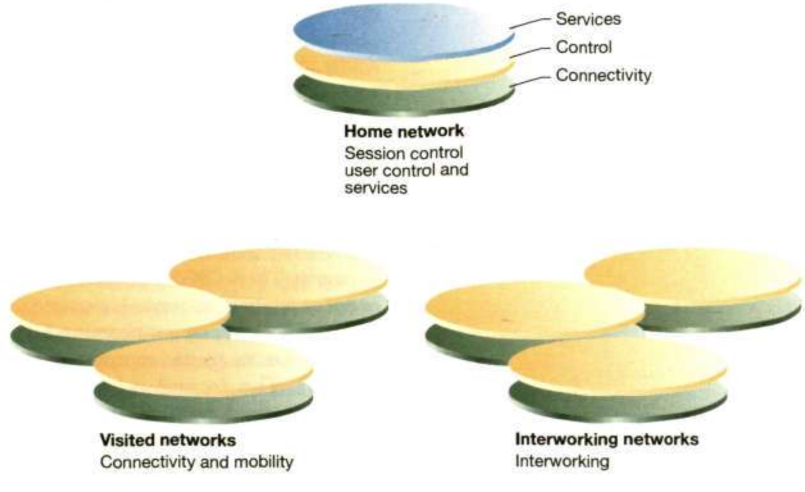
If the IMS is to interconnect with legacy networks, some functionality must be provided to enable the systems to interwork. To allow for the optimal user-plane transport path (for example, by keeping the session in its own network as long as possible), this functionality can be placed in an interworking network outside of the home and visited networks. Home, visited and interworking networks can be physically different networks or they can be implemented in one and the same network.

BOX C, IP MULTIMEDIA PROTOCOLS

The session initiation protocol (SIP) has been defined in IETF RFC 2543. It describes a way of supporting session-based applications over IP networks that involve one or more participants. In mobile terms this means that it allows a mobile client to set up an IP session to another mobile client. An updated version of SIP is expected in 2002.

SIP "uses" the session description protocol (SDP) to describe the nature of the multimedia sessions—that is, which media are included in the session, in which format the media will be transported, and so on. SDP has been defined in IETF RFC 2327.

Figure 3
The visited network provides connectivity to the home network.



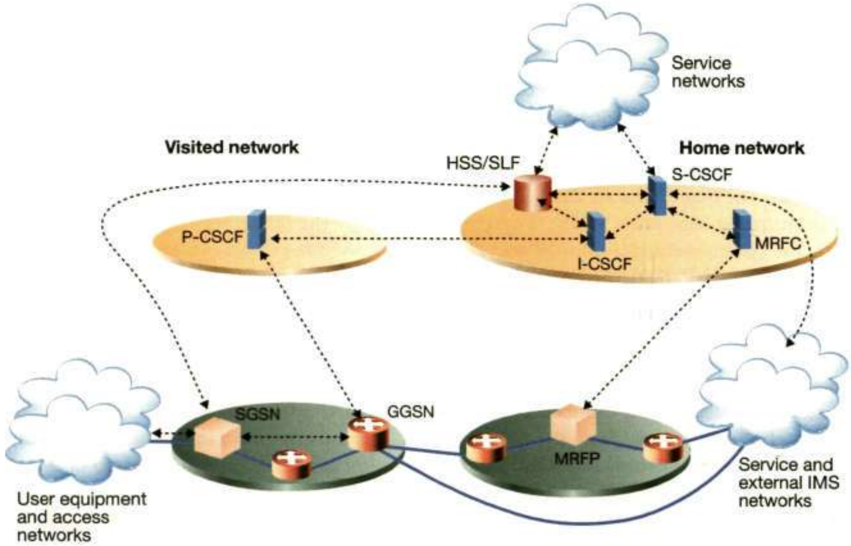


Figure 4
Basic IMS architecture.

Architecture and functional entities

The IP multimedia system is built around the call/session control function, of which there are three different kinds:

- the interrogating CSCF (I-CSCF);
- the proxy CSCF (P-CSCF); and
- the serving CSCF (S-CSCF).

The P-CSCF is the UE's first point of contact with the IMS. The P-CSCF forwards SIP messages received from the UE to a SIP server in the home network (and vice versa). The P-CSCF might also modify an outgoing request according to a set of rules defined by the network operator (for example, address analysis and potential modification).

The I-CSCF function, which forms the entrance to the home network, hides the inner topology of the home network from other networks and provides flexibility for selecting an S-CSCF.

The S-CSCF performs the session control services for the UE. This includes routing originating sessions to external networks and routing terminating sessions to visited networks. The S-CSCF also decides whether or not an application server is required to receive information on an incoming SIP session request to ensure appropriate service handling. This decision is based on information received from the HSS (or other sources, such as an application server).

All CSCF functions can generate call detail records for input to the charging process.

The HSS, which is an evolution of the home location register (HLR) and authentication center (AUC), holds the subscriber profile and keeps track of which core network node is currently handling the subscriber. It also supports subscriber authentication and authorization functions (AAA).

In networks with more than one HSS, the subscriber location function (SLF) provides information on the HSS that contains the profile of a given subscriber.

The media resource function (MRF), which contains the functionality for manipulating multimedia streams, supports multiparty multimedia services, multimedia message playback and media conversion services. The Third-generation Partnership Project (3GPP) has split the MRF into a control part (MRFC) and a processing part (MRFP). Figure 5 depicts a scenario in which a multimedia session interworks with a general switched telephone network (GSTN).

The BGCF selects the network in which the interworking is to be performed. If the interworking is performed in the home network, the BGCF selects an MGCF. If the interworking is to be performed in another network, the BGCF selects another BGCF or an MGCF.

The MGCF provides interworking func-

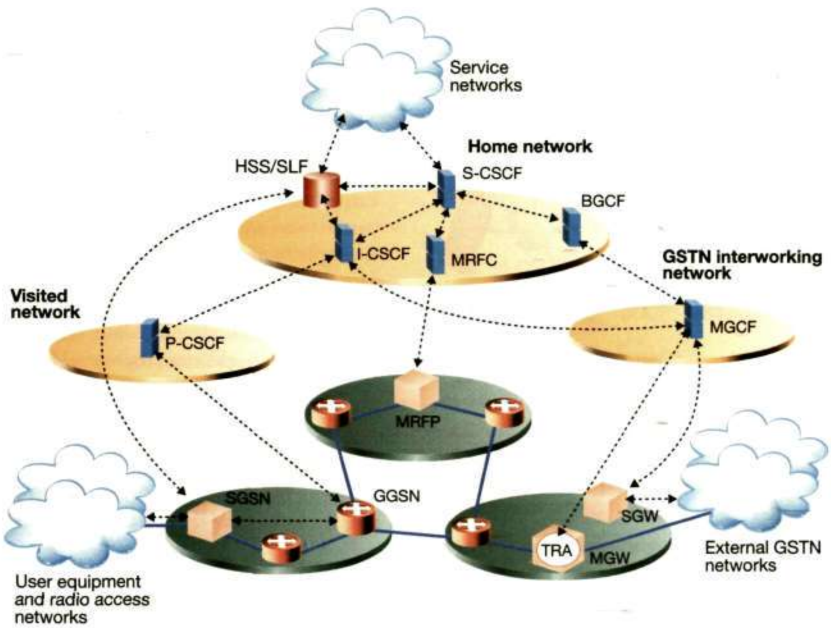


Figure 5
A multimedia session interworks with a general switched telephone network (GSTN).

tionality between SIP session control signaling from the IMS and ISUP/BICC call control signaling from the external GSTN networks. It also controls the media gateway that provides the actual user-plane interworking functionality (for instance, for converting between AMR- and PCM-coded voice). The signaling gateway provides bearer interworking functionality for the control signaling (ISUP/IP – ISUP/TDM).

Services

The 3GPP is working to define the IMS architecture, but its concepts and protocols were derived from the Internet Engineering Task Force (IETF). The IMS enables the convergence of, and access to, voice, video, messaging, data and Web-based technologies for the wireless user. The main tools that the IMS provides to build these services are

- peer-to-peer addressing architecture for IP-based sessions;
- flexible integration of "any" type of media into sessions;
- integration with other IP applications, such as RTSP, HTTP, and so on; and
- end-to-end QoS charging and security architecture.

Although the IMS mainly targets client-to-client services, its tools can be used to facilitate or enhance client-server-based services, such as those found in gaming. Moreover, besides providing real-time services (such as video-conferencing), the IMS provides non-real-time services (such as instant messaging).

Unlike second-generation mobile systems (such as GSM), specific IMS services will not be standardized. The 3GPP is merely defining the architecture framework and service capabilities that can be used to build services. The actual services are implemented on top of these capabilities by network vendors, operators or third parties.

In keeping with IETF principles, the endpoints of the system contain considerable intelligence for supporting services with little or no assistance from the network. However, there are also scenarios in which the network provides value-added services—for instance, to provide a presence service or to optimize the use of resources for conferencing services.

To support these network-controlled or network-assisted services, the 3GPP is defining an IMS service creation environment. The S-CSCF and the HSS each feature one or more vertical service creation interfaces, and the 3GPP is deliberating whether or not it will also include them in the MRFC.

Horizontal interfaces might also be considered. A horizontal interface would apply to a scenario in which the S-CSCF routes control of the user plane to an external ap-

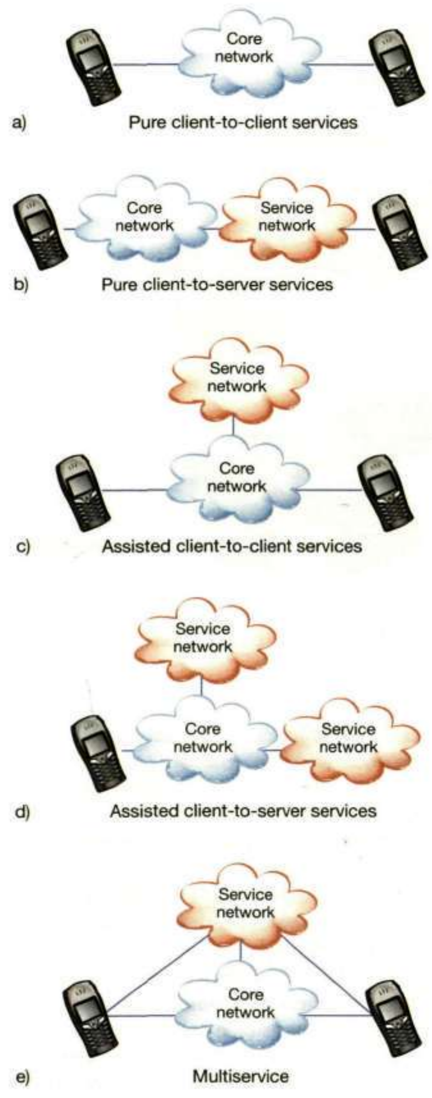


Figure 6
Overview of five service execution scenarios.

plication server. In that case, the application server would have full control over session routing and media. Figure 6 gives an overview of five service execution scenarios:

a) The main service functionality lies in the user equipment (UE). The core network can provide assistance for some value-added services. Example: a video-phone session between two clients.
b) The core network passes the session via a horizontal interface to an application server. Example: a gaming session.
c) The core network passes session control via a vertical interface to an application server. After session control has been returned to the core network, the session is established to the terminating user equipment. Example: a voice session with session forwarding.
d) The core network passes session control via a vertical interface to an application server. When session control has been returned to the core network, the session is passed to another application server. Example: a gaming session in which the service network determines the closest gaming server.
e) The main session runs between user equipment. In addition, there may be one or more sessions running between the user equipment and the service network. The user equipment coordinates these multiple sessions. Example: a voice session during which both clients view the same streaming video.

Core network products

Given the complex functional architecture of the IMS, careful consideration ought to be given to the mapping between functions and actual products. The result of implementing each function as a separate physical box would be a complex system that is extremely hard to operate.

In all probability, the first commercial IMS products will be deployed on a small scale. Therefore, cost-efficient, entry-level solutions should exist that provide enough capacity to handle the initial IMS traffic. As the market evolves, more flexible and powerful configurations will also be required.

The Ericsson IMS product line will fulfill near-term and long-term requirements by implementing the IMS functions in an integrated yet modular way, by

- adhering to a horizontally layered architecture;
- providing scalability from entry-level (integrated) to high-end (distributed) configurations;
- using a modular software architecture (for example, separate S-CSCF, I-CSCF and P-CSCF modules) that allows for distributed solutions;
- implementing servers on TSP; and
- implementing gateway and user-plane functions on CPP.[4]

Figure 7 shows some example IMS product configurations. The trial site provides full IMS functionality, including support for video-conferencing and interworking with GSTNs. It also provides a local service execution environment on the CSCF.

The figure also shows some long-term

configurations that enable handling larger volumes of IMS traffic. In this particular scenario, we can differentiate access, service and gateway sites. The access sites can host P-CSCF and possibly also I-CSCF functionality to handle the interface to the UE and to forward sessions to the appropriate service sites.

The service sites host the functions that provide the actual session control and services. The gateway site handles the interface to external GSTNs.

Figure 8 shows only a handful of many possible configurations. Given the flexibility of the IMS product offering, many other configurations can be created to suit specific network deployment scenarios.

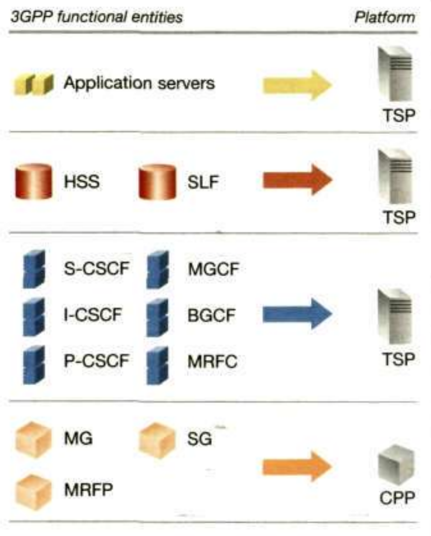


Figure 7
Example IMS product configurations.

IP connectivity

General

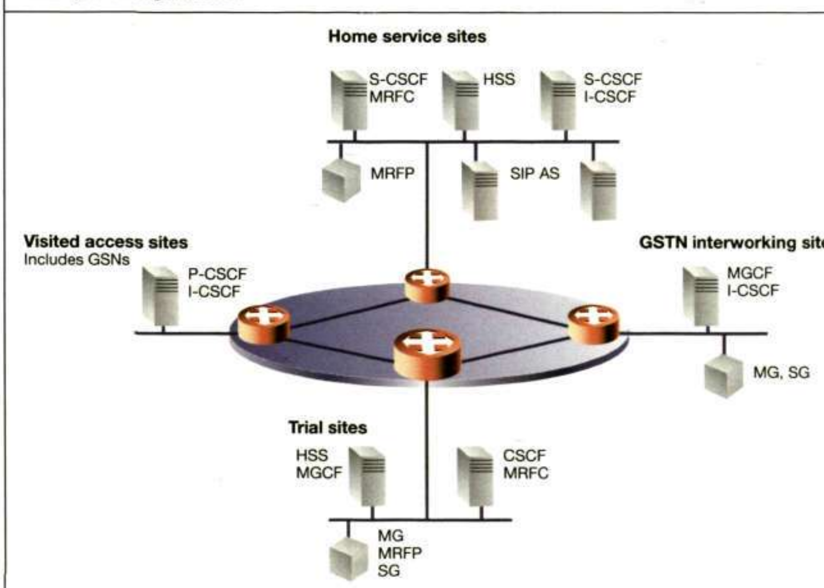
Most mobile operators have introduced (or are in the process of introducing) packet-based transport technologies into their networks. IP networks are being introduced for the intranet when WAP servers are introduced—following the introduction of GPRS—and when operators become Internet service providers (ISP). At the same time some operators are introducing asynchronous transfer mode (ATM) backbones to decrease transport costs for traditional voice services. Ordinarily, this is done at the transit level of the network.

More and more operators are seeing a need to harmonize the introduction of packet-based networks and have concluded that the best move is to invest in a single multiservice network. But the question remains: which basic technology should be used, ATM or IP? ATM is a mature technology for traditional voice services and for running IP. But for operators to run IP over ATM they must still invest in routers.

IP is mature technology for best-effort traffic. An operator who wants to deploy IP technology as the basis for a multiservice network that also supports traditional voice services (real-time) must consider several new aspects. At this time, there is no well-established dominant network design for this deployment. Emerging technologies, such as IP-VPNs, differentiated services, and resilience mechanisms, are now available in router products for these kinds of carrier-class IP network. These technologies also provide wire-speed routing for acceptable delay and jitter.

Ericsson's work with these issues has resulted in the multiservice IP network presented in this article. The basic assumption is that an operator who is building a multiservice IP network is not willing to sacrifice the characteristics associated with traditional services that are provided in TDM and ATM networks. We must thus solve the

Figure 8
Example configurations.



problem of running telephony over IP multiservice networks while preserving the characteristics of present-day networks, such as quality of service, congestion control, security, and resilience.

The objective is to run all services over the same IP network, which makes the IP layer the converging layer.

In the near future we expect to see a shift from synchronous digital hierarchy (SDH) to IP on wavelength division multiplexing (WDM). Today there is a lot of surplus capacity in fiber networks. The deployment of solutions based on, for example, Ethernet technologies, instead of SDH on layer 2 can give a dramatic reduction in transmission costs. To realize this potential, the services must be transported in a packet-based form.

IP-based core network architecture

General

In an IP-based WCDMA/GSM core network, all core network elements use the connectivity services of a common IP infrastructure to interconnect user traffic and internal signaling.

A key component of the IP infrastructure is an IP backbone network that is used as a common backbone for WCDMA/GSM services, ISP services or fixed network services (Figure 9). The challenge is to find an IP network solution which integrates security, resilience, QoS, dual-stack IPv4/IPv6, and bandwidth efficiency, and which can handle connection-oriented services in a "connectionless" network. One possible network solution is described below, but variations of this solution are possible (depending on the operator's specific prerequisites).

The structure of the IP infrastructure

The IP infrastructure is made up of two main tiers: a backbone tier, which is used to carry all traffic between sites; and a site infrastructure tier, in which the site IP infrastructure extends IP connectivity to the core network elements at the site. Each site IP infrastructure is attached to the backbone tier through one or more edge routers, which serve as traffic aggregation points and demarcation points between the local IP network and the backbone IP network domain.

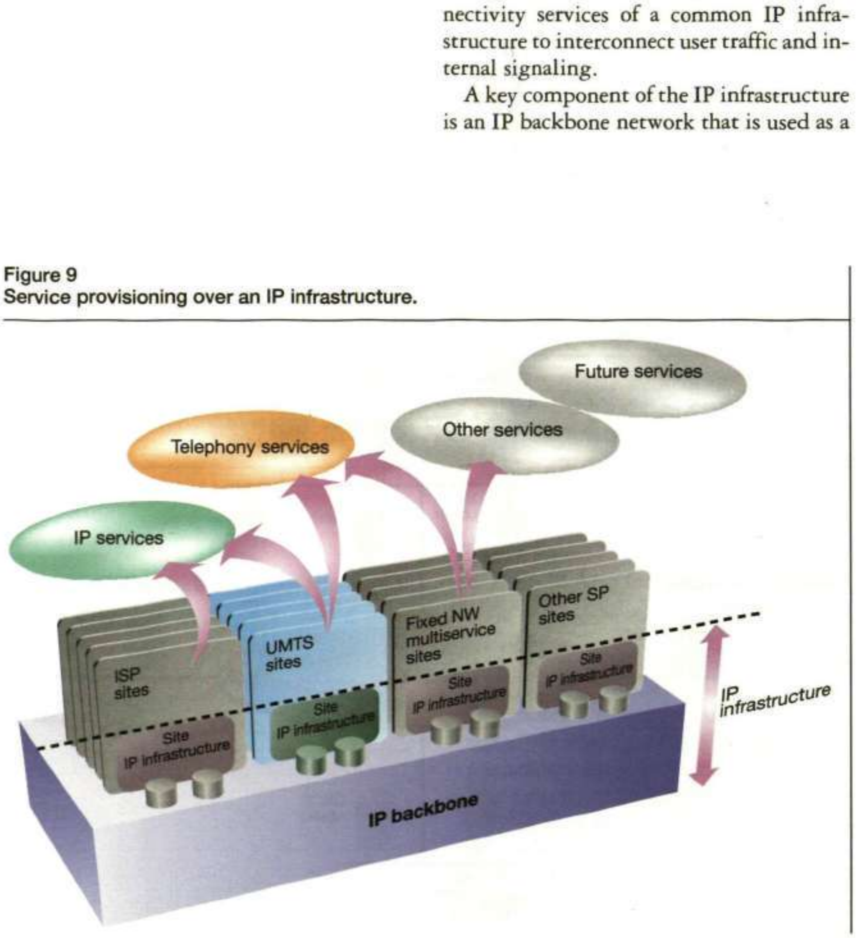
The backbone IP network is a shared network, which also interconnects sites belonging to other service networks (non-WCDMA/GSM networks, such as ISPs or fixed service network providers). The IP backbone tier provides wide-area IP connectivity between sites. It is designed for simple high-speed packet transport and is typically built with large backbone routers interconnected with fast links, such as Ericsson's AXI 520/580 routers interconnected with Gigabit links.

In this article, we assume that the backbone IP infrastructure is run by a single operator. The routers support different classes of service and layer-3 VPNs. They also support layer-2 encapsulation, which can be used in parallel to IP traffic to carry Frame Relay or ATM between sites.

The site IP infrastructure tier extends IP connectivity to the various WCDMA/GSM network elements. Each site has a local IP infrastructure that is connected to the various elements at the site and serves as a liaison to the IP backbone through edge routers. The site IP infrastructure is adapted to low-cost, high-capacity traffic capabilities within the site, typically using Fast Ethernet, Gigabit Ethernet, and LAN switches that can make use of virtual LAN (VLAN) techniques. The site IP infrastructure is duplicated to guarantee full availability.

The edge routers connect the site IP network to the IP backbone. In terms of struc-

Figure 9
Service provisioning over an IP infrastructure.



ture, they belong to the site IP network domain and to the IP backbone domain, and participate in routing protocols in both domains. The edge routers contain advanced functions (for instance, MPLS LER function, 2547*bis*, BGP, and filtering) for defining a service agreement between the site IP network and the IP backbone. A site typically uses a pair of edge routers connected to different core routers in the IP backbone.

Different types of site

In a practical network design the physical equipment is grouped together in sites. Sites can be grouped by type, depending on their role in the network. For the core network, three types of site can capture the needs of an operator. Other types can also be defined to suit specific operator conditions.

- The *primary site* includes a complete set of functions needed for a WCDMA/GSM network (control servers, media gateways, GPRS support nodes, and radio access network controllers). A primary site might also include a service network configuration. To distribute redundant load, a network can have several primary sites.
- The *secondary site* contains media gateways, GPRS support nodes, and radio access network controllers. If necessary, secondary sites can also have peering connections to other networks.
- The *concentrator site*, which includes media gateways and radio access controllers, is used for concentrating load far out in the network.

Peering connections to other networks can be made from any site. Figure 10 shows the mapping of different site types on the IP backbone. The primary site is the most important type. In fact, a complete network can be built exclusively from primary sites. Figures 11 shows example configurations of a primary site.

Logical networks and VPNs

Logical networks

Different kinds of information are exchanged between sets of network elements. Different types of network (STM/TDM, ATM, IP and SS7) can be used to handle different kinds of information flows, each with well-defined quality of service and little or no connectivity between the networks. It is thus possible to have complete separation of traffic between information flows.

In the context of a multiservice IP network, the IP infrastructure must be able to

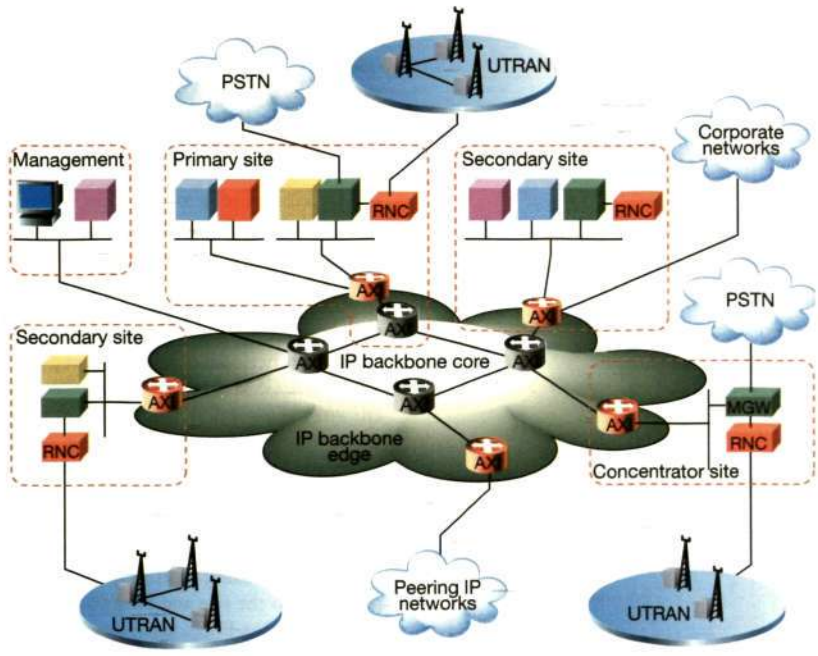
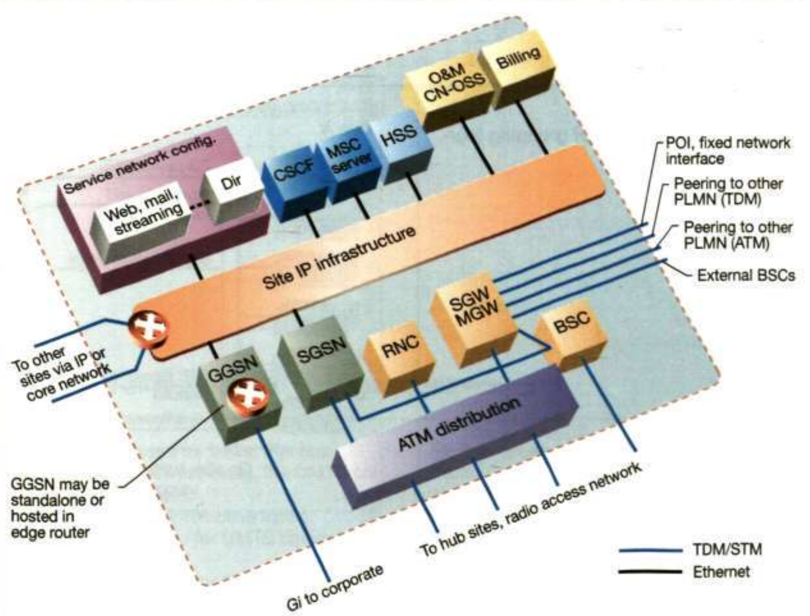


Figure 10
Mapping of different site types on the IP backbone.

Figure 11
Example configurations of a primary site.



handle this exchange of information flows. To facilitate traffic separation and ensure quality of service for the different traffic types, the WCDMA/GSM core network is conceptually divided into a number of logical networks.

Each logical network encompasses a particular kind of information flow between a designated set of functional entities in the WCDMA/GSM network elements. Furthermore, each logical network has a set of requirements with respect to connectivity, QoS, network availability, and so on.

The functional entities reside in WCDMA/GSM network elements at different sites. To support the logical networks, the IP infrastructure is configured into several virtual networks. Logical networks are implemented as virtual networks. Depending on the requirements for the supported logical network, the virtual networks are implemented using an appropriate set of capabilities in the site and backbone IP infrastructures (Box D).

Mapping of logical networks to VPNs

Logical networks can be grouped according to their specific characteristics into virtual networks (security, redundancy and resilience, addressing, QoS, and scaling). This helps operators to decide which VPN technology to apply for each network.

Figure 12 shows one possible implementation. In this example, BGP/MPLS IETF RFC 2547 *bis* is used to separate logical networks into layer-3 VPNs in the backbone. Virtual LAN tagging gives characteristics that are similar in nature to those of an ATM network and gives the operator good tools for controlling the different traffic flows through the IP network. Another simple mapping would be to avoid the use of VPNs and instead use filtering and BGP communities in the routers and client nodes.

Information flows can be encrypted using IPSec in the client nodes or IPSec VPNs between edge routers.

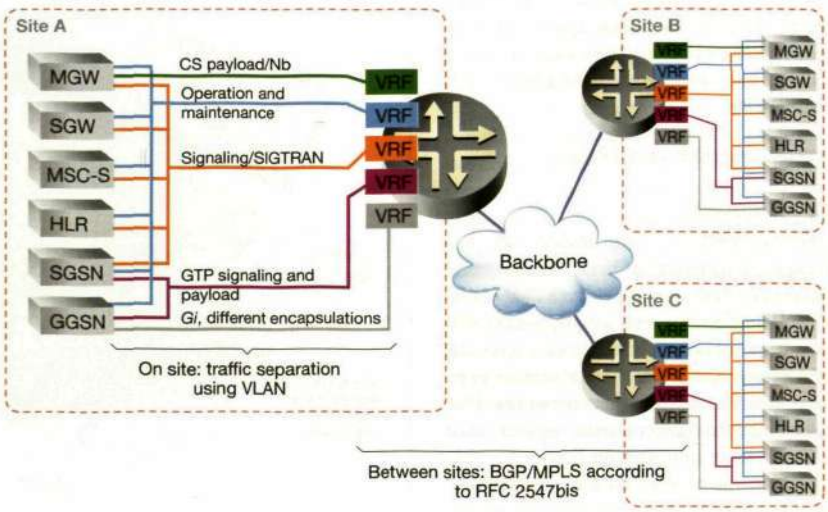
Network redundancy and resilience

The network redundancy principles are based on the assumption that the network should be able to withstand single-failure situations and resume service to its users with very short restoration time. Furthermore, it is assumed that

- the IP infrastructure—that is, the site IP network and the IP backbone network—can be configured to support alternative paths through the IP infrastructure; and
- redundant access can be provided to the infrastructure.

The main routing protocols in the IP backbone will be OSPF or IS-IS and BGP-4. At present, large networks require anywhere from a few seconds to tens of seconds to converge link state protocols, such as open shortest path first (OSPF) and IS-IS. This is too long for time-critical WCDMA/GSM

Figure 12
Possible implementation of grouping logical networks into VPNs.



BOX D, EXAMPLES OF LOGICAL NETWORKS

Signaling network
Used to carry UMTS signaling traffic such as H.248 and BICC signaling, making use of SCTP as transport protocol. The signaling network has application level awareness of alternative addresses to signaling recipients. Traffic volume is relatively low (a few Mbit/s to a site). Signaling traffic must be well protected from external traffic to avoid possibilities of network intrusion.

ISDN voice network
Carries circuit voice and data traffic between media gateways. The traffic is characterized by short packet traffic, with high QoS requirements.

***Gi* toward ISP**
Carries user traffic between GGSN and the Internet. The traffic must be well separated from the internal UMTS traffic. Traffic is currently of best-effort class only, but future traffic can be of any class of service.

***Gi* toward corporate networks**
Carries user traffic between GGSN and an appropriate access point to the corporate network. Separate virtual networks, thus allowing for overlapping addresses, will be required for each corporate network with several access points in the IP infrastructure.

***Gn, Gp* traffic**
User traffic carried in GTP tunnels between GSN nodes. QoS requirements are depending on the carried user traffic.

O&M network
Used for operation and maintenance of UMTS core network components. Very high availability and security requirements exist.

traffic. Therefore, to reduce fail-over times to 50 ms or less, an operator should not solely rely on layer-3 redundancy in the IP backbone. One solution is to rely on underlying SDH mechanisms. Another is to use MPLS with redundant secondary LSPs and the MPLS fast reroute mechanism.

The OSPF protocol with equal cost multipath routing (ECMP) is the recommended method of applying the network-provided multipath principle in the site IP infrastructure. ECMP distributes the traffic between multiple paths between routers and allows fast fail-over.

Quality of service

IP quality-of-service capabilities are implemented using overprovisioning, admission control, and differential services (DiffServ or DS).

Overprovisioning

Within the site, overprovisioning gives simple management and is the cheapest way of guaranteeing QoS. In the backbone, however, other means must be added due to the cost of bandwidth.

The extent to which overprovisioning can be reduced depends on how sophisticated the congestion control mechanisms are. The amount of overprovisioning needed is determined by fault situations (link or router failure), traffic concentration in conjunction with "abnormal" events, and the provision of best-effort capacity (capacity must never be completely starved).

Admission control

Various mechanisms and policies are used for controlling the amount of traffic that is injected into the IP backbone. Admission control is exercised at the edges of the network and serves to protect the backbone from being overloaded. An overloaded backbone results in packet loss and increased delays. There are three main ways of controlling admission:

- admission control in client nodes;
- policing of external interfaces; and
- policing of internal interfaces.

In principle, every node that generates a substantial amount of non-best-effort traffic should perform admission control—for example, GSN, MGW, data-intensive applications servers, such as streaming servers and O&M nodes.

BOX E, USE OF DIFFERENTIAL SERVICES CODE POINTS

EF PHB will be used for traffic that has requirements on lowest delay, the UMTS conversational QoS class.
AF4 is reserved for interactive traffic.
AF3 PHB is proposed for UMTS streaming QoS class.
AF2 PHB is proposed for CS and PS signaling.
BE is proposed for UMTS background QoS traffic.
Network control signaling is assumed to have a separate code point for PHB.
Different kinds of O&M traffic may have different requirements on the DiffServ PHB. Thus multiple DiffServ PHBs will be used for O&M.

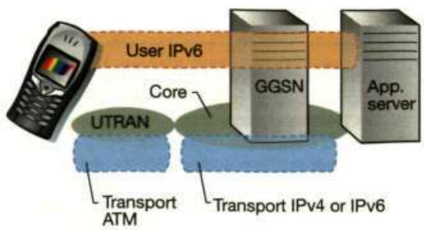


Figure 13
Coexistence of IPv6 and IPv4.

Admission can be controlled in different ways and according to different principles/algorithms. For instance, a trunk-based model can be used for voice traffic between media gateways. The media gateway controls the amount of traffic on the routes to other media gateways. Before new sessions are accepted, the originating node must check that the required bandwidth is available for the destination.

DiffServ

In the proposed QoS solution, differential services constitute a cornerstone for handling quality of service at the IP layer. Various applications in the client nodes and end-user applications or clients in terminals mark IP packets. Examples of applications in client nodes are the ISDN application in the media gateway, the GTP encapsulation function in the SGSN/GGSN, and the SIGTRAN application in the HLR. To differentiate between independent traffic flows, several DiffServ per-hop behaviors (PHB) have been proposed (Box E). When tunneling traffic over GTP, the DSCP in the SGSN and GGSN is marked as follows:

- The DSCP in the end-to-end IP header can be set by the UE.
- In the uplink this setting can be overwritten by the GGSN in accordance with the PDP context (APN) when the DSCP is forwarded over *Gi*.
- The DSCP in the outer IP header is set according to the PDP context (APN) in the SGSN (uplink and downlink) and for the GGSN (downlink).

The different routers are configured to schedule and prioritize traffic packets according to their DSCP. Ericsson's AXI routers and the embedded router in the media gateway and GSN provide rich mechanisms for this.

IPv6

Ericsson's IP solutions support IPv4 and IPv6. To have enough IP addresses for every connected terminal, operators will need to use IPv6 between end-users and applications. IPv4 and IPv6 will coexist in the IP backbone for a long time. Therefore, Ericsson products include dual-stack IPv4/IPv6 implementations.

Ericsson products for the IP infrastructure

The IP solution described above can be supported using Ericsson products. The RXI 820 real-time router capabilities are

TRADEMARKS

NetScreen is a trademark of NetScreen Technologies, Inc.

Extreme Networks is a registered trademark of Extreme Networks, Inc, Santa Clara, California in the United States, and may be registered in other countries.

being embedded in the media gateway (AXM 101) together with SIGTRAN capabilities and signaling gateway functions. The media gateway will also include transformation functions for payload between ATM, TDM and IP. In addition, IP and SIGTRAN capabilities are being included in server nodes, such as AXE.

The packet platform for GSN nodes continues to evolve to include new functionality and improved performance.

For the site IP infrastructure, Ericsson uses products from partners: NetScreen for firewalls, and Extreme for LAN switches. These products are well suited for carrier-class networks with high-availability architecture and hardware-supported filtering and forwarding. Ericsson's partnerships include the identifying and implementing of specific mobile network requirements.

Ericsson IP-Works develops DNS and DHCP products adapted to mobile networks.

Ericsson's partnership with Juniper Networks aims at providing a carrier-class router family with enhancement for the mobile core network. Ordinarily, edge routers are based on the AXI 520 series (equal to Juniper M20, M40 series). A new GGSN product J20 has been developed on the Juniper router platform. Either the AXI 520 or AXI 580 (equal to Juniper M20, M40, M160) can be used for the core routers in the backbone. The partnership between Ericsson and Juniper combines unique competence in mobile systems with that of building carrier-class routers.

Conclusion

Mobility and the Internet, the two most dynamic forces in communications today, converge in the design and implementation of the Ericsson mobile core network. Support for new services and a common transport technology are the main drivers for the integration of IP technology into the core network.

Ericsson is fully committed to the introduction of

- IP-based technologies in its products; and
- the mobile core network solutions that address mobility and the Internet.

Ericsson's flexible core network architecture allows operators to address mobility and the Internet independently.

The introduction of GPRS is the first step toward supporting IP-based applications. Support for new IP multimedia services is now being implemented according to the ongoing 3GPP standardization work outlined in this article.

New IP-based connectivity solutions are being introduced step by step with the objective of decreasing transport costs. The basis of these solutions are IP-based multiservice networks based on carrier-class routers. Other IP technologies, such as VPN, resilience, quality of service, and security mechanisms, are needed to provide the characteristics we associate with carrier networks.

The new IP-based multiservice network also opens up the way for much cheaper transmission techniques in the long-haul network.

REFERENCES

1 IP Multimedia Subsystem Stage 2, 3GPP technical specification 3G TS 23.228
2 Peck, M., Granström, P and Olson, S.: The future of communication using SIP, Ericsson Review Vol. 80(2002):1, pp. 30-37.
3 Ekeroth, L. and Hedström, P-M.:GPRS Support Nodes, Ericsson Review Vol. 77(2000):3, pp. 156-169
4 Reinius, J.: Cello—An ATM transport and control platform, Ericsson Review Vol. 76(1999):2, pp. 48-55
5 Session Description Protocol, RFC 2327
6 Session Initiation Protocol, RFC 2543

The future of communication using SIP

Peter Granström, Sean Olson and Mark Peck

In this article, the authors describe the session initiation protocol (SIP), which is the new standard for establishing multiparty, multimedia communication in IP-based networks, and some of the implications that its adoption carries for the future of communications. SIP-based solutions are shown to highlight Ericsson's competitive position in this new technology space.

SIP has been called the integrated services user part (ISUP) of next-generation networks. Certainly, the two protocols share many characteristics, but whereas ISUP embodies the evolution of common-channel signaling over the span of many years in the history of telephony, SIP has the potential to be a groundbreaking, disruptive force for change. We foresee that the adoption of SIP will usher in a new era of multimedia communications that leverage the strengths of Internet technologies to provide opportunities for innovation.

The authors discuss the basic concepts behind SIP, including some of the technologies on which it builds. This leads into a discussion of the features and capabilities in the protocol and its defined behavior, which enables operators to deliver rich functional content to their users.

The authors also consider the business drivers behind the adoption of SIP-based networks. These business drivers gave rise to the specification and the development of the IP Multimedia System (IMS), which is surveyed briefly in this article.

The authors conclude the article with an examination of a SIP server that can function as the basis for several SIP-based products in the near future. Although the product is initially targeted for the IMS, other products that use this server as a base can be positioned for a wide array of offerings and operating scenarios in the core networks of Ericsson's key operator customers. Ericsson is poised to capitalize on the revolution that SIP heralds, which offers an exciting promise for the communication networks of the future.

How SIP works

The session initiation protocol (SIP) is an application-level control protocol for establishing, modifying, and terminating multimedia sessions between one or more participants. It supports multimedia conferencing, Internet telephone calls, registration and redirection services, and is easily extended. It traces its roots to several multiparty conferencing initiatives in the history of the Internet Engineering Task Force (IETF) as well as to the World Wide Web (WWW) and Internet e-mail. As a result, SIP embodies a distinct protocol, with syntax and semantics for the messages to be exchanged, and a philosophy of end-to-end control over session establishment with support from servers in the network. Figure 1 shows the relationship of SIP to various other protocols in use in the Internet.

Basic session establishment

In the simplest case, two users want to communicate with one another using a variety of media types (such as audio, video, and text messages) over an IP network. The software application that enables the communication of each party is known as a user agent (UA). The UA could be running as a "soft client" on a PC, as the operating software in a mobile device, or in the firmware of a desktop SIP phone.

Party A, who wants to communicate with Party B, sends an INVITE message to Party B, who is listening on the official SIP port (5060). The INVITE body contains information encoded with the session description protocol (SDP), which indicates the types of media the receiving party is willing and able to use. Party B's response indicates his preferred media types. Once Party A returns an ACK message, each party is aware of the other's IP address and port numbers where the media streams will be received. Similarly, each knows which types and bandwidth of media the other is able to receive. When ACK has been sent and received, both ends begin transmitting data to the corresponding receiver ports, via a separate media connection using the real-time protocol (RTP) or some other appropriate transport protocol. Throughout the session, either party can make updates (indicating a new set of media types, addition of new parties to the session, or other changes), by sending additional SIP messages. At the end of the communication, either Party A or B can send a BYE message to indicate termination. When the other party responds the session is ended.

The user datagram protocol (UDP) is a required transport protocol primarily for performance reasons (TCP and other protocols are optional). But because of the unreliable nature of the datagram service in UDP, SIP contains its own retransmission mechanisms, including the three-way exchange between nodes for establishing sessions.

Syntax and addressing

Instead of being IP addresses, destinations in SIP can be represented with uniform resource indicators (URI), which have the same format as e-mail addresses. Accordingly, a valid SIP address might be *sip:Mark.Peck@ericsson.com*. This implies the use of the domain name service (DNS) to map host and domain names to IP addresses. Support for DNS is a key aspect of the integration of SIP with Web- and mail-enabled technologies, which are already familiar with the concepts of URIs and their interpretations.

The close connection between SIP and DNS facilitates interoperability with telephone systems and addressing mechanisms. Support for E.164 numbering in DNS (ENUM) allows SIP servers and clients to send and receive telephone numbers in place of SIP URIs in messages, and to route them in a sensible fashion.

SIP is a text-based protocol—it reuses the message structure found in the hypertext transfer protocol (HTTP) and simple mail transfer protocol (SMTP)—with numerous informational headers followed by a body (possibly multi-part). As a result, scripting languages such as Perl or Python are well suited for automating many session-processing tasks in a SIP server.

That SIP addressing lends itself to finding resources and people is an important part of the philosophy behind the protocol. When we consider real-time communications and the future of such services, we see that our focus is still mainly on communication between people. SIP is very useful for finding and "connecting to people," regardless of where they happen to be or what they are doing.

Indirection

Indirection is a key concept for understanding how SIP works. According to this concept, the current location of a user is hidden behind a permanent user identity or uniform resource locator (URL). A network-based SIP server binds the user's mobile identity to the permanent URL much in the same way a home agent functions in a mobile IP network. The two principal mechanisms in SIP that support this are redirection and proxying.

If Party A does not know the address of Party B, then Party A can send an INVITE to a redirect server, which returns a response indicating where Party B can be found (usually in the form of a SIP URI). Party A can then send a new INVITE to Party B.

The use of a proxy server allows users to have a node in the network that performs some intermediary function before the SIP messages are routed to their destination on behalf of the UA. If such a node exists, the SIP messages that it receives are forwarded to the appropriate destination and responses are forwarded in the reverse direction. Thus in terms of signaling, the proxy appears to each endpoint as if it were the other endpoint.

In cases that involve either a redirect or a proxy server, a location server might be consulted for information on the current SIP address of the indicated destination. The interface between the proxy or redirect server and the location server is not defined in the SIP RFC, but can be some appropriate querying interface, such as LDAP, HTTP, or DIAMETER. SIP supports real-time updates to the location server database via a

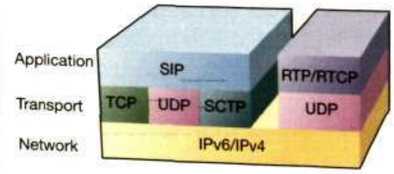


Figure 1
Relationship of various protocols used on the Internet.

BOX A, TERMS AND ABBREVIATIONS

3GPP	Third-generation Partnership Project
3GPP2	Third-generation Partnership Project 2
AAA	Authentication, authorization and accounting
ALG	Application layer gateway
API	Application program interface
CDMA	Code-division multiple access
CORBA	Common object request broker architecture
CPL	Call processing language
CSCF	Call/session control function
DIAMETER	IETF-defined protocol for AAA functions, successor to RADIUS (Remote authentication of dial-in user services)
DNS	Domain name service
ENUM	E.164 numbering in DNS
GPRS	General packet radio service
HSS	Home subscriber service
HTTP	Hypertext transfer protocol
IETF	Internet Engineering Task Force
IMPP	Instant messaging and presence protocol
IMS	IP multimedia subsystem
IP	Internet protocol
ISP	Internet service provider
ISUP	Integrated services user part
LDAP	Lightweight directory access protocol
MGCF	Media gateway control function
MRF	Media resource function
NAT	Network address translator
OSA	Open service architecture
PBX	Private branch exchange
QoS	Quality of service
RFC	Request for comment
RTCP	Real-time control protocol
RTP	Real-time protocol
SCM	Service control manager
SCTP	Stream control transmission protocol
SDP	Session description protocol
SGW	Signaling gateway
SIMPLE	SIP for instant messaging and presence leveraging
SIP	Session initiation protocol
SMTP	Simple mail transfer protocol
SNMP	Simple network management protocol
SRV	Server location records extension to DNS
SS7	Signaling system no. 7
TCP	Transmission control protocol
TSP	Telecom server platform
UA	User agent
UDP	User datagram protocol
UMTS	Universal mobile telecommunications system
URI	Uniform resource indicator
URL	Uniform resource locator
VPN	Virtual private network
WAP	Wireless application platform

REGISTER message that indicates the user's current location via a SIP URI. The SIP server that receives the REGISTER messages and updates the location database is called a SIP registrar. Through information obtained from the registrar, the redirect or proxy server is able to reroute a SIP request to the destination where the user wants to be reached.

Indirection is not limited to the user—by using DNS it can also be applied to the SIP servers themselves. A number of DNS mechanisms exist to map a symbolic name for a SIP server into an IP address where that server can be reached. Particularly interesting are SRV records (server location records extension to DNS), which allow the definition of one or more SIP servers that are the first point of contact for a given domain. For example, four separate SIP servers might share the load for the *ericsson.com* domain. The use of DNS makes it very easy to establish these network topologies. Therefore, the process by which a user is contacted using SIP involves

- identifying the server for that user; and
- determining the location of the server.

Indirection can be used in each of these steps to yield a very flexible and fluid communications network.

Forking

By means of a procedure known as forking, SIP proxies can simultaneously forward a SIP message to multiple destinations. A user can thus have multiple destinations registered (say, a mobile device as well as a desktop phone) and have each destination alerted simultaneously when a new session request arrives. The proxy server correlates the responses received from various branches and ensures that only a single upstream response is sent to the client.

Using SIP in various environments

To bring useful SIP implementations to market, some hurdles must be overcome in particular environments. Knowing these hurdles and their solutions helps explain Ericsson's architectural choices for the solutions currently under development.

SIP on the public Internet

SIP on the public Internet is the default case, given the high bandwidth and low latency between network elements (making QoS mechanisms less important) and the lack of differentiated domains that would introduce elements such as firewalls, network address translators (NAT) or other gateways. In this environment, emphasis can be put on terminal-based or end-to-end solutions that embody the "pure" philosophy of the Internet. Most of the general descriptions of SIP in this article are based on this scenario, and it is from here that we will add domain-specific extensions that accommodate the unique needs of the other scenarios.

SIP in protected domains

For many users, the default scenario described above is incomplete. Most notably, the network is often not "transparent" end-to-end at or above the IP layer. Provisions must thus be made to ensure unchanged behavior from an end-user perspective.

For security purposes, many IP networks are protected from external traffic by the installation of a firewall, which solely allows packets to flow through designated secure points. For security purposes and address conservation, many ISPs and enterprises al-

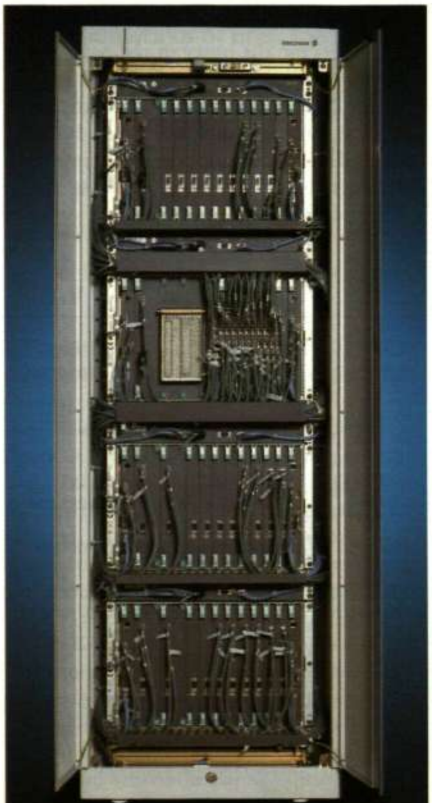


Figure 2
Telecom server platform (TSP).

locate private IP addresses to their users. These private addresses are then mapped to public IP addresses by a network address translator. For "pure" client-server applications, such as the Web, file transfer, or e-mail, in which the precise real-time network address of the end-user is irrelevant, application-level proxies can provide unimpeded access to the services needed, and firewalls and NATs need not be show-stoppers. However, although SIP differentiates the signaling from the media in establishing a real-time communication session, the two are highly interdependent, and firewalls and NATs can pose a problem. A SIP proxy that is employed to pass messages through a firewall, for example, must provide indirection support for locating users on either side of the firewall, and have some means of instructing the firewall to allow media to pass through it (since the media might use an entirely different combination of IP addresses and port numbers). Similarly, if a NAT transforms the private IP address used by a SIP UA into a public address, then the private media address described by the session description protocol (SDP) in the SIP message will be unusable by the remote UA. Therefore, an application layer gateway (ALG) must be coupled to the NAT to ensure that the SDP is rewritten to reflect the public address that will be used to transit the media across the domain boundary.

SIP in a mobile environment

Limited bandwidth over the air interface means that the amount of signaling must be kept to a minimum. The size of SIP messages might become quite large due to the number of steps taken in routing between endpoints and because SIP is a text protocol rather than binary. This necessitates the establishment of a node that can customize signaling for efficiency without affecting end-to-end transparency. Such a node (SIP proxy) can also serve as a useful default inbound and outbound proxy for address-translation services, the invocation of locally significant services (such as taxi locators, weather, or news), and guaranteeing quality of service (QoS) for network resources.

Consequences

Because of the particular requirements imposed on SIP-based networks in these environments, there is great benefit to be derived from an architecture that supports some or all of the requirements in a flexible manner. Several specialized nodes have been identified by (among others) the Third-generation Partnership Project (3GPP) to accommodate these requirements and provide operators with a powerful framework for adding value to the end-user experience.

Capabilities supported by SIP

In the paragraphs that follow we list some examples of services and service building blocks that are made possible by the basic mechanisms of SIP. We will also describe scenarios in which they add value to the end-user experience.

Add/drop media

SIP supports the ability to add new media types (or remove unwanted media) in the middle of a session. Additionally, disparate media types can be supported at the various endpoints without degrading the user experience (that is, lowest-common-denominator communications are not necessary). Consider a surgeon who, while in the middle of a procedure, wants to consult with a colleague in another part of the country. They begin discussing the case over a traditional voice connection, but midway through the session, the specialist determines that she needs to see the patient's condition. A one-way video feed is established from the operating room, and the specialist begins to type in a series of instructions that show up in a separate chat window on the surgeon's viewing screen. The basic SIP mechanisms that support this exchange require no extensions or specialized hardware apart from an image-capture device in the operating room.

Find me/follow me

The specialist from our previous example can be simultaneously registered (with a SIP registrar) in multiple locations according to her daily patterns. Perhaps she has a fixed SIP-enabled terminal on her desk that rings when an INVITE message is sent to her public SIP address. She can also register

- with her mobile terminal, which can receive audio calls or instant messages; and
- with her secretary, who might be instructed to take any audio calls that have not been answered after a given number of signals.

Finally, she can register with a SIP video client on her desktop PC to receive video feeds from surgeons while she is talking to them on the phone. Audio calls might ring simultaneously at all four locations. When she answers one of the devices the other three devices stop ringing. Note that none of what we have described thus far requires specialized PBX equipment!

Presence and instant messaging

In the example above we mentioned the possibility of sending and receiving instant messages via SIP, and its use for instant messaging and presence has also been specified by the SIMPLE working group of the IETF. SIP is currently the leading candidate to fulfill the requirements of the instant messaging and presence protocol (IMPP) working group, which has specified an overall instant messaging and presence framework. Several major operators and vendors, including AOL and Microsoft, have announced plans to support SIP for interworking between presence and instant-messaging domains. Because it is well suited for this purpose, SIP will be one of the leading choices for transport of presence and instant messaging information between users.

Conferencing and distance working

The operating-room scenario above could be extended to include a class of medical students who observe the surgery in progress on a remote video feed and ask questions of the specialist over an audio connection. A lecturer located on a college campus could provide distance-learning support to students spread around the city. The necessary equipment includes little more than video capture equipment, a conferencing server to mix the audio, and a chat-room environment in which students' questions can be queued and answered.

The abolition of Class-5 services

Traditional voice services that previously provided substantial revenue to incumbent local operators are made trivial by SIP. Services such as caller ID, call waiting, and call hold are handled by basic SIP mechanisms within the UA and require no input or control by the network. While this poses a threat to the existing business model, in the new business model of SIP-enabled networks, it becomes a competitive advantage since users are given greater control over the behavior of their communications services at minimal cost to the operator. For this reason, the simplification and trivialization of formerly significant services gives operators with new networks a competitive advantage over their legacy competition.

Multiparty gaming

SIP can also be used to transport a wide range of real-time information, including gaming events for multi-player games. Additionally, SIP (whether supported by the game client itself or via a separate communicator application) can be used during the game for audio, video, and chat between players.

VPNs made simple

The use of DNS, ENUM, and the indirection capabilities of SIP make it easy to develop and manage virtual private networks (VPN). A single SIP proxy can provide address mapping and forwarding services for a remote location, giving users the perception that they are located within the same corporate domain as their colleagues (even when they are not).

Business motivation for SIP-based systems

We are currently witnessing a dramatic change in the way we communicate. The monolithic incumbent telecommunications operators, who until only a few years ago provided all or most of our communications needs, are now under constant pressure from

- technology turning one of their core assets—network bandwidth—into a commodity;
- regulators and governments encouraging competition in all areas of operator businesses;
- dramatic price drops being driven by regulatory changes, competition, and technology;
- the Internet, and IP in general, which is moving communications into a totally new era of packet-switched technologies;
- the need to make large investments to meet users' demands for broadband access;
- the threat that mobile telephony will overtake and replace fixed telephony, which challenges operators' core service and revenue sources; and
- the relative ease with which a user can obtain services from third parties in an IP-based network, relegating the role of the operator to that of "bit-pipe" provider.

Within the next decade operators will need a new generation of networks, primarily based on IP technology. These will pave the way for new opportunities to earn revenue and make customer-driven services the key to profitability. All the same, these new networks do not have the same value chain as traditional telephony, and the business models have not yet been established. Operators will thus need to learn how they can take advantage of these new networks and move their businesses into a totally new space. Simply providing the same set of services using new technology will not enhance the end-user value. In fact, as we have seen, a re-implementation of traditional services will only drive margins down and eventually lead to operator losses. The deployment of SIP-based networks can be a remedy. Compared to existing systems, elements of the SIP technology and business value proposition include but are not limited to

- presence;
- combinational services;
- access independence;
- new charging models;
- quality of service; and
- security.

Figure 3
Prototype SIP telephone.

Presence

By presence we mean that a group of individuals can share information (status) on their current availability. SIP provides new and creative ways of developing services based on presence information. The value of presence information is not found in providing it as an individual service, but rather in combining presence with other multimedia capabilities, such as combinational service offerings.

Combinational services

With SIP, it will be easy to combine conversational multimedia services with other categories of services, such as directory information, Web browsing, positioning, and presence. For example, a location-based service can be developed to combine a conversational communication session with positioning information and maps to provide information that is pertinent to the geographical location of the parties involved.

Access independence

Being an application-layer protocol in the IP-based suite, SIP is access-independent and offers seamless service capabilities between fixed and mobile networks. This is a key element in making the promise of fixed-to-mobile convergence a reality.

New charging models

Operators will be able to define new charging models based on actual media usage. For example, if the communication between two parties begins with a real-time voice session and video is later added, it is possible to charge for the sessions individually and for actual media usage during the sessions.

Quality of service

Given the importance of conversational multimedia services (where sessions which are real-time sensitive, and which involve voice and video streams, are shared with streams that are not as time-sensitive) then sufficient QoS mechanisms must be in place to guarantee a rich end-user experience. New product offerings will provide a robust and flexible architecture that supports the required QoS and security requirements in mission-critical applications. SIP-based products sit on top of the IP network and take advantage of the capabilities of the underlying network to provide QoS.

Security

Security mechanisms are major concerns for operators who deploy IP-based networks. SIP can encrypt and authenticate signaling messages; RTP supports the encryption of media. Together these two protocols provide cryptographically secure communications. Important benefits will be found in providing the necessary security functionality for the operator, including authentication, access control, confidentiality and integrity. These capabilities enable operators to deploy secure next-generation networks built on IP technology.

Ericsson's IMS delivers on the promise of SIP

Recognizing the opportunities and challenges of SIP-based networks, and seeing the

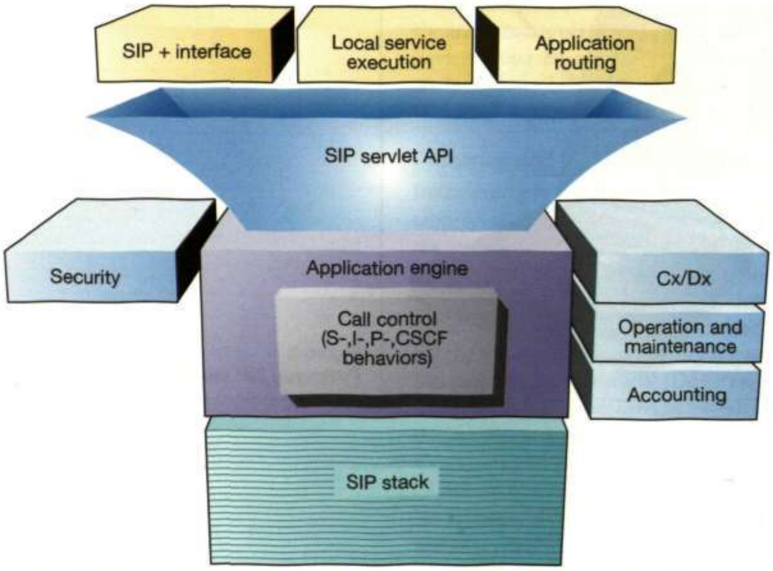


Figure 4
The CSCF architecture.

potential for dramatic change in mobile communications, the 3GPP has defined a new domain that takes advantage of the packet-switched capabilities provided by GPRS and EDGE. This domain is called IP Multimedia. Although the leading mobile standards body drives the standardization work, IP Multimedia is independent of any access technology and is also intended to be suitable for fixed-access networks.

Components of the IMS

Ericsson is building a 3GPP-compliant IP Multimedia System (IMS) that will bring the benefits of SIP-based communications to operators and end-users in the near future. The commercial IMS is not a standalone offering, but is intended for delivery as part of solutions from several of Ericsson's major business units.

The remarkable success of the ENGINE concept has given Ericsson a powerful market message for fixed operators around the world. The ENGINE concept continues to evolve from an ATM product focused on telephony into an IP-based solution: ENGINE Integral + IP. The next major step, called ENGINE Multimedia, offers a multimedia-enabled solution. At the heart of this solution sits the IMS.

The standards work for CDMA2000-based mobile systems is carried out in 3GPP2, which has defined a network architecture for multimedia that is very similar to that of 3GPP. The products and applications deployed in the IMS will also be part of the offerings from Ericsson to CDMA-based operators in the US and elsewhere.

Delivering on the promise of interoperable multimedia

During 2002, Ericsson will release IM 0.9, which will allow operators and end-users to see

- close alignment with the 3GPP architecture, including the CSCF, MRF, MGCF, SG, HSS, subnetwork manager, and other nodes;
- integrated operations and management;
- improved system characteristics, including capacity, reliability and throughput; and
- new services.

Beyond release IM 0.9 lies the reality of commercial systems delivered as part of total solutions that offer robust security, flexible accounting and charging, differentiated QoS, and support for the wide array of media types and devices that will be connected to the system.

SIP Server example: CSCF

The call/session control function (CSCF) is to be deployed as part of Ericsson's IMS, and as such it can be used in next-generation wireline and wireless networks using SIP as the signaling protocol. However, it is not intended to be a standalone commercial offering. Instead, it has been designed to be suitable for a wide range of applications, including the call/session control function defined by 3GPP, and the session control manager (SCM) defined by 3GPP2, as well as a number of call control and application servers described in various Internet drafts.

The CSCF offers an ideal platform for deploying next-generation communication services in multimedia-enabled networks. It provides support for popular Internet protocols and open, industry-standard APIs on a scalable, high-performance platform that draws from Ericsson's experience of telephony systems (Figure 4).

High-level functionality

The CSCF, which supports the establishment, modification and release of IP multimedia sessions using the SIP/SDP protocol suite, provides the following capabilities:

TRADEMARKS

CDMA2000 is a trademark of the Telecommunications Industry Association (TIA).

- subscriber registration;
- invocation of multimedia sessions (originating and terminating)—the CSCF supports SIP mechanisms for invoking one or more IP multimedia sessions;
- capability negotiation at session invocation—the CSCF supports SIP mechanisms for establishing the capabilities of a session;
- modification and clearing of multimedia sessions;
- forwarding, redirection, and rejection of multimedia sessions;
- notification of multimedia session events to the service network; and
- interfaces to an IP policy control function.

Carrier-class technology

The CSCF has been designed to take advantage of the capabilities of TSP. This is considered a critical part of the value proposition of this product to Ericsson's customers, because of their need for high availability, cost-effective scalability, and best-in-class capacity for running large-scale networks. Notwithstanding, a focus on carrier-grade characteristics does not limit the functional capability of the server.

Routing and addressing

The CSCF is capable of routing according to standard SIP mechanisms for session establishment, clearing, and modification. It can query DNS servers to map E.164 numbers (using ENUM) or SIP URIs to network addresses. The CSCF also supports specialized routing mechanisms (as defined in 3GPP, 3GPP2, or other standards bodies) that involve other nodes in the network (such as the HSS or AAA server). It is also able to route multimedia session attempts to and from non-3GPP and non-IMS systems. This gives operators the ability to provide their customers with open and interoperable services.

Customized services

The CSCF supports the invocation of services either remotely (remote invocation) or locally (local invocation). For remote invocation the CSCF supports an external API (such as Parlay/OSA) carried over CORBA for access to remote application servers. Applications that want to make use of the capabilities of the CSCF can also use this interface. For local invocation the CSCF supports the execution of services on the node by means of at least two mechanisms:

- through the execution of scripts written in the call processing language (CPL); and
- through the execution of Java SIP servlets. The SIP servlet engine allows programmers to write applications in Java that control the behavior of the CSCF in response to SIP messages. This gives operators extensive opportunities to customize and deploy basic services that enhance the establishment of multimedia sessions.

Charging

The CSCF supports the collection of accounting information for time- and event-based charging. Using DIAMETER it forwards this information to the accounting server, where it can be aggregated with data collected from other nodes.

Availability/reliability

The CSCF has been designed for uninterrupted operation—it takes advantage of the facilities offered by TSP to enable high availability of hardware and software. All hardware and software upgrades can be performed on the node while in operation. Availability is expected to exceed 99.995%. A single hardware failure will not stop the operation of the CSCF. Moreover, the network architecture of the IMS provides additional redundancy in the event of catastrophic failure of a single node—alternate CSCFs can be selected without loss of service.

Conclusion

SIP is a central part of the value proposition of future multimedia networks. This value stems from several key aspects of the protocol, including

- the flexibility of addressing, routing and modifying messages using the protocol;
- support for a wide range of media types, simultaneously invoked or selectively added as the need arises;
- the wealth of information that can be communicated to all nodes in the network, fostering an end-to-end view of services and applications; and
- ready integration with Web-enabling technologies that springs from the origins of SIP in the IETF.

Ericsson is bringing several new products and solutions to bear in this new technology space. These offerings will enable operators to position themselves in new markets and to realize new revenue streams while protecting their investment in legacy systems.

REFERENCES

1 Handley, Mark et al. The Session Initiation Protocol, draft-ietf-sip-rfc2543bis-05.txt, http://www.ietf.org, October 2001
2 Camarillo, Gonzalo. SIP Demystified, McGraw-Hill, 2002
3 Rosenberg, J., et al. SIP extensions for presence, Draft-ietf-simple-presence-04.txt, http://www.ietf.org, November 2001
4 Rosenberg, J., et al. SIP extensions for instant messaging, Draft-ietf-simple-im-01.txt, http://www.ietf.org, July, 2001

Ericsson Mobile Operator WLAN solution

Tomas Boström, Tomas Goldbeck-Löwe and Ralf Keller

Wireless LAN is a complementary service offering for mobile operators. The Ericsson Mobile Operator WLAN solution combines the wide-area benefits of second- and third-generation mobile systems, including unlimited roaming and mobility, with additional throughput and capacity in indoor hot-spots via wireless LANs. The solution gives broadband mobile public access to the Internet and to corporate intranets from relatively little additional investment.

The authors present WLAN as a complement to GPRS and UMTS and explain the market drivers of the solution. They describe the Ericsson Mobile Operator WLAN – Release 1 in detail and give an overview of ongoing standardization.

WLAN as a complement to GPRS and UMTS

Wireless LAN (WLAN) technology is being used more and more in homes, offices and indoor public areas. Mobile service providers are exploring opportunities to extend their service portfolios by providing limited, indoor WLAN hot-spot access. The same basic configuration—that is, a laptop computer with a WLAN adapter—can be used to gain access in indoor public and private environments. End-users can thus access their office environments without any noticeable change in network performance.

Market drivers

The WLAN market is currently undergoing very rapid expansion. At the office it is being used to increase organizational flexibility, and in homes it is replacing obtrusive cables. While these market segments are conceptually relatively straightforward, there is a third market segment that is still in its infancy: indoor public WLAN access—that is, the ability to access the Internet from indoor public places, such as airports lounges, hotels and conference centers. This segment has attained considerable interest over the past few years and several entrepreneurs have set up public WLAN networks in major US and European cities. So far, however, the number of subscribers is still low, due to

- poor coverage—users might be able to use WLAN services at the airport of departure, but not at the airport of arrival, or at the hotel;
- a lack of brand recognition—the service operators are often new start-ups, which causes end-users to hesitate to use the service; and
- a lack of roaming agreements—end-users are forced to locate different service providers at the places they roam to.

As has been demonstrated time and again during the history of wireless service, coverage is the single most important factor for making a new wireless technology a success. This is where cellular mobile operators come into the picture. They already have an infrastructure that covers wide areas. Therefore, with little extra investment, they can add indoor WLAN access to their present offerings. In addition, they have management systems for billing, authentication and subscriber handling. They also have a very large base of mobile subscribers who would be prime targets for a high-speed data offering. Given this background, Ericsson believes that mobile operators are in a very good position to add indoor WLAN service as a complement to their existing wide area service. Business professionals are expected to be the first important group of users of a combined cellular/mobile WLAN service. They already use mobile phones and they usually bring their laptop computers along when they travel.

Ericsson Mobile Operator WLAN solution

WLAN is a complementary service offering for mobile operators. GPRS and UMTS operators can provide two kinds of access to mobile packet-data services: via the wide-area GPRS/UMTS network, and through

BOX A, ABBREVIATIONS

3GPP	Third-generation Partnership Project
AP	Access point
APIS	Application program interface server
ASN	Access serving node
BGW	Billing gateway
CABS	Customer administration and billing server
CDR	Charging detail record
ETSI	European Telecommunications Standard Institute
GPRS	General packet radio service
GSM	Global system for mobile communication
HiperLAN	High-performance LAN
HiSWAN	High-speed wireless access network
HLR	Home location register
IAPP	Inter-access point protocol
IETF	Internet Engineering Task Force
IP	Internet protocol
ISDN	Integrated services digital network
ISP	Internet service provider
LAN	Local area network
MAC	Medium access control
MMAC	Multimedia mobile access communication
MSISDN	Mobile station ISDN
OTP	One-time password
PDA	Personal digital assistant
RADIUS	Remote authentication dial-in user service
SAS	Statistics and accounting server
SCS	Service control server
SIM	Subscriber identity module
SMS	Short message service
SMS-C	SMS center
SOHO	Small office/home office
SSL	Secure socket layer
UMTS	Universal mobile telecommunications system
WAN	Wide area network
WLAN	Wireless LAN

the high-capacity WLAN access network. WLAN access is a particularly suitable form of alternative access at indoor public hot-spots, such as airport lounges, hotels, and conference areas. What is more, the packet-data service can be provided seamlessly (with some trade-off in performance and possibly service) between GPRS/UMTS and WLAN. Ericsson's Mobile Operator WLAN solution, for example, fully integrates WLAN into the mobile operator service offering. It ensures easy subscriber handling and integrates WLAN network management into existing network management routines.

Requirements and basic principles

Requirements for the Mobile Operator WLAN solution

The most important requirements pertaining to the Mobile Operator WLAN solution[1], users, operators, and system are as follows:

- WLAN interworking with GPRS and UMTS must adhere to standards;
- it must be possible to reuse GPRS/UMTS authentication mechanisms for WLAN access without degrading the security of the GPRS/UMTS network or its subscribers;
- an enhancement to the GPRS/UMTS subscriber base must be specified so that existing subscribers can easily obtain WLAN services;
- roaming must be specified between wide-area cellular radio access and WLAN radio access networks. Moreover, roaming between different Mobile Operator WLANs must be supported; and
- user data in the WLAN must be protected to the same extent that it is in GPRS/UMTS.

Basic principles of interworking with WLAN

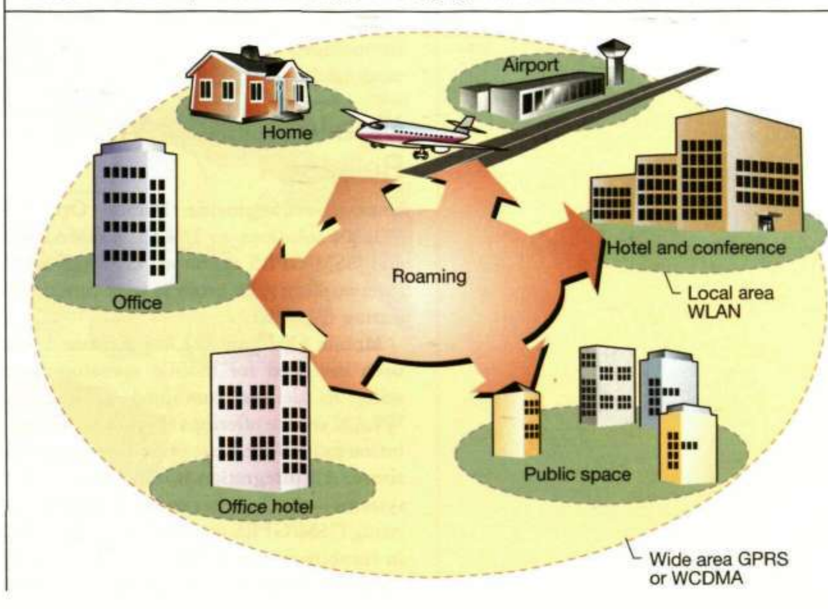
As stated above, the Mobile Operator WLAN solution combines the wide-area benefits of second- and third-generation mobile systems, including unlimited roaming and mobility, with additional throughput and capacity in indoor hot-spots via WLANs. Users of public WLAN will be part of the mobile operator's subscriber base. Thus, with little additional investment, mobile operators can further expand a second- or third-generation packet service.

BOX B, TERMINOLOGY

Access point
Device responsible for the centralized control of the resources in a radio cell.
Authorization
The act of determining if a particular right, such as access to some resource, can be granted to the presenter of a particular credential.
BRAN project
ETSI project preparing standards for equipment providing broadband (25 Mbit/s or more) wireless access to wire-based networks in private and public environments, operating in either licensed or license-exempt spectrum. These systems address both business and residential applications.
HiperLAN/2
High-performance radio LAN type 2 is a short-range wireless LAN that provides local broadband access. HiperLAN/2 is being standardized by ETSI (see BRAN project).
Home network
A network where a user is known to the authentication system of that network.
Mobile terminal
End-system equipment that provides the interface to human beings through a set of applications. In the context of this document, the MT includes the functions and protocols necessary to provide and handle communication with the WLAN network and other networks (GPRS, UMTS), services and applications.
Private network
A network under the administrative control of a single entity, such as a corporation or family. Network services are basically only offered to the users employed by that entity.
Public network
A network under the administrative control of a legal entity, often referred to as the network operator. The network services offered to the customer base are defined by the operator (and possible partners).
Roaming
The ability of a user to function in a serving network other than the home network.

However, the standardization of WLAN interworking needs to be limited to a very small set of items, such as the reuse of existing subscriber management mecha-

Figure 1
Ericsson Mobile Operator WLAN vision—bridging local-area and wide-area networks.



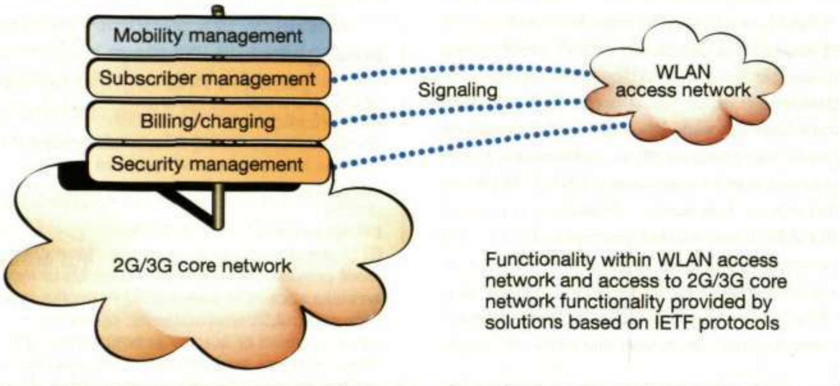


Figure 2
Reuse of core network functionality.

nisms, authentication and security functions, and billing functions (Figure 2).

In this way, the WLAN becomes a complementary IP access network of the current GPRS/UMTS packet-switched domain. If the cellular mobility management is not reused, then only signaling messages need to be exchanged between the WLAN radio access network and the GPRS/UMTS core network. This minimizes the impact of WLAN on GPRS/UMTS and reduces the need for standardization work within the 3GPP. In addition, the architectural solution, with its IETF-defined interfaces to GPRS/UMTS networks, has the advantage of being generically suitable for all WLAN technologies. This is in line with ongoing work in the ETSI BRAN project.

Mobile Operator WLAN – Release 1

Ericsson will provide Mobile Operator WLAN solutions to UMTS, CDMA2000 and GSM/GPRS operators. GPRS-WLAN interworking pilot projects were conducted during Q4/2001.

Mobile Operator WLAN Release 1 has been designed for mobile operators who want to deliver combined GPRS and WLAN service offerings (Figure 3). The solution follows the basic principles described above. All integration is done in back-end systems—that is, it has no impact on the existing GSM/GPRS network. This approach, in combination with the reuse of the GSM transport network, billing gateway, and GPRS RADIUS server, reduces operating and capital expenses. The system architecture assumes that a GPRS infrastructure has been deployed and is in service. The supported WLAN standard is IEEE 802.11b, which is already widely deployed in the market.[2] However, the limited number of frequencies for WLAN and other systems at 2.4 GHz complicates the issue for systems that reuse frequencies. Moreover, massive use of frequencies at 2.4 GHz can severely degrade service. Consequently, plans have been made to expand the Mobile Operator WLAN offering into the 5 GHz range where more spectrum has been reserved for WLAN systems and where the spectrum requirements in Europe enable contention-free access to spectrum.

Functionality and features

The system gives mobile operators a set of tools that they can use to exploit the WLAN business opportunity. These tools include different authentication and billing models, service differentiation, roaming, corporate network access, localized content, and support for captive portals.

Billing models

The system gives operators a choice of four billing models: flat-fee, consumption-based, voucher-based (similar to pre-paid subscription), and free service.

- Flat-fee billing—the user has a subscription with the operator and is charged a fixed amount each month, regardless of usage.
- Consumption-based billing—the user

has a subscription with the operator and is charged according to usage. Charges can be based on time or volume (minutes logged in or megabytes of data transferred).
- Vouchers—the system supports the generation and recognition of vouchers, which enable time-limited accounts (per day, week, and so on) with randomly generated login and password information. This information can be printed on scratch cards, which can be distributed and sold, for example, at the WLAN hotspot. The voucher concept can be employed to advantage when the WLAN service is first being introduced, to heighten user awareness of the capabilities of the service.
- Free services—the system supports services for non-paying users. The scope of the services restricts access to specific sites. The user is granted access to local information, such as gate info at an airport, special offers from local shops, or redirection to the operator portal with information on subscription types and pricing.

Authentication models and WLAN service access

The system uses Web login and supports two authentication models:

- SIM-based authentication by a one-time password (OTP) delivered via SMS; and
- static password-based authentication.

The SIM-based authentication model uses a secure and authenticated channel to distribute one-time passwords for WLAN service access.[3] The concept, which uses the GSM subscription (SIM card + GSM phone) as an authentication token, uses SMS to distribute OTPs via GSM to SIM-authenticated users. The WLAN service-access scenario is as follows: The user starts a browser that is redirected to a WLAN service login page. The login page prompts the user to provide his credentials. After the user presents his identity (GSM phone number), the Web page is updated and prompts the user to provide the OTP. In the meantime, the system delivers an OTP to the user's phone via SMS. If the OTP has been entered correctly on the Web page, the system grants the user access.

With static password-based authentication, the user launches the Web browser application that is automatically redirected to a WLAN service login page. The login page prompts the user for username and password. The user enters the username and password that was supplied with the subscription confirmation or voucher. SSL in server-authentication mode is used during the transfer to protect sensitive data con-

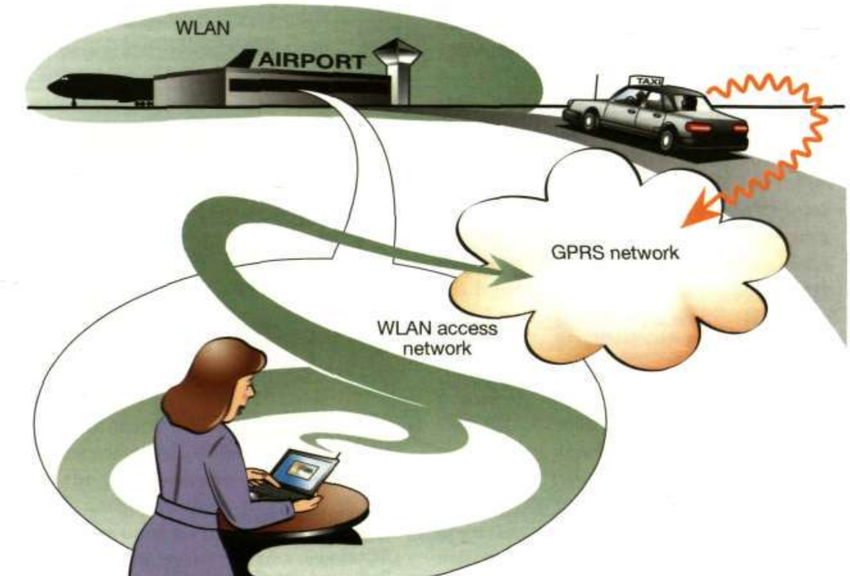


Figure 3
Complementing GPRS with WLAN service.

taining user credentials. If the login is successful, a session control window is displayed. The user terminates the session by clicking on a logout button in the session control window. A logout window then appears displaying a summary (statistics) of the session.

Localized content and captive portals

The login page for the WLAN service can adapt according to user location. The user is always presented with localized information, such as departure information at the airport, the menu of the day at the hotel, and local advertisements. The system can also be configured to present a specific home page for authorized users. The home page functions as a captive portal that can be tied to the user's subscription type and current location.

Service differentiation—product profiles

The system can offer different services to different users—for instance, Gold, Silver or Bronze subscriptions. Differentiation is made according to application type and geographical availability. Occasional users (bronze) get limited Web access, whereas no limitation of traffic type is put on regular users (Silver). Similarly, users with a special subscription (Gold) might receive additional services. Another example could be vouchers sold through a hotel chain that solely grant access on the hotel premises.

Roaming

The system allows roaming with other service providers who offer WLAN services that support RADIUS.[4] Roaming can be achieved with coherent service differentiation, provided that the roaming partners define the product profiles in the same way.

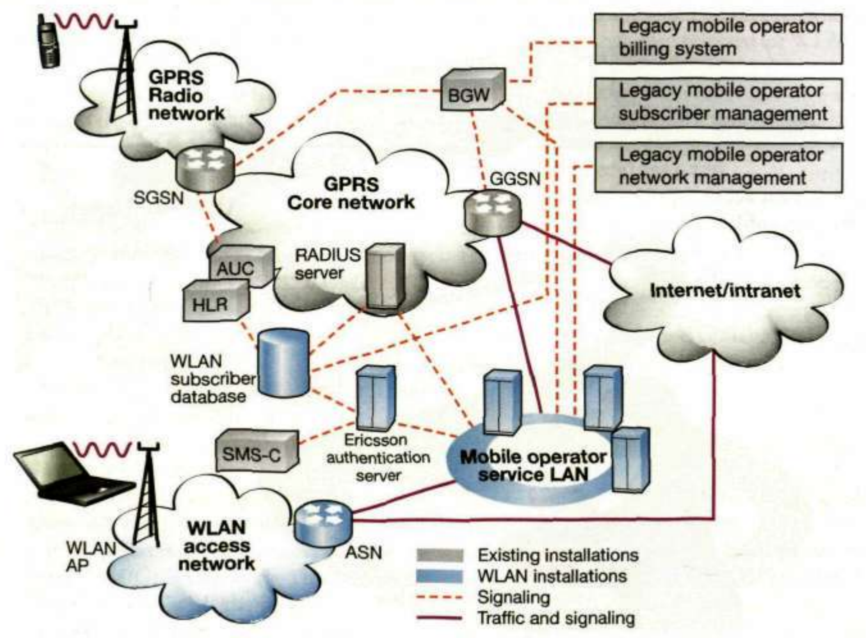
Corporate access

Corporate access is a vital end-user service. Indeed, business professionals will be the first to adopt WLAN service. Corporate access gives users secure access to e-mail, the Web and company file structures (documents and presentations). Ericsson provides verified corporate-access solutions that operate over GPRS and WLAN.

System overview

The Ericsson Mobile Operator WLAN Release 1 assumes that a GSM or GPRS infrastructure has been deployed and is in service. To deploy the WLAN service, an additional set of components must be added to the GSM/GPRS infrastructure. These components are installed at

Figure 4
Ericsson's Mobile Operator WLAN solution (Release 1).



- indoor sites where the WLAN service is to be offered (hot-spots); and
- a central location where the access management system is located.

Figure 4 illustrates the system architecture of the Ericsson Mobile Operator WLAN solution (Release 1).

A set of IEEE 802.11b access points, installed to give indoor hot-spot coverage, is connected through an Ethernet network to the access serving node (ASN). The ASN, in turn, is connected to the IP backbone. The prime functions of the ASN are to restrict services to authorized end-users and to enforce the corresponding service policies (product profiles). The ASN is a generalized access server that authenticates users who want to access services from a hot-spot site. It controls access by filtering all packets coming from, and directed to, the WLAN access network. The ASN also creates accounting data. The access management system consists of the following components:

- service control server (SCS);
- authentication server;
- statistics and accounting server (SAS);
- customer administration and billing server (CABS);
- application program interface server (APIS); and
- WLAN manager.

The SCS assists the ASNs, performing authentication and authorization. For redundancy purposes one ASN is associated with two SCSs. Each time a user logs in, the ASN contacts the SCS using RADIUS to authenticate the user. The SCS either directly answers the authentication request (voucher account) or functions as a RADIUS proxy to the authentication server or the external RADIUS server (post-paid subscriber). The SCS also provides the ASNs with access policies, in the form of product profiles, using LDAP.[5] The product profile is enforced in the ASN after the user has been authorized.

The authentication server enables SIM-based authentication. When it receives an authentication request from the ASN, it looks up the user in the subscriber database, creates a one-time password, and communicates with the associated SMS center (SMS-C) in order to deliver an SMS with the one-time password. When the user has provided the password, the ASN performs a second authentication request. If the entered password matches the OTP delivered via SMS the system grants access. The authentication server functions as a back-end RADIUS server in the GPRS home network.

The SAS provides accounting and statistical functions. It aggregates accounting data from all ASNs and compiles data for statistics. The aggregated billing data is sent to the GSM/GPRS billing gateway (BGW) as WLAN-specific CDRs and retrieved by the legacy billing system.

The CABS is used as the administrative interface for vouchers and product profiles. It includes a Web-based interface and a CORBA interface to the APIS.

The core function of the APIS is to store original data (product profiles, voucher parameters such as identities, passwords and limits). It constitutes the storage from which the SCSs replicate data. However, configuration data on regular post-paid subscribers is administered and stored in the authentication server or the RADIUS database of the GSM/GPRS network. The authentication server or GSM/GPRS RADIUS server contains a duplicated set of information from the HLR (that is, the MSISDN). All WLAN post-paid users must have additional tags or fields which relate to the WLAN service entered and which specify the product profile that should be enforced. The APIS comprises a CORBA interface that connects it with the CABS and which can also be used for integration with legacy back-end subscriber management and provisioning systems.

The WLAN Manager is the integration point for all WLAN-related management in the Ericsson Mobile Operator WLAN solution Release 1, including network management, element management, and systems management.

Standardization

Ericsson has strong presence in standardization bodies and is one of the initiators of the 3GPP work item on WLAN-UMTS interworking. Besides the 3GPP, other standardization bodies actively defining essential functionality for WLAN-UMTS interworking are the IETF, IEEE, ETSI and MMAC. Ericsson has the goal of ensuring that any WLAN terminal has the technical possibility of securely connecting to any GPRS/UMTS-WLAN network (subject to commercial agreements), and that all required functionality is in place to allow global roaming, charging and billing, and operation and maintenance. The WLAN systems will complement the current GPRS/UMTS packet-switched domain.

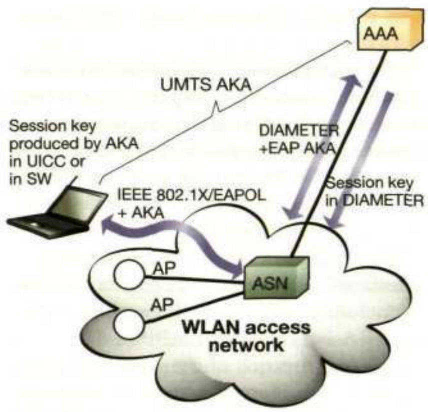


Figure 5
Example of UMTS AKA authentication between a mobile terminal and WLAN service.

Work item in the 3GPP

The first 3GPP activity within the WLAN area is to conduct a feasibility study

- to define service scenarios;
- to define the interworking requirements put on UMTS;
- to specify the interworking functionality; and
- to identify the need for enhancements in the UMTS specifications.

Following the feasibility study, work will begin on drafting the actual standards. The first set of standards is expected in mid-2003 as part of 3GPP Release 6.

To minimize the workload in 3GPP, a five-phase approach has been defined. The first phase covers common billing between WLAN and UMTS. Later phases will cover interworking (where subscribers can roam between UMTS and WLAN), security aspects, intersystem session continuity, and service mobility. The last step might imply the complete integration of WLAN access into UMTS, where the WLAN access points are connected directly to an evolved RNC.

Service aspects will need to be assessed in terms of service requirements and the support of UMTS services over the WLAN radio access. Man-machine interface aspects should define a minimum set of functions needed to support the choice of access system when both access systems are available.

Security requirements should be specified in such a way that

- the security level of the UMTS platform is not compromised; and
- the security level offered users in the WLAN mode is comparable to that of UMTS.

WLAN standardization in Europe, the US and Japan

Several WLAN standards have been drafted in Europe, the US, and Japan. In Europe, the BRAN project (ETSI) has specified HiperLAN/2 in the 5 GHz frequency band. HiperLAN/1 was a predecessor specification that was never commercially launched.[6]

In the US, the IEEE has specified the 802.11 family of standards with a single medium access control (MAC) protocol and several physical layers[7-8]:

- one infrared physical layer (802.11);
- three 2.4 GHz physical layers (two of which are in 802.11; and one supporting 11 Mbit/s in 802.11b[7]); and
- one 5 GHz physical layer (802.11a).

Further specification work is ongoing within

- 802.11e—to add QoS support to the MAC protocol;

- 802.11f—to specify an inter-access point protocol (IAPP) to transfer information between access points at handover;
- 802.11g—even higher throughput is envisioned on the physical layer for 2.4 GHz;
- 802.11h—the 5 GHz physical layer is being enhanced to support radar detection, dynamic frequency selection, and to fulfill spectrum requirements for power control in Europe; and
- 802.11i—improving the security in the standard.

In Japan, the multimedia mobile access communication (MMAC) specification consists of three different systems in the 5 GHz frequency band:
- high-speed wireless access network (HiSWAN)—similar to HiperLAN/2;
- wireless Ethernet—similar to IEEE 802.11a; and
- wireless home link (based on IEEE 1394, also called FireWire).

The Mobile Operator WLAN solution can be adapted to work with all WLAN standards. Release 1 supports the dominant WLAN standard, IEEE 802.11b, which operates in the 2.4 GHz band at data speeds of up to 11 Mbit/s.

Conclusion

The unlicensed frequencies and high user data rates of WLAN systems make them interesting for cellular operators and ISPs. By connecting a WLAN system to their core network, operators can cover crowded hotspot areas in indoor public environments and provide complementary high-speed data access.

The Ericsson Mobile Operator WLAN solution complements the portfolio of value-added radio-access technologies. The increasing number of WLAN deployments in the corporate and SOHO market segments support this development. Both market segments are early adopters of WLAN technology, in particular because the same technology and set-up apply in the office, at home, and on the road. Public operators can thus offer the same service quality and the low latency currently experienced in the fixed office infrastructure. In addition, the support of roaming between WLAN deployments and between WLAN and the cellular wide-area coverage will further boost growth.

The Mobile Operator WLAN solution will give broadband mobile public access to the Internet and to corporate intranets with relatively little additional investment. WLAN will support all GPRS and UMTS packet services on standard user equipment (laptops).

Work on the architecture and protocols of the Mobile Operator WLAN solution is currently being specified in 3GPP and WLAN standardization bodies. GPRS and UMTS specifications will be enhanced to permit the reuse of subscriber management, security management, and charging and billing functionality.

WLAN technology serves as a broadband complement to GPRS, UMTS and other second- and third-generation systems, bridging the gap, as it were, between different environments and applications. The perceived quality of service over the public WLAN access will not differ from WLAN access in the office.

REFERENCES

1 Telenor and Ericsson: Wireless LAN as a complementary UMTS access technology. White paper, 2001
2 "Information Technology – Telecommunications and Information Exchange between Systems – Local and Metropolitan Area Networks – Specific Requirements – Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," ISO/IEC 8802-11:1997(E), 1997IEEE 802.11, 1997
3 B.H. Walke, "Mobile Radio Networks." John Wiley and Sons Ltd, 2001
4 C. Rigney et al., "Remote Authentication Dial In User Service (RADIUS)," RFC 2865
5 M. Wahl et al., "Lightweight Directory Access Protocol (v3)," RFC 2251
6 HIPERLAN/2 Global Forum: http://www.hiperlan2.com
7 "Information Technology – Telecommunications and Information Exchange between Systems – Local and Metropolitan Area Networks – Specific Requirements – Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," ISO/IEC 8802-11:1999(E), 1999
8 "Information Technology – Telecommunications and Information Exchange between Systems – Local and Metropolitan Area Networks – Specific Requirements – Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, High Speed Physical Layer in the 5 GHz Band," ISO/IEC 8802-11:2000(E), 2000

Previous issues

No. 4, 2001

Secure electronic transactions—The mobile phone evolution continues
On-demand mobile media—A rich service experience for mobile users
WAP 2.x architecture—Features, services and functions
RF multicarrier amplifier for third-generation systems
Ericsson's family of carrier-class technologies
Building a reliable, cost-effective and future-proof fiber optical access network

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

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

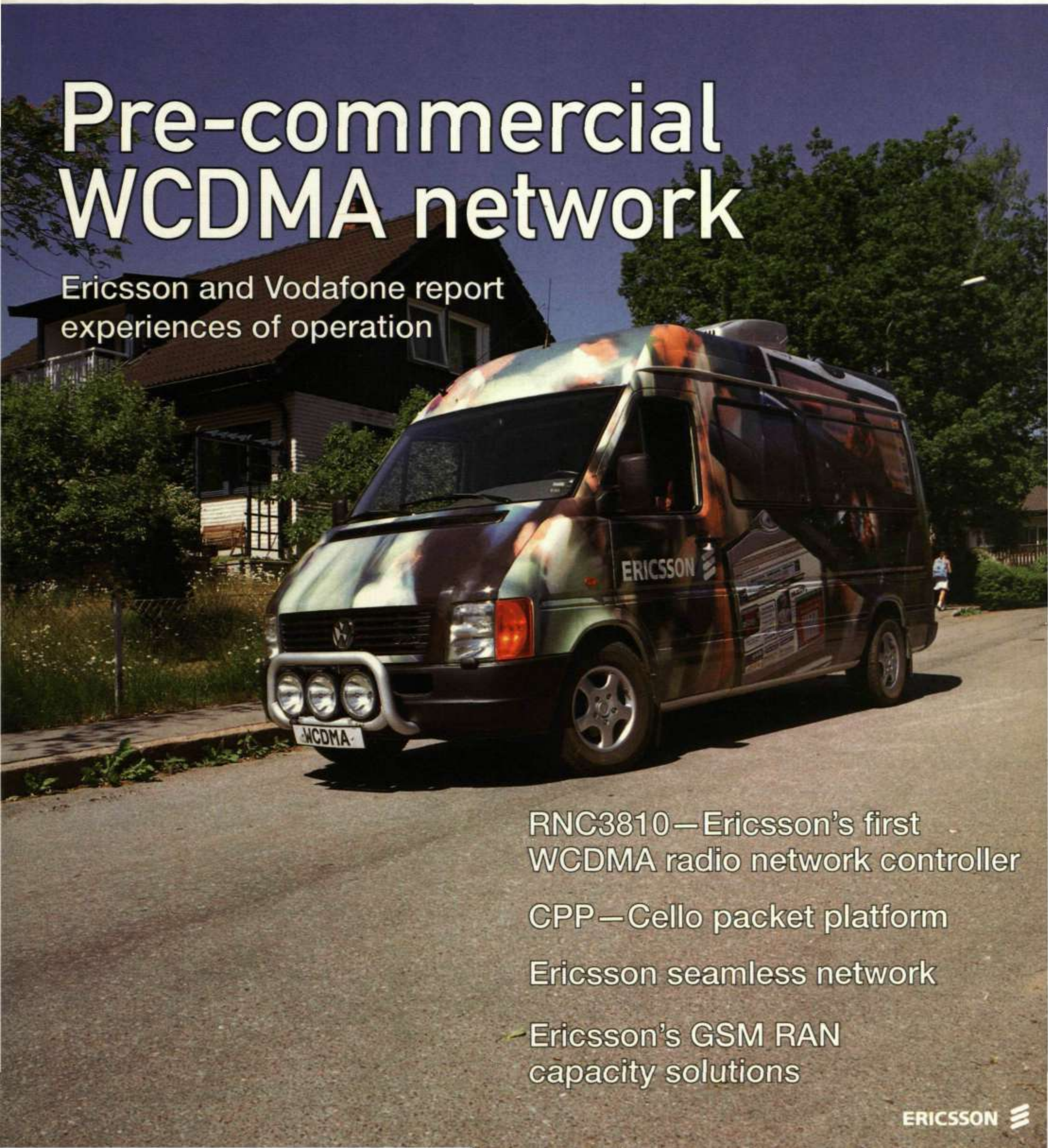
ISSN 0014-0171
Ljungföretagen, Örebro 2002

Ericsson

2/2002

REVIEW

THE TELECOMMUNICATIONS TECHNOLOGY JOURNAL



Pre-commercial WCDMA network

Ericsson and Vodafone report experiences of operation

RNC3810—Ericsson's first WCDMA radio network controller

CPP—Cello packet platform

Ericsson seamless network

Ericsson's GSM RAN capacity solutions

ERICSSON

Cover: To evaluate various aspects of operating a joint, pre-commercial WCDMA network, Ericsson and Vodafone fitted numerous cargo vans with roof-mounted antennas, GPS receiver, uninterruptible power supply and a workbench environment. One van served as a "test" vehicle; the others served as "interferer" vehicles. Each interferer vehicle contained several mobile terminals. Drive tests were conducted under realistic operating conditions in a good mix of rural, suburban and dense urban environments.

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 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:
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: 79, 2002

Contents

Contributors

In this issue

Anders Birkedal joined Ericsson in 1986, initially working with the development of satellite communications equipment. From 1992 to 1995 he was in charge of standardization of the PDC mobile system. From 1995 to 1998 he served as general manager for the Ericsson branch office in Nagoya, Japan. In 1998 he was appointed general manager of marketing for WCDMA systems at Ericsson in Japan. Since returning to Sweden, in 1999, he has worked with project management for WCDMA development, and managed the evaluation project for the experimental WCDMA system, and the subsequent evaluation project for the pre-commercial WCDMA system. He is currently senior project manager for the market introduction of WCDMA Release 3. Anders holds an M.S. in engineering physics from the Lund Institute of Technology and an MBA from the University of California at Berkeley.
anders.birkedal@era.ericsson.se

Peter Blom is currently a product manager for GSM radio access networks. Since joining Ericsson, in 1990, he has held various positions in mobile systems, including three years as strategic product manager for high-capacity radio network solutions for the GSM radio access network. He holds an M.S. in electrical engineering from the Chalmers University of Technology, Sweden.
peter.blom@era.ericsson.se

Eddie Corbett is a master systems engineer at Ericsson Systems Expertise in Ireland. He joined Ericsson in 1994 and has worked as a system design engineer developing radio network algorithms for Ericsson's products for TDMA, CDMA2000 and WCDMA. In 1998, he joined WCDMA Radio Access Network Design to work on the development of power control algorithms. During the field trial project he worked with the verification of radio network algorithms. He currently works with WCDMA radio resource management. Eddie holds a B.E. in electrical engineering from University College Cork, Ireland.
eddie.corbett@era.ericsson.se

Bengt Gestner is presently product manager for the WCDMA RNC node at Core Unit Radio Network Development (CRND). From 1979, when he joined Ericsson, through 1995, Bengt held various positions working in AMPS, and later, TDMA design organizations for radio network products. Between 1995 and 2000, he served as product manager for different TDMA RBS products and for transmission solutions in the access network.
bengt.gestner@era.ericsson.se

Roland Heickerö joined Ericsson in 2000, serving as senior product manager at Ericsson Radio System, WCDMA and GSM total product management. He is responsible for the roadmap and end-to-end solution portions of the Ericsson Shared Network project. Roland has extensive experience of working with technical and economic aspects of research and development strategies. He holds an M.S. in mechanical engineering and a Ph.D. in industrial economy and organization/works science from the Royal Institute of Technology. He is also associate professor in innovation technology at the University of Mälardalen, Sweden.
roland.heickero@era.ericsson.se

Karim Jamal joined Ericsson Research in Stockholm, Sweden, in 1990, to work on adaptive receiver algorithms for second-generation mobile systems. In 1994, he was involved in research on third-generation radio access technology. During 1996 and 1997 he was with Ericsson Japan, where he assisted with the WCDMA standardization efforts within ARIB and continued research on WCDMA physical layer procedures and algorithms. In 1998, he joined the Radio Network Development unit at Ericsson in Stockholm, Sweden, where he contributed to ETSI and 3GPP standardization work, and helped develop commercial WCDMA products. Since November 2000 he has been working with WCDMA field trials at Ericsson UK. He was appointed senior specialist in CDMA signal processing in 1998. Karim holds an M.S. in engineering physics and electronic engineering from the Linköping Institute of Technology.
karim.jamal@era.ericsson.se

Stefan Jelvin currently works with strategic product management at the Ericsson Mobile Systems WCDMA and GSM busi-

ness unit. In particular, from a systems and market perspective, he works with the evolution of second-generation systems to third-generation systems, defining the opportunities and benefits of evolving GSM and integrating it with WCDMA. Stefan joined Ericsson in 1996, and has held marketing and market strategy positions for TDMA and GSM. He holds an M.S. in mechanical engineering from the Royal Institute of Technology, Stockholm.
stefan.jelvin@era.ericsson.se

Bodil Josefsson, who joined Ericsson in 1991, is program manager for Ericsson Seamless Network at the Mobile Systems WCDMA and GSM business unit. From 1991 to 1993, Bodil worked as product manager for transport networks. In 1993, she moved to Ericsson Radio Systems, where the main focus of her work has been on marketing and business management. She holds an M.S. from the Linköping Institute of Technology, Sweden.
bodil.josefsson @era.ericsson.se

Lars-Örjan Kling is an expert in forwarding technology at the CCND Connectivity Development Center, Ericsson AB, Sweden. He joined Ericsson in 1980 to work as a computer architecture designer. In the years since, he has been active in the areas of multiprocessing, mathematical specification methods and neural computation. He entered the datacom area in 1997. He is currently involved in the introduction of IP to the Cello packet platform. Lars-Örjan holds an M.S. in electrical engineering from the Royal Institute of Technology, Sweden.
lars-orjan.kling@uab.ericsson.se

Åke Lindholm joined Ericsson in 1974, and has been working with data communication since 1977. During his time at Ericsson he has worked with the design of local area networks, fiber-optic networks, interprocessor communication, ATM, and IP routers. Today, he is an expert on IP transmission protocols at the CCND Connectivity Development Center, Ericsson AB, Sweden. Åke holds an M.S. in electrical engineering from the Chalmers University of Technology, Sweden.
ake.lindholm@uab.ericsson.se

Lars Marklund joined Ericsson in 1981 as a software developer for AXB-30. He has also helped develop software and design systems for the AXE-10 group switch and ISDN. In particular, he helped develop the SMS subsystem in AXE-10 for GSM. In 1999, Lars became system manager for the Cello packet platform with responsibility for IP-related development.
lars.marklund@uab.ericsson.se

Gunnar B. Nilsson joined Ericsson in 1982. During his 20 years with the company he has served in various positions, including product manager, technical manager and project manager. Most recently, he has worked on Cello packet platform (CPP) development projects. At present, for example, he is in charge of the project to implement IP in CPP. He holds an M.S. in electrical engineering from Chalmers Institute of Technology, Göteborg, Sweden.
gunnar.b.nilsson@uab.ericsson.se

Bengt Persson has been a product manager for the WCDMA RNC and its functionality since 1997. He joined Ericsson in 1978 to help develop machine-to-machine communication, but soon moved to system design and, later, product management. His early responsibilities included standardization and solutions for wide area networks. Over time the emphasis moved to products for local area networks and IP networks. In the mid-1990s, Bengt served as system design manager for work on the Cello packet platform. He holds an M.S. in engineering physics from the Chalmers Institute of Technology, Sweden
bengt.persson@era.ericsson.se

Keith Woodfield joined Ericsson in 1996, working as a product manager for Ericsson's point-to-multipoint radio access products. In 1998, he joined Ericsson's UMTS group in the UK to work on the experimental WCDMA system. This assignment involved establishing a live test network, working with GSM operators in the UK to promote the use of WCDMA, and performing many test cases in support of the WCDMA evaluation project. Keith is currently a technical manager at the WCMDA RAN Competence Center in Guildford, UK. He holds a B.S. in electronic engineering from University College London.
keith.Woodfield@etl.ericsson.se

Ericsson Review is also published on the Web. Visit us at http://www.ericsson.com/review

Editorial

Eric Peterson

I spend most of my working hours at the editorial office or at home (I telecommute quite a lot). But occasionally, I also pay a visit to one or another of Ericsson's research or development centers to meet with authors and to be briefed on various projects. I cherish these visits for the knowledge and contacts they give me, but also (and perhaps more especially) because I always come away feeling totally energized and enthused. Why? Because the people I meet and see are totally energized and enthused. Simply put, they love their work and it shows! This is a side of Ericsson that most people do not get to see. And in my opinion that is a real shame.

When I was a young boy, my father gave me a crystal radio set. Okay, I admit it. I didn't "get it" at first. The kit was simple and, in terms of esthetics, it was completely plain. Plain boring, I thought. The circuit consisted of an inductor (also called a coil), a variable capacitor, a germanium diode (formerly called a crystal), a filtering capacitor, and a very high-impedance earphone. "This is really odd. Dad seems unusually excited about this project. Could there be more to it than meets the eye? I guess I'd better humor him." And so, my father and I sat down together and began to assemble my new radio. "No speaker, huh? Nope. Apparently, you have to listen through the earphone. No power cord or place for batteries either. Hey, wait a minute! How does this thing work?" And then my father explained that this radio required no power other than that provided by the transmitting antenna from the radio station. Now that was something different. That was exciting! (Still is for that matter.)

It is this sense of discovery, of wonder, which I first felt as a young boy with a crystal radio set, that has characterized Ericsson since the days of its founder, Lars Magnus. Today, this same feeling continues to fill and invigorate Ericsson employees around the world. It is the fuel in Ericsson's motor. I have witnessed it in my travels to Canada, Germany, Japan, Singapore, the UK, the USA, and here at home in Sweden, and I am certain that the same holds true for the rest of Ericsson. If you haven't seen this side of Ericsson before, please take the opportunity to get better acquainted. You won't be disappointed. Under its cool, Swedish exterior, this company is still bubbling with excitement!

Eric Peterson
Editor

Experiences of operating a pre-commercial WCDMA network

Anders Birkedal, Eddie Corbett, Karim Jamal and Keith Woodfield

Ericsson and Vodafone have jointly built a pre-commercial WCDMA network in the UK. During 2001 and the first half of 2002, the network has been used to evaluate various aspects of operating a WCDMA network, and for preparation for commercial services scheduled to start in 2002. The work mainly focused on evaluating the basic radio network characteristics of the WCDMA standard in a live environment.

The testing of a complete network under realistic operating conditions based on commercial hardware and pre-releases of software has been extremely valuable to Ericsson and Vodafone.

The authors describe the network configuration, the test setup, various areas of testing, and give some examples of test results.

Introduction

During 2002, operators around the world are planning to introduce third-generation mobile services to their subscribers. Many have selected the WCDMA system from Ericsson for building the infrastructure to provide these services. The WCDMA standard for the mobile network is the core technology for efficient mobile multimedia services and for realizing the vision of true Mobile Internet access. The standard, which has been developed after many years of research at Ericsson, is now the state-of-the-art telecommunications technology. One area of focus of the research has been to get adequate performance in the air interface (between the network and user equipment) in order to handle the new mobile data services such as Internet access, MMS, video calls, and IP multimedia calls.

For commercial operations of third-generation mobile networks with paying customers, Ericsson is currently introducing WCDMA Release 2. Release 1 was introduced in 2001 for use as a pre-commercial service by operators. It was used for the main part of the joint Ericsson-Vodafone pre-commercial tests. Release 1 has also been used for demonstrating WCDMA features, network roll-out, network tuning, and for showing compliance with technical requirements from regulators.

To ensure the expected performance at the commercial introduction of the WCDMA system, Ericsson built a pre-commercial network in cooperation with Vodafone UK. The network has been used to test various operating aspects in a realistic environment before delivering hardware and software to

BOX A, TERMS AND ABBREVIATIONS

2G	Second-generation mobile telecommunication systems
3G	Third-generation mobile telecommunication systems
3GPP	Third-generation Partnership Project
ALEX	Active library explorer
ATM	Asynchronous transfer mode
BCCH	Broadcast control channel
BLER	Block error rate
BSC	Base station controller (GSM standard)
BTS	Base transceiver station (GSM standard)
CPP	Cello packet platform
CTR	Cell traffic recording
DHCP	Dynamic host configuration protocol
DNS	Domain name server
E_c/N_o	Chip energy per noise spectral density
FTP	File transfer protocol
GPEH	General performance event handling
GPRS	General packet radio service
GPS	Global positioning system
GSM	Global system for mobile communication
GSN	GPRS support node
IP	Internet protocol
ISDN	Integrated services digital network
ISUP	ISDN user part
Iu	Interface between CN and RAN in the WCDMA standard
Iub	Interface between RNC and RBS in the WCDMA standard
Iur	Interface between RNC and another RNC in the WCDMA standard
MMS	Multimedia messaging service
MSC	Mobile switching center
Mub	O&M interface for RBS (between RBS and WCDMA OSS)
Mur	O&M interface for RNC (between RBS and WCDMA OSS)
NBAP	Node-B application part
Node-B	Radio base station in 3GPP standard
NTP	Network time protocol
O&M	Operation and maintenance
OSS	Operations support system
PLMN	Public land mobile network
PSTN	Public switched telephone network
RAB	Radio access bearer
RAN	Radio access network
RANAP	RAN application part
RBS	Radio base station (in WCDMA system)
RNC	Radio network controller (in WCDMA system)
RNS	Radio network subsystem
RNSAP	RNS application part
RRC	Radio resource control
RSCP	Received signal code power
SIR	Signal-to-interference ratio
TEMS	Test Mobile System (product name)
TU	"Typical urban" channel model
UE	User equipment (handset, terminal)
UETR	User equipment traffic recording
Uu	Interface between RAN and UE (air interface) in the WCDMA standard
VA	Voice activity
VAF	VA factor
VP	Virtual path
WCDMA	Wideband code-division multiple access

operators. The core network parts of the system have been developed as an ongoing evolution from second-generation systems, whereas the radio access network represents a fundamental shift in technology, to achieve substantial gain in performance. Therefore, the activities associated with the pre-commercial network have focused on basic radio characteristics that have a significant influence on overall system performance and represent a major technical challenge for third-generation systems. The main objectives have been

- to evaluate the behavior of a complete WCDMA system in a real-life environment;
- to evaluate the performance and characteristics of the WCDMA radio access network (RAN); and
- to tune RAN system parameters under realistic radio conditions.

The network, which consists of 30 radio base stations (RBS) distributed to cover different types of radio environments, was taken into operation with WCDMA Release 1 software in April 2001, when the first live WCDMA call was made. Release 2 of the software was introduced for testing in the spring of 2002.

Network configuration

General

The pre-commercial network consists of 30 radio base stations, two radio network controllers (RNC), a supporting core network, and service network nodes. Twenty-five of the RBSs, located in the Thames Valley region west of London, have been arranged to provide coverage over an area of approximately 300 square kilometers. The coverage area includes a good mix of rural, suburban and dense urban environments. Motorway coverage is also available to perform drive tests at higher velocities. Apart from the 25 RBSs, all other equipment is located at Ericsson's and Vodafone's premises in Newbury, UK.

The network was split into two separate radio network subsystems (RNS) so that a variety of tests could be executed simultaneously. Adequate flexibility was provided to enable the RBSs to move between RNSs, thus allowing each network to grow or shrink according to requirements. For some tests, the entire network of 25 field RBSs was parented to one RNC.

A third RNC was added to test the interface between RNCs (*Iur*).

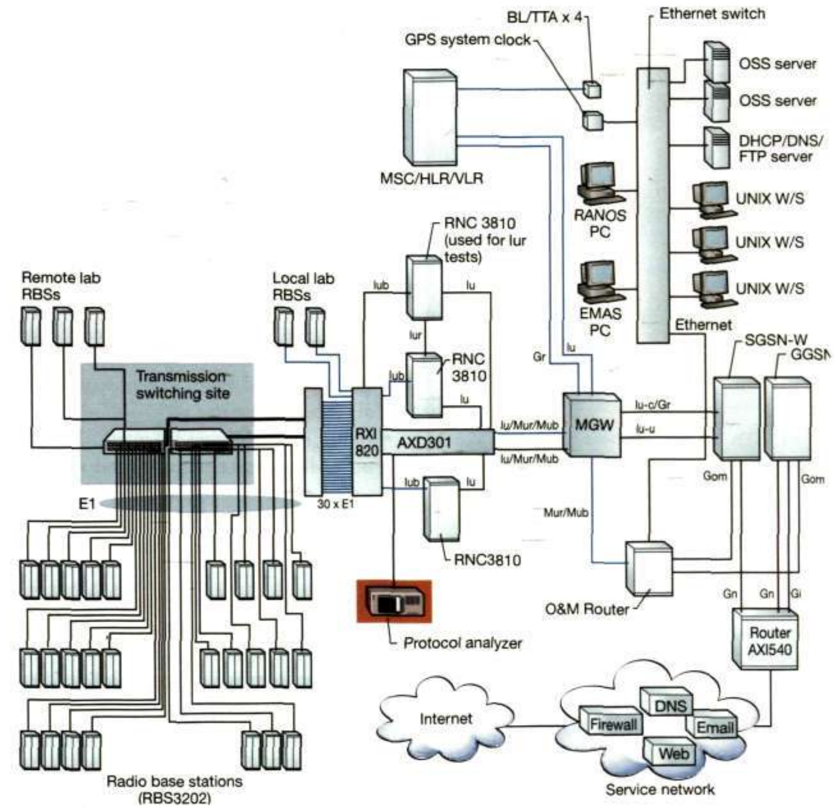


Figure 1
Block diagram of the pre-commercial network.

The key objective was to evaluate Ericsson's WCDMA radio access network in a live environment. The core network, transport network, and service network nodes were installed to allow this evaluation to take place (Figure 1).

Radio base station

The network is composed of 30 radio base stations configured with one carrier frequency per sector (Figure 2). Twenty-eight of the RBSs have three sectors (3x1). At two field sites the RBSs have been configured for two sectors. The Ericsson RBS3202 is based on the Cello packet platform (CPP), a high-performance ATM switch.[1-3] The key functions of the RBS are

- layer-1 termination of the air interface;
- inner-loop power control;
- softer handover combining and splitting; and
- operation and maintenance (O&M).

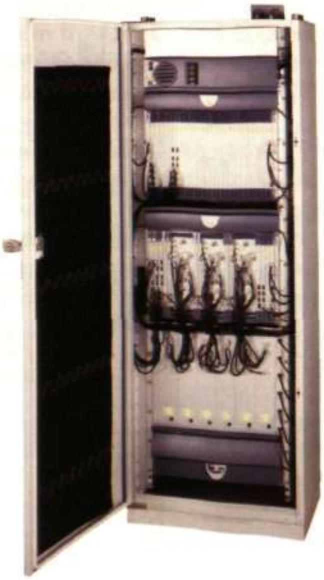
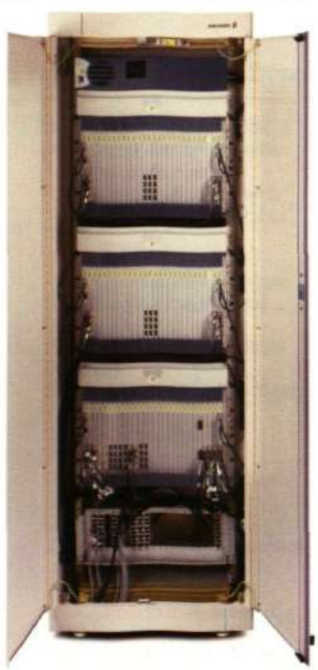


Figure 2
The Ericsson RBS3202.

Figure 3
The Ericsson RNC3810.



Radio network controller

The radio network controller—RNC3810—is also based on CPP (Figure 3). Early test cases were executed on a single subrack configuration; during the trial, however, the RNC was upgraded to a two-subrack configuration to keep it aligned with the majority of commercial roll-out projects. The RNC terminates the *Iub* interface to the RBSs, the *Iu* interface to the core network, and the *Iur* interface to other RNCs (to support inter-RNS handover). The key functions of the RNC are

- layer-2 and the radio resource control (RRC) part of layer-3 termination of the air interface;
- radio resource management, including outer-loop power control and admission and congestion control (capacity management);
- control of mobility, including soft handover;
- the provision of transparent bearer services between the core network and user equipment (UE);
- the management of the supply of bearer services with traffic classes to the core network; and
- operation and maintenance.

ATM switch—AXD301/RXI820

Each RBS is connected to the network control center by a dedicated 2 Mbit/s (E1) leased line. On the *Iub* interface, all traffic and signaling between the RNC and each RBS are carried over a single ATM virtual path (VP). The AXD301 is used for switching virtual paths from all RBSs onto one of two 155 Mbit/s (STM-1) connections to the RNCs (*Iub* grooming). Reconfiguration of these virtual-path cross-connects and ATM and IP configuration changes in the RNC and RBS allow the RBSs to be re-parented between the two RNCs. All connections from the RNC to the core network via the *Iu* interface are also connected via the AXD301. This means that *Iu* and *Iub* interfaces for all nodes are available in the AXD301. The addition of extra virtual-path cross-connects and STM-1 interfaces made it possible to monitor these interfaces on a protocol analyzer in an efficient manner. During the project, the AXD301 was supplemented with a CPP-based RXI820, which took over the function of *Iub* grooming.

Media gateway

The media gateway (also based on CPP) provides the interface between the RNC and the core network. In the pre-commercial network, the media gateway was used as an ATM cross-connect. All *Iu* user plane and control, and O&M interfaces of the RNC (*Mur*) and RBS (*Mub*) are passed to the media gateway via a single connection from the RNC. The media gateway serves to split the *Iu* traffic between the

- mobile switching center (MSC), which handles voice and circuit-switched traffic; and
- GPRS support node (GSN), which among other things, handles packet-switched traffic in WCDMA networks.

The media gateway, which is a single subrack solution, also provides interfaces to the O&M router.

Core network

The core network consists of

- MSC;
- GSNs;
- home location register (HLR);
- authentication register;
- flexible numbering register;
- equipment identity register; and
- ISUP/ISDN protocol converter and PSTN interface for testing 64 kbit/s circuit-switched data calls.

WCDMA OSS

The WCDMA operations support system (WCDMA OSS) is a set of software for handling operation and maintenance tasks for the WCDMA radio access network. It gives a consolidated view of information on the radio access network, such as alarms, configurations and basic performance. It also provides several interfaces for easy integration into the existing management environment.

Operation and maintenance infrastructure

O&M router

The main purpose of the O&M router is to act as an interface between the IP-over-ATM-based WCDMA radio access network and the IP-over-Ethernet-based O&M network, which consists of network servers and workstations.

Network servers

A variety of UNIX and Linux-based servers operate as

- WCDMA OSS servers;
- FTP/DNS/DHCP server;
- NTP server, including GPS (for network synchronization);

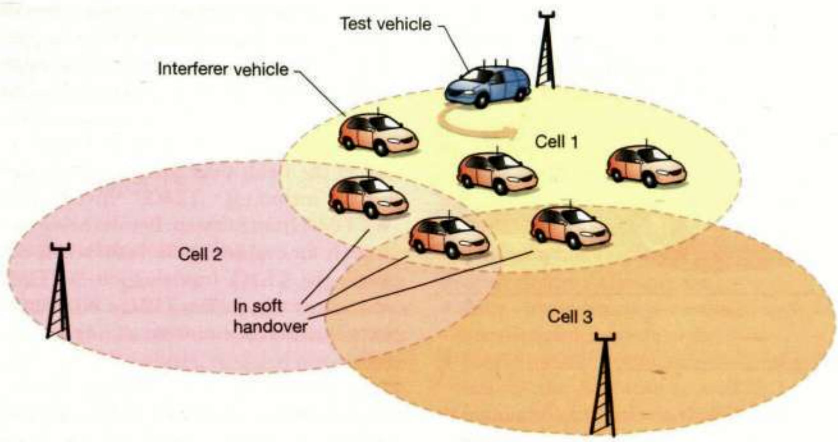


Figure 4
Setup for testing system load.

- network address translation (NAT);
- Internet firewall;
- Web server; and
- active library explorer (ALEX) server—for storing customer product information.

Workstations

A variety of UNIX, Linux and Microsoft Windows-based workstations are used as
- RBS and RNC element managers;
- WCDMA OSS clients; and
- UNIX workstations for command line interface for data logging and debug tools for the RNC and RBS.

Test setup and tools

When testing traffic functionality, we generally distinguish between
- unloaded tests, where only a single UE is used; and
- load tests, where many UEs are used.

The basic WCDMA traffic functions, such as power control and soft handover are first assessed in an unloaded situation. This ensures that the functions work as expected when the system is not under stress—that is, when interference is largely stationary and the main contributor is thermal noise.

In a loaded WCDMA system, interference, as seen by one radio link, consists of thermal noise and other user signals. Interference can thus vary considerably, in particular when data users are few. Since the inner-loop power control mechanism in both links is based on the signal-to-interference ratio (SIR), the power control characteristics need to be evaluated during load. The same is true for soft handover functionality, as well as the combination of soft handover and power control. Another area of assessment is the basic behavior of the capacity control functions (admission control and congestion control) of the WCDMA radio access network.

Generating real load on the air interface for controlled tests on a live pre-commercial WCDMA system involves considerable logistical planning, since numerous UEs are needed, and each UE must be controlled to some degree. Data is logged extensively for one specific UE. The other UEs generate interference in a realistic way. Figure 4 shows a typical setup for a load test, where several "interferer" vehicles are used. Each interferer vehicle contains several UEs; the "test" vehicle contains the UE being used for data logging. The interferer vehicles either hold stationary positions throughout the cell or drive at random while the test vehicle typically drives along a specific drive route. The same basic setup can be repeated for various types of environment.

The capacity control functions are tested by adding calls in a cell using stationary or moving UE clusters within the cell and observing the network-side measurement entities used by the capacity control functions, and the actual behavior of the functions themselves. The unloaded tests are easier to perform, since only the test vehicle is needed.

To accommodate convenient UE-side handling and data logging, several cargo

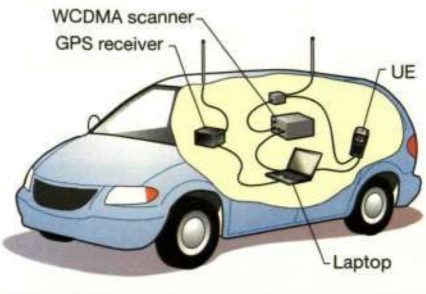


Figure 5
Test equipment installation in the cargo vans.

vans were fitted to include roof-mounted antennas, GPS receiver, uninterruptible power supplies and a workbench environment (Figure 5). The UEs were pre-commercial prototypes of WCDMA terminals from Sony Ericsson and a variety of other UE suppliers. All logs on the UE side were collected using laptop computers connected to the UEs and running special logging software. The internal clock of each laptop computer was synchronized to GPS time for subsequent time correlation to network-side logs.

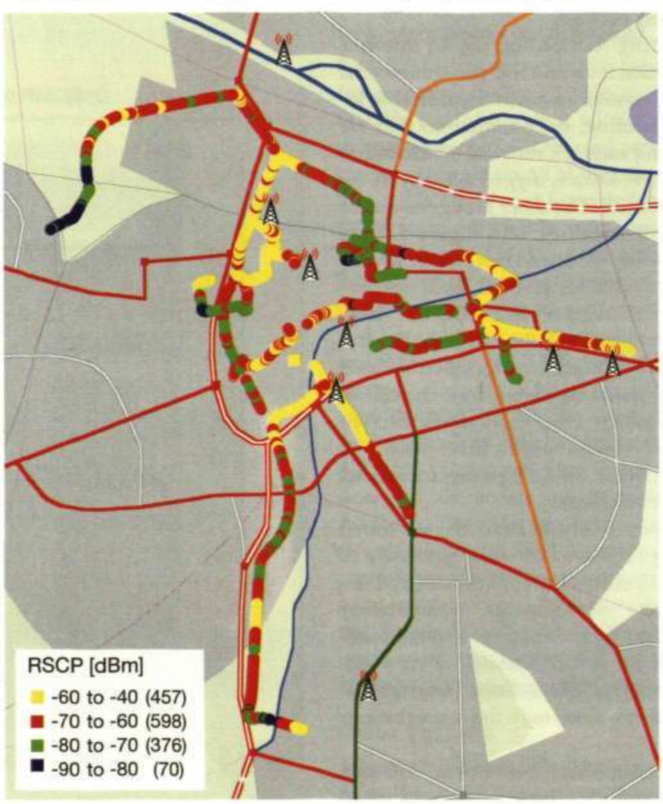
For enhanced observability, and to get a second independent reading of the cell status in the field, pilot scanning tools were used, including TEMS Investigation WCDMA from Ericsson. Besides being used as tools for evaluating the radio access network, the TEMS Investigation WCDMA and other tools, such as TEMS Cell Planner, were themselves evaluated during the live tests.

To increase the interference generated during testing, the interferer UEs used the built-in antenna instead of the roof-mounted antenna—the use of built-in antennas increases path loss.

On the network side, the main observation points were the RBSs, the RNCs, and the *Iub* and *Iu* reference points. A protocol analyzer was used to monitor the *Iu*, *Iub* and *Iur* reference points (note also that *Uu* protocols can be monitored by connecting a protocol analyzer to *Iub*). The main protocols of interest were RRC (*Uu*), NBAP (*Iub*), RANAP (*Iu*) and RNSAP (*Iur*). In addition, the UE traffic recording (UETR) function was used for some entities. With these procedures it was possible to log layer-3 messages and to measure signal quality—for example, block error rate (BLER). All nodes in the network and the protocol analyzer ran an NTP client that was synchronized to NTP server software. This, in turn, was synchronized to GPS to ensure that the time-and-calendar function in each node was synchronized. The network logs could thus be correlated with those collected on the UE side. Special parser software was developed to extract relevant entities from the network logs. All resulting log files were saved together with the UE log files for further processing and analysis.

In WCDMA, the source activity directly influences how much power is being transmitted from the UE or RBS. Therefore, it also directly affects interference in the system. For this reason, to reproduce tests and to interpret the results properly, the source activity for both the "test" and "interferer" UEs had to be controlled. Known source data statistics as well as tools that generate reproducible data streams were used for packet-data calls. For UE-to-fixed-line voice calls, an announcement machine was dialed. The announcement machine was fed by a voice recording with controlled voice

Figure 6
Measured RSCP throughout the drive route through an urban RBS cluster.



activity (VA) statistics that correspond to typical voice. In the uplink, a similar recording was fed to the UE microphone. This solution was also used at both ends for UE-to-UE voice calls.

Test objects

Radio network planning

Tools for planning and dimensioning WCDMA radio networks are included as part of the TEMS family of products from Ericsson. To evaluate, tune and improve the models used to describe the WCDMA radio network, early feedback from the pre-commercial network is essential. Measurements were made

- to verify that the coverage for voice and data services meets expectations as predicted by the planning tools;
- to verify that areas where the mobile terminals are in soft or softer handover state are in line with predictions;
- to evaluate performance and usability of radio network planning tools;
- to create an understanding of the effect of key parameters on the radio network; and
- to investigate suitable measures of radio network performance.

To verify the network plan, we measured the pilot signals from the radio base stations. Although limited to a few common channels, the pilot survey can give an indication of the coverage, interference and soft handover that can be expected during real commercial operations.

We can verify the cell plan by studying the results of the pilot survey. We can also identify potential interference situations through the detection of many equally strong pilot signals. Figure 6 shows a plot of the received signal strength from the radio base stations for a part of the network. Figure 7 shows the handover regions in the same area using the same drive route. This test case shows an urban cluster made up of 22 active cells.

Operation and maintenance

Another objective for testing was to assess the basic O&M functionality and characteristics in a live network environment. Areas of particular relevance include the stability of the system, the efficiency of upgrading software, and the general ability to observe the system in terms of performance management and alarms.

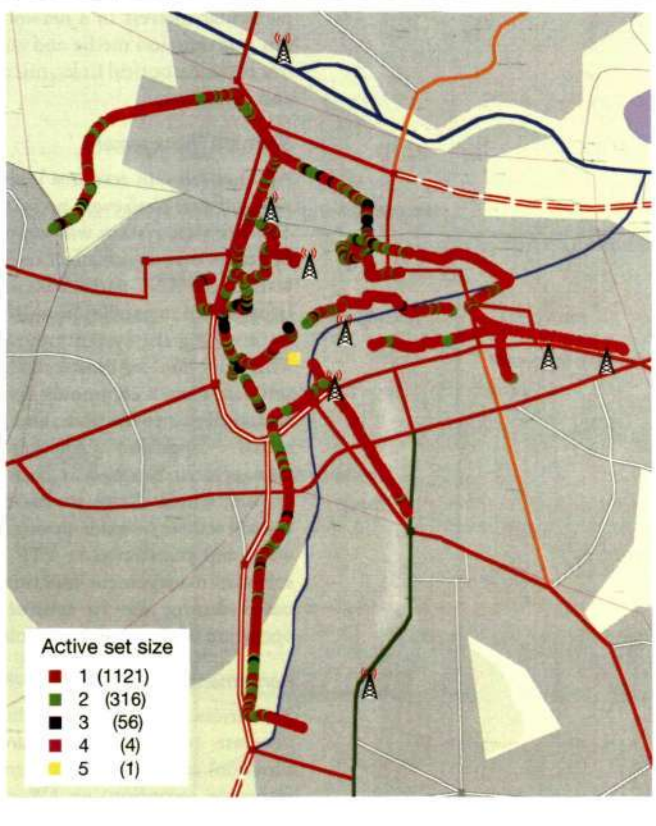
Ericsson's management solution provides an integrated management view of the overall network. The architecture structure contains embedded element management and subnetwork management. The embedded element management is a fundamental principle in the O&M system. Each network element (RNC or RBS) contains software for element management—that is, the management of that specific network element.

The main functions supported by the embedded element manager are

- fault management—viewing and acknowledging active alarms;
- configuration management—configuration of the hardware as well as setting radio and transport network parameters; and
- software management—upgrading the network element with new software and taking backup copies of the active configuration.

Ericsson's subnetwork manager for the WCDMA radio access network (WCDMA

Figure 7
Number of legs in the active set throughout the drive route through an urban RBS cluster. The active set includes legs within 3 dB from the main signal.



OSS), supports the management of a single network element or coordinated management of multiple network elements. As with element management, the main functions supported by WCDMA OSS are fault management, configuration management, and software management. However the scope of management is the entire radio network including the transport network. In addition, WCDMA OSS embodies several other important features including product inventory functions and performance management functions, such as performance statistics and performance recording.

Stability and fault management

Typical tests of stability and fault management include *Iub* link breakage—for example, to simulate a microwave link in bad weather. Another interesting type of test is to remove boards from nodes during operation and when calls have been set up. When removed, only directly affected traffic should be lost and proper alarms should be generated. After re-insertion, the boards should become fully operational. Other areas include recovery testing, such as node restarts and the accuracy and stability of node synchronization. This latter test is of particular interest in a network with various transmission media and configurations (for instance, optical links, microwave links, and cascading).

Software management

Another relevant test area is the loading of new software onto every node in the system. The Ericsson system will support software upgrades that do not affect traffic. The software of the RNC and RBSs can be remotely upgraded in parallel from WCDMA OSS by invoking the system upgrade function. This function automatically fetches new software from a commonly accessible server, transfers it to the node, and performs the actual upgrade. Another software-management function is backup administration, which allows for backup copies of the software to be made throughout the network and transferred to FTP servers. The software-management functions were evaluated during use in routine day-to-day operation of the pre-commercial network.

Performance management

The Ericsson system allows the operator to generate statistics (performance statistics function) and make event recordings (performance recording) on UE or cells—UE traffic recording (UETR) and cell traffic recording (CTR). The operator may also set up events to trigger recordings or statistics collection—general performance event handling (GPEH). The performance statistics function collects samples over a given RNS from various operator-defined entities. Examples include number of call setup failures, number of admission rejections, and their causes. The results are collected in files that may be retrieved from the nodes through WCDMA OSS. After having verified the performance-management functions, they were used as tools to test traffic functions.

Mobility functions

Cell selection/reselection

The cell selection/reselection function ensures that the UE successfully camps on the cell that offers the best quality. Note that this procedure is used when the UE is in idle mode as well as when there is a packet-switched connection.

Given a predetermined PLMN, the UE uses a cell-selection procedure to find a cell on which to camp. Immediately after the UE has powered on, the initial cell-selection procedure executes a cell-search procedure, which leads to the synchronization of frequency and time, and acquisition of the downlink scrambling code of the selected cell. On the other hand, when the UE is already camping on a cell and has received a neighbor cell list from the network, the cell search procedure may be simplified, since the scrambling codes of the other cells are already known.

From the camped normal state, the UE evaluates whether or not a new cell should be selected. The measurements are allowed/triggered when

- the quality (measured as E_C/N_0) drops below a given threshold;
- internal UE triggers are used; or
- information pertaining to the broadcast control channel (BCCH) is modified.

When the measurements have been completed, cells are ranked according to standardized criteria and the UE reselects the cell with the highest ranking.

If the UE does not perform cell selection and reselection properly, then it might select a cell with high path loss. This could result in an unnecessary increase in interference in the cell, or a subsequent call set-up might fail. Therefore, the main benefit of testing cell selection/reselection came from assessing

- the probability of camping on the best cell when in an environment that is "polluted" with pilot signals; and
- the execution time of cell selection/reselection, and parameter dependencies.

Soft and softer handover

During soft handover, the UE is simultaneously connected to, is power-controlled by, and controls the power of, two or more cells that belong to different RBSs. The downlink and uplink signals are split and combined in the controlling RNC node. By contrast, during softer hand-over the UE communicates with two or more cells that belong to the same RBS, and the downlink and uplink signals are split and combined in that node. The UE uses the same method for combining downlink signals during soft and softer handover. The set of cells with which the UE is communicating simultaneously is called the active set.

In WCDMA systems, handovers are triggered by the UE in response to events observed among the cells it monitors. When the quality of the downlink of a certain cell fulfills the stipulated criteria, the UE sends a measurement report to the RNC, suggesting, as it were, that the active set should be changed. The RNC then determines whether or not the active set should be changed.

The tests of soft and softer handover were primarily evaluated in terms of

- voice quality—to verify that handovers remain non-audible in various scenarios;
- stability—to tune parameters, to give appropriate levels of active set changes under different conditions;
- handover delay—to assess the time after the radio conditions trigger a measurement report until a change has been made to the active set;
- dropped calls—to assure call retainability at handover; and
- handover gain—to study the reduction in transmitting power obtained with soft and softer handover.

Handover delay is a measure of the time that elapses from the occurrence of an event in the UE until the active set has been updated. There are several components of this delay that can generally be grouped into

- UE delay, which results from internal UE processes;
- network delay, which results from internal radio access network processing and signaling; and
- UE network signaling delay, which results from communication between the radio access network and the UE.

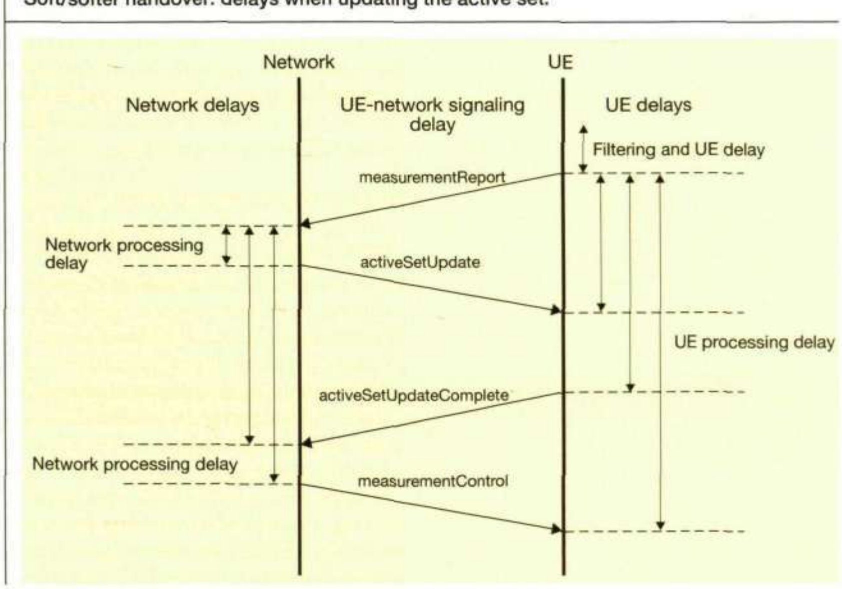
Figure 8 shows the delay. Many components of the delay can be estimated by observing the timing of messages sent between the UE and the network. The measurement control message, which is sent after the active set has been modified, contains an updated list of the cells that the UE is to monitor. The UE cannot generate a new measurement report—and initiate a new handover—until it has received this message and updated its measurement processing accordingly.

The handover delay tests showed that the various components of delay were reasonable-sized contributions to the total handover delay, and that changes to the active set were typically completed within the expected timeframe for the parameter settings used in the test.

Handover between third- and second-generation systems

An important function for the migration from second- to third-generation systems is handover between WCDMA and GSM. In

Figure 8
Soft/softer handover: delays when updating the active set.



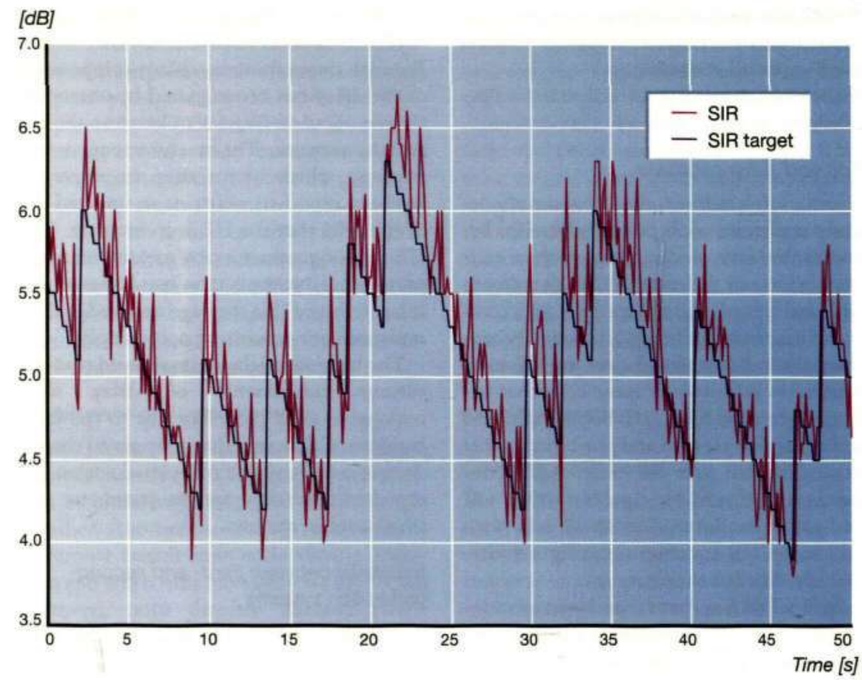


Figure 9
Typical time fragment of uplink SIR target and SIR for an urban environment, 50 km/h, with 100% VAF.

this context, the main characteristics to evaluate are the execution time and end-user perception of handover in various radio-related scenarios. To test this function in the field, a standard operational GSM radio network is used covering the same area as the WCDMA radio network. Some of the GSM BTSs are co-located with WCDMA RBSs. Various dual-mode 3G/2G UEs will be used for tests in this network when they become available from handset vendors.

Radio resource management

Power control

Since every cell uses the same frequency, each user's signal interferes with the signals of other users in the area. Therefore, power control algorithms have a fundamental role in WCDMA. The output power of both the UE and the RBS must be held to the minimum level needed to support each radio link with sufficient quality.

The power control of dedicated physical channels of the uplink and the downlink work in more or less the same way, and generally constitute an inner-loop and outer-loop algorithm. The task of the outer loop is to keep a measured BLER level at a specified BLER target level. This is accomplished by continuously updating an SIR target level, which in turn is used by the inner loop. With the given SIR target level as input, the inner loop strives to maintain a measured SIR level at the SIR target level by requesting that the transmitting side should increase or decrease its output power level.

Thus, by using the outer loop, the connection quality is maintained in terms of keeping the BLER at a predetermined value, rather than keeping the SIR constant. This is relevant since the BLER versus SIR relationship differs for different radio environments and UE speeds.

For uplink power control, the inner loop is managed by the RBS to control the power of the UE. The outer loop is managed by the RNC—that is, the algorithm works on the BLER measured after soft handover combining. The downlink power control loops are managed by the UE, where the inner loop controls the power of the RBS. One objective of the tests was to evaluate the perfor-

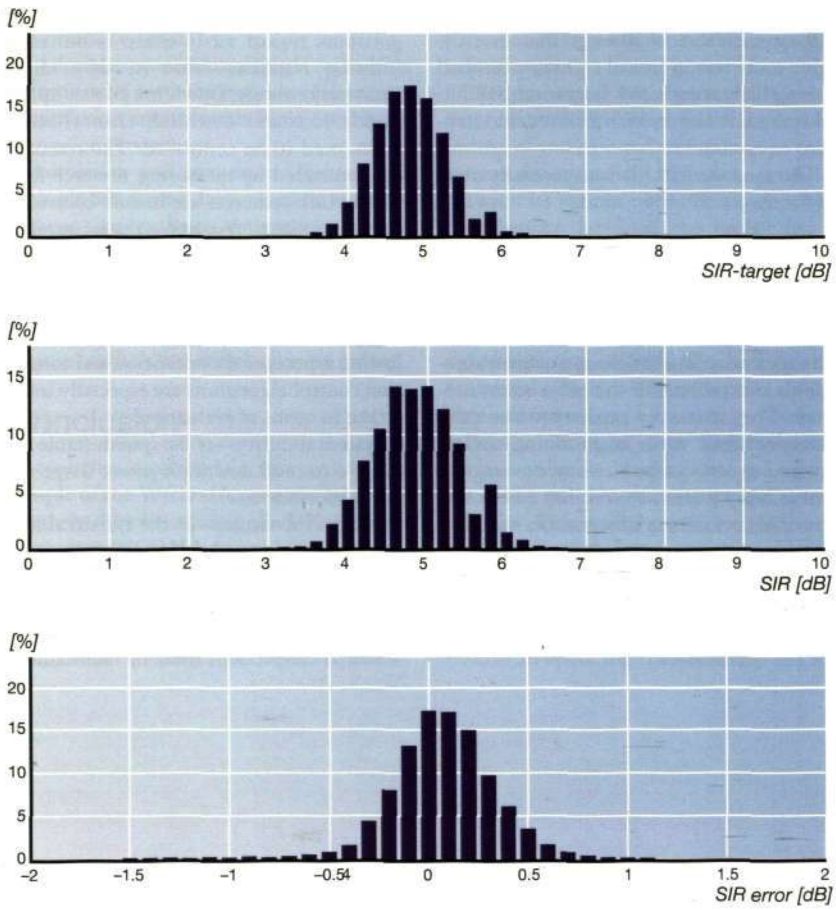


Figure 10
Distribution of SIR target, SIR and SIR error for an urban environment, 50 km/h, with 100% VAF.

mance of the downlink inner-loop power control by plotting BLER versus the SIR target and recording the difference between the SIR target and measured SIR for different environments. We monitored the standard deviation of the measured SIR and BLER and compared it to the test during which downlink outer-loop power control was activated.

Another objective of the test was to evaluate the performance of the downlink inner- and outer-loop power control. From these tests we were able to determine the necessary SIR and SIR target, and standard deviation of measured SIR and BLER for different environments.

The performance of the uplink inner-loop power control was also evaluated. We were able to characterize BLER versus the SIR target, and to observe the difference between the SIR target and measured SIR for different environments. We monitored the standard deviation of the measured SIR and BLER and compared it to the test during which the uplink outer-loop power control was activated.

Similarly, we evaluated the performance of the uplink inner- and outer-loop power control. From these tests we could determine the needed SIR and SIR target, and the standard deviation of measured SIR and BLER.

The uplink tests illustrated in Figures 9 and 10 were performed with the uplink outer-loop algorithm activated with a BLER target of 0.5% and while driving in an urban area at 50 km/h with no antenna in line of sight using a single UE. The variation in path loss along the drive route was estimated to be less than 20 dB. The average SIR target in this case was around 5 dB. For lower speeds, slightly lower values were measured, as expected. A highway drive test gave an SIR target of as low as 4 dB, but since the drive route was in line of sight at nearly all times, this was expected. In general, SIR follows the SIR target, with a reasonable standard deviation.

The measured BLER was generally close to the target of 0.5%.

Admission and congestion control

Admission and congestion control are algorithms at the cell level that prevent and resolve overload situations on the air interface. The admission and congestion control algorithms control load in the radio access network. They maximize capacity in the radio access network while maintaining the requested quality of service and coverage—that is, they guarantee stability in the system. The core network determines the quality of service parameters for each radio access bearer (RAB). The admission and congestion control algorithms control the load in the radio access network by regulating the use of resources.

The admission and congestion control algorithms regard air-interface resources as limiting resources—that is, they might limit performance. Downlink power, uplink interference and downlink channelization codes need to be controlled. The resources are controlled by measuring resource load. The uplink resources are mainly controlled through a virtual resource—the air interface speech equivalent. Which resources might be limiting performance depends on the conditions of the specific cell, including system deployment, and operator control. The following aspects of the admission and congestion control algorithms are especially interesting in terms of evaluation:

- system stability—is the system stable relative to total downlink power usage and uplink noise rise?
- system robustness—is the system able to recover from overload situations?
- system capacity—what is the average capacity provided in a cell? and
- connection quality—the quality of individual connections must be maintained.

TRADEMARKS

Microsoft is a registered trademark of Microsoft Corporation.
UNIX is a registered trademark of the Open Group.

Three types of test case were specified for the performed tests:

1. Admission control in a single cell with a static radio environment—to evaluate the admission control with focus on the admission process.
2. Admission and congestion control in a single cell with a dynamic radio environment—to evaluate congestion control with focus on the congestion process.
3. Admission and congestion control in multiple cells with a dynamic radio environment—to evaluate the mobility settings of admission and congestion control.

Conclusion

The testing described in this article is a sample of the WCDMA system evaluation conducted by Ericsson in the UK, in order to prepare WCDMA technology in general, and the air interface in particular, for full-scale commercial operations. The evaluation can be seen as the last link in a long chain of verification made at different levels of the WCDMA system to ensure that it is ready for release to paying subscribers.

The evaluation has shown that simulations cannot fully replace full-scale, real-life testing of complicated technology to verify system functionality and interoperability between its various parts. This exercise has considerably decreased the probability of unforeseen problems when commercial systems are taken into operation. Moreover, it has given Ericsson the opportunity to tune the default system parameters stipulated in the operating guidelines. This will save operators valuable time when they take the system into operation.

Finally, Ericsson and Vodafone each gained valuable know-how concerning WCDMA radio characteristics and system behavior. This can be applied to improve

- radio network planning tools;
- parameter-setting guidelines;
- system performance in upgraded system releases; and
- skills in operating the network.

ACKNOWLEDGEMENTS

The authors want to thank the staff from Vodafone who participated in this project, and without whose contributions this article would not be possible.

REFERENCES

1 Reinius, J.: Cello—An ATM transport and control platform. Ericsson Review Vol. 76(1999):2, pp.48-55.
2 Berglund; b:, Nygren, T. and Sahlman, K.-G.: RF multicarrier amplifier for third-generation systems. Ericsson Review Vol. 78(2001):4, pp. 184-189.
3 Kling, L.-Ö., Lindholm, Å., Marklund, L. and Nilsson, G. B.: CPP—Cello packet platform. Ericsson Review Vol. 79(2002):2, pp. 68-75.

RNC3810—Ericsson's first WCDMA radio network controller

Bengt Gestner and Bengt Persson

While the role of a WCDMA radio network controller (RNC) is similar to that of a GSM base station controller (BSC), the differences between the two are big enough to justify a new RNC design, even among vendors who have the best BSC technology available.

The authors describe the RNC3810, its hardware and software architecture, and explain how it differs from a BSC.

The role of the RNC

As the cellular industry has evolved from first-generation to third-generation cellular systems, system architects have refined the role of nodes in the network. In hindsight, one can say that the architects of first-generation cellular systems, such as AMPS and NMT, had quite modest expectations regarding the scope of future traffic volumes. Accordingly, a simple architecture could meet these modest needs: all control resided within a single node—the mobile switching center (MSC)—which managed the call, the subscriber, mobility, and the base transceiver station (BTS).

When the system architects defined the architecture for GSM, they drew on experiences of first-generation networks, revising expectations for growth, improving multi-vendor capabilities, operations, and so on. New demands pointed to the need to separate subscriber aspects from radio aspects. As a consequence, the MSC, BSC and BTS were designed as separate nodes. This architecture has served the industry well and allowed GSM networks to grow far beyond early expectations. As the name implies, the base station controller (BSC) is responsible for controlling base stations (BTS), including most handover tasks. All subscriber-related and most call-related tasks have been retained in the MSC.

Later, when work was begun on the architecture for third-generation systems (WCDMA), significant experience had been gained from second-generation systems. Numerous services that had not even been envisioned at the inception had been introduced successfully. However, the radio technology had shown itself to be more noticeable to the core network nodes than was desirable. Given the multitude of services envisioned in the Mobile Internet, the architects understood that they would need to better isolate the radio resources from the core network. They selected GSM as the base for the new architecture, but saw the shift in generations as a window of opportunity for making certain desirable adjustments (Figure 2). In the third-generation specification, the successor to the BSC has a modified role and a new name: radio network controller (RNC). The key external interfaces of the radio network are

- the air interface to the terminal or handset, designated *Uu*; and
- the interface to the core network, designated *Iu*. The *Iu* interface corresponds to the *A* and *Gb* interfaces in GSM and GPRS.

The internal interfaces of the radio access network are as follows:

- *Iub*—which corresponds to the *Abis* interface in GSM; and
- *Iur*—this is a new interface. It is required for performance in CDMA networks and to make mobility transparent to the core network.

RNC and BSC

At first glance the architecture of the GSM radio access network appears to be quite similar to that of the WCDMA radio access network. Therefore, one might assume that the BSC and RNC have more or less the same role or function. However, closer examination reveals a number of differences, most of which relate to the WCDMA radio access network and its architecture.

BOX A, TERMS AND ABBREVIATIONS

AAL2	ATM adaptation layer 2	NMT	Nordic mobile telephone
AMPS	Advanced mobile phone system	O&M	Operation and maintenance
ATM	Asynchronous transfer mode	OC	Optical carrier—SONET standard
BSC	Base station controller	OC-3	SONET transmission frame format for 155 Mbit/s
BTS	Base transceiver station	PDH	Plesiochronous digital hierarchy
CORBA	Common object request broker architecture	RANOS	Radio access network operation system
CPP	Cello packet platform	RNC	Radio network controller
CPU	Central processing unit	SCB	Switch core board
E1/T1/J1	PDH transmission frame formats for 2 Mbit/s (E1) alt. 1.5 Mbit/s (T1/J1) transmission rates	SDH	Synchronous digital hierarchy
ET	Exchange terminal	SONET	Synchronous optical network
GPB	General-purpose processor board	SPB	Special-purpose processor board
GSM	Global system for mobile communication	STM	Synchronous transport module
HTML	Hypertext markup language	STM-1	SDH transmission frame format for 155 Mbit/s
IP	Internet protocol	SXB	Switch extension board
IT	Information technology	TCP	Transmission control protocol
MPC	Main processor cluster	TEMS	Test Mobile System (product name)
MSC	Mobile switching center	TUB	Timing unit board
MSP	Multiplex section protection (G.783)	WCDMA	Wideband code-division multiple access
NBAP	Node B application protocol		

Soft handover

For reasonable radio performance, WCDMA—a direct-sequence CDMA technology—requires soft handover between cells, base stations and RNCs. This results in radio functions in the RNC, and requires interfaces to the RNC user plane and signaling between RNC control planes. Consequently, for a given call, the RNC has two logically separate operating and serving roles. In its operating role, the RNC manages cells and base stations (Node B). In its serving role, the RNC manages radio connections. Ordinarily, both roles belong to the same RNC, but after handover the operating role may move to another physical RNC, controlling the base station used by the terminal.

Integrated service and future-proof interfaces

Some of the services that have been added to GSM have led to the creation of new interfaces, for example, for positioning services. These services have been integrated into third-generation systems and interfaces. The interfaces have been restructured to limit the risk of having future services affect them. As part of this clean-up, the responsibility for "mobility" in the RNC and the use of radio resources have been clarified, and differences in the interfaces to the packet-switched and circuit-switched domains of the core network have been reduced.

Packet transport network

Many third-generation applications are expected to generate traffic as bursts of TCP/IP packets. To use transmission resources efficiently, packet transmission technology is needed. Initially, while preparing for other technologies (such as pure IP) the ATM/AAL2 transmission technology has been employed. The architecture solutions also enable the use of a dynamic switching infrastructure, which further reduces the cost of maintaining the network. The ongoing IT evolution also influences the architecture. Faster, smaller and cheaper semiconductors allow intelligence to be built into base stations, making them much more self-contained.

Implementation consequences

In terms of implementation, the most striking differences between an RNC and a BSC are as follows. The RNC

- uses packet-centric transmission (ATM);
- makes use of the *Iur* interface;
- consolidates the *Gb* and *A* interface in the *Iu* interface;

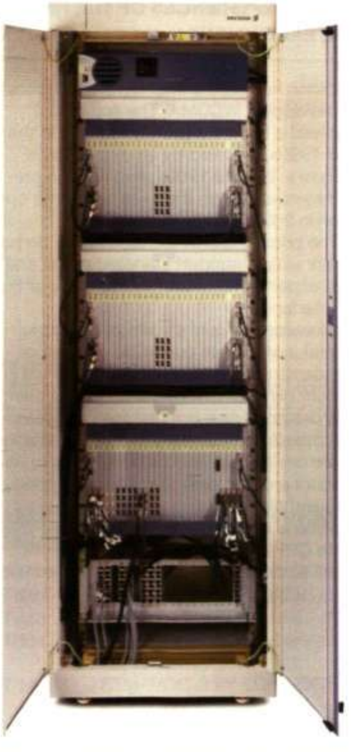


Figure 1
RNC3810, C-configuration.

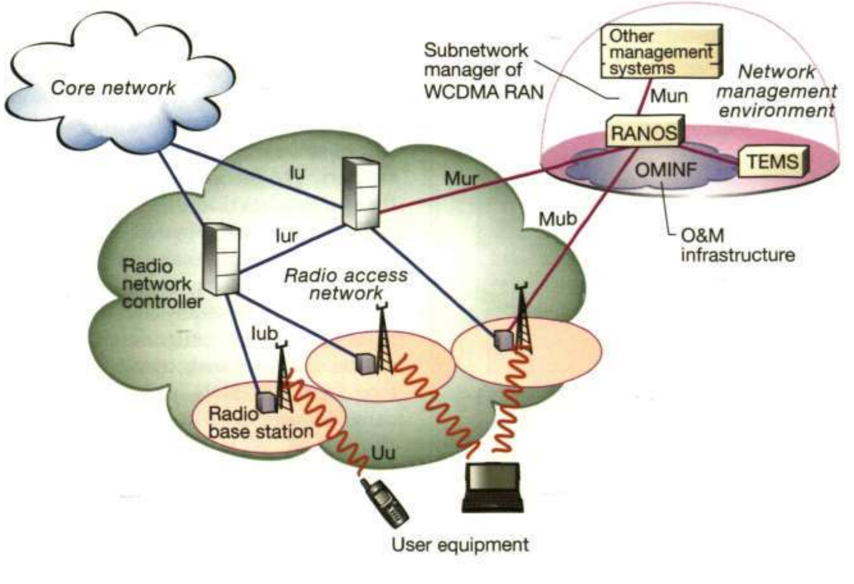


Figure 2
WCDMA radio access network.

BOX B, EXPERIENCES OF THE RNC IN THE EXPERIMENTAL WCDMA SYSTEM

Ericsson has used a step-by-step approach to developing WCDMA. The first step involved a technology demonstrator in the mid-1990s, followed by an experimental system in the late 1990s, and finally a pre-commercial system that was taken into operation in 2001.

The primary aim of the technology demonstrator was to support standardization and to gain expertise regarding the behavior of the air interface.

Likewise, the primary aim of the experimental system was continued support of standardization and to gain competence in system behavior. The experimental system was designed as a pre-commercial system that modeled a real network implementation. Therefore it could also be used to verify new product architectures, development principles and technologies.

Several of the technologies verified related to CPP. The experience gained from these activities was applied to the development of the commercial systems and platform.

- requires considerably more signaling per call; and
- has no transcoder, but instead uses the soft-handover splitting/combination to manipulate the user plane.

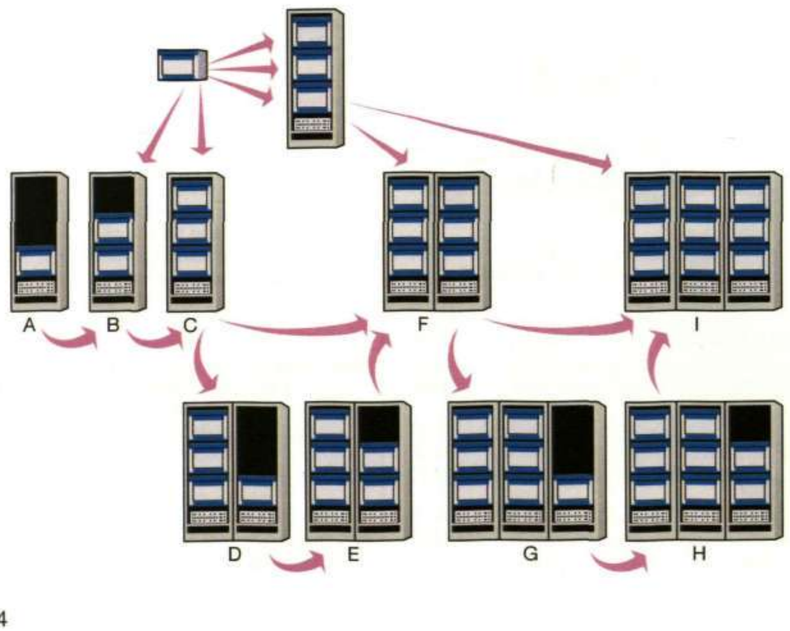
Most current BSC implementations are based on traditional telecommunications circuit switches for wire-line phone services. Major changes would be required to modify these to accommodate new third-generation requirements.

Packet-centric transmission can be used with second-generation BSCs provided an ATM interface is added and the switching matrix is greatly overdimensioned. As an alternative, the circuit switch can be replaced with an ATM switch. This is what Ericsson has been doing in its ENGINE solution.

The difference in switching versus CPU capacity is perhaps an even bigger challenge. The amount of signaling (and thereby required CPU power) per call varies at a ratio of 1:10:100 between wire-line switch, BSC and RNC applications. The main reasons for the variance between the BSC and the RNC are frequent soft handover events and adjustments of the radio channel (channel switching) in response to the packet nature of applications. In GSM there is approximately one handover per call, whereas in a loaded WCDMA network more than 10 handovers and other radio channel changes can be expected per call.

Although the total processing power required for the user plane in the RNC is greater than that of the BSC, nearly 80% of the hardware in the RNC3810 is employed for user-plane-related processing. Therefore, the difference in required processing power has little effect on the hardware architecture of the RNC. More significant is the difference in the nature of processing. In the BSC, transcoders with signal processors account for the majority of processing resources, whereas in the RNC the brunt of processing entails shuffling data using general-purpose processors. Another difference lies in the internal data paths. The BSC works with a strict multiplexing architecture, whereas the RNC requires full switching capabilities. Ericsson has thus developed a completely new node platform for the RNC based on the Cello packet platform (CPP). CPP is also used in other products, such as the RBS3000 and the media gateway.

Figure 3
RNC hardware expansion.



RNC3810

The RNC3810, Ericsson's first RNC for WCDMA, has been optimized for scalability and easy upgrading—even during full-service operation. For example, a small configuration used during roll-out can easily be expanded to larger configurations having a throughput that surpasses that of many large BSCs.

The two central building blocks for the RNC3810 are the main subrack and the extension subrack. The smallest configuration contains only the main subrack. However, up to eight extension subracks can be added to provide capacity for serving some 375,000 subscribers on 500 base stations. The subracks are housed in one, two or three cabinets. The footprint of the largest configuration is less than 0.75 m2 (Figure 3).

The logical and physical separation of transport and transmission interfaces from user plane processing makes it simple to connect the RNC to any core network and any access network configuration. The RNC can be equipped with transmission interfaces for ATM-, SDH- and PDH-switched networks on channelized as well as non-channelized STM-1/OC-3, E3/T3 and E1/T1/J1. Likewise, higher-order transmission and connectivity to IP networks can easily be introduced when the marketplace so requires.

Cello packet platform

The RNC3810 is built almost entirely from products based on CPP—for example, for computing, transmission interfacing, and the ATM switching core.[1] CPP also provides a real-time operating system that features support for multiprocessing hierarchies and extensive solutions for redundancy handling. The following CPP hardware building blocks are used in the RNC:

- The switch core board (SCB) handles ATM switching. It is also the interface to four extension subracks using inter-subrack links (ISL).
- The switch extension board (SXB), which is an extension to the SCB, facilitates the connection of four more extension subracks.
- The general-purpose processor board (GPB) is the computing heart of the RNC (Figure 4). Several GPBs can form a main processor cluster (MPC).
- Special-purpose processor boards (SPB) are used for special, resource-intensive tasks, such as processing user-plane data.
- The timing unit board (TUB) provides the timing reference for the system.
- Exchange terminals—a family of transmission interface boards with standardized internal interfaces to the processing parts of the RNC (Table 1).

Figure 4
General-purpose processor board (GPB).

RNC3810 hardware architecture

RNC3810 has a four-tier processor hierarchy (Figure 5):

- At level one (hub level) three GPBs form the hub MPC. This configuration is duplicated in hot standby set-ups (6 GPBs). The processing tasks—O&M handling (*Mu* interface), signaling (*Iu* and *Iur*), and node control—are allocated to the processors in the multiprocessor cluster by configuration at software build.
- Level two (device level) is an RBS MPC that is physically implemented as a pool of GPBs whose main task is to provide the *Iub* interface.
- Level three (board processor level) constitutes a standardized internal computing interface. Each board contains a board processor regardless of the application. This level is not visible to the end-user of the RNC.
- Level four (user-plane level) consists of a pool of SPBs for intensive user-plane data processing for the dedicated channel, broadcast channel, and packet handling.

The board processor level (level three) can only been seen by the platform level in CPP. The WCDMA radio access network application solely uses levels one, two and four.

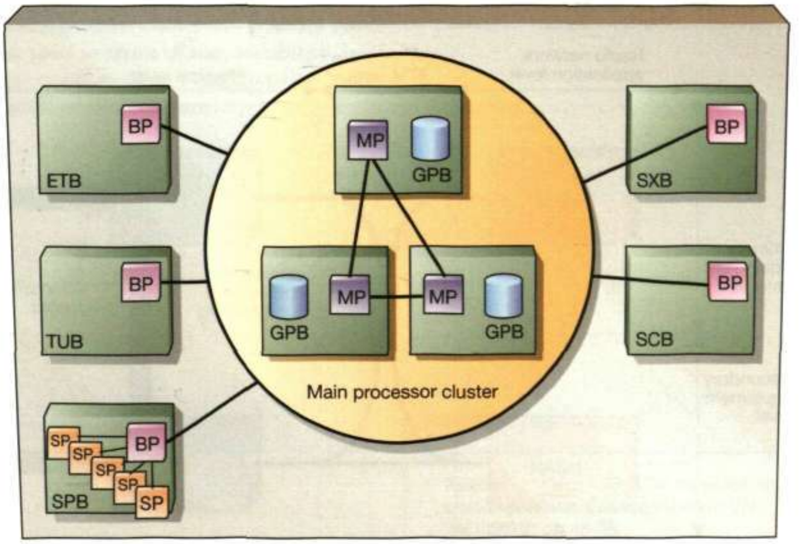
RNC module

The RNC module, which provides another way of looking at the RNC architecture, is a processor cluster made up of level-two and level-four processors by a logical definition in software. For instance, one level-two GPB and five level-four SPBs can be grouped into

TABLE 1, EXAMPLES OF EXCHANGE TERMINALS.

ET-M1	E1/T1/J1 for PDH networks
ET-M3	E3/T3 for PDH networks
ET-M4	STM-1/OC-3 for ATM-switched networks
ET-MC41	STM-1/OC-3 for SDH/SONET-structured E1/T1/J1

Figure 5
Processor hierarchy.



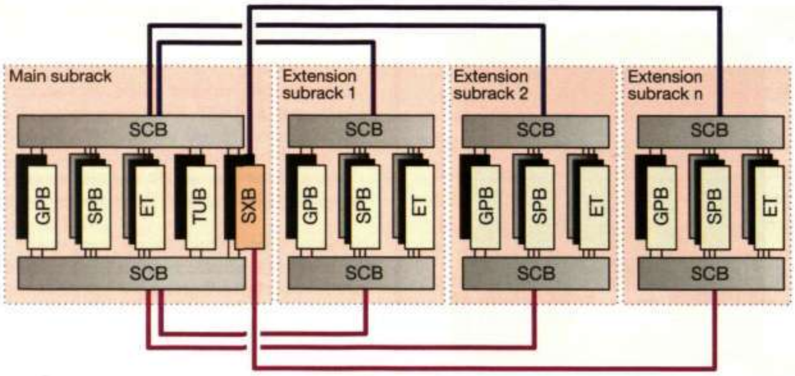


Figure 6
Duplicated interconnections.

a processor cluster—the RNC module. This is actually a small RNC that handles all system interfaces (*Iu*, *Iur* and *Iub*) using hub-level computing resources and the standardized internal transmission interface for external access. More RNC modules means more capacity (near-linear relation). Although the RNC is a logical entity it can be physically mapped to the circuit boards of the extension subrack. This means that the addition of extension subracks also gives a linear growth in capacity.

In-service performance

Due to the large service capabilities required of the RNC, designers have gone to great lengths to ensure that it can operate continuously. One way of obtaining high availability in fault situations is to implement a high degree of redundancy (Figure 6). Therefore, the RNC3810 has no singular units. Besides redundantly configured processors, the RNC3810 has duplicated internal connections, fan units, and power feeding and distribution.

Redundancy is also used for updating hardware. Duplicate units can be forced into standby state, replaced, and then taken back into operation while the executive unit guarantees full operation. Both processors can be updated by toggling the state of an executive unit to standby and the standby unit to executive.

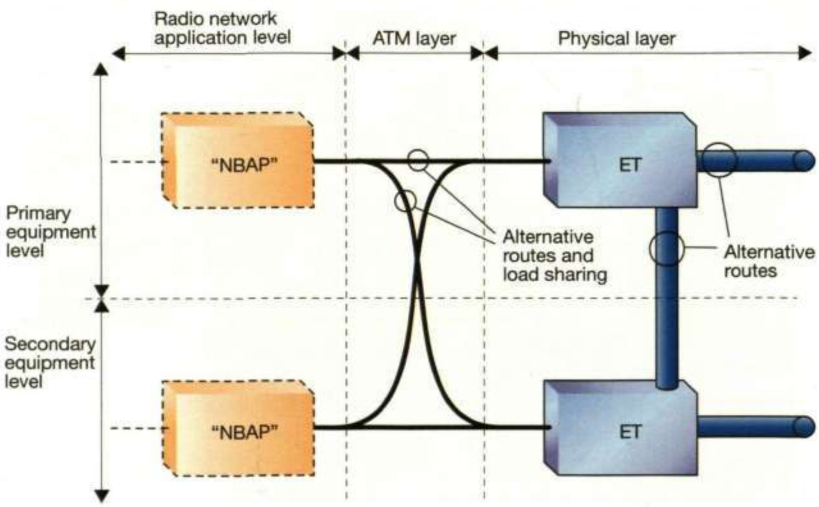
Transmission redundancy

The information transported on STM-1 links corresponds to large service areas. Therefore, continuous connectivity must be ensured. Two concepts guarantee high transmission availability to equipment and transmission links (Figure 7):

- alternative processing; and
- physical link protection.

Alternative processing is set up as duplicated instances of processing parts on the radio network application level. Each part, executing on separate processors, is connected to separate transmission branches using ATM switching functionality. The ATM switching functionality is also used for load sharing by dividing one flow of information onto two physical transmission lines. Physical link protection is implemented as MSP1+1 protection, which is to say that a fatal failure on one transmission link results

Figure 7
Transmission redundancy.



in automatic switchover to an alternative transmission line. MSP1+1 is also used to protect equipment on channelized STM-1 interfaces.

Restart mechanism

When recovering from a fault or from a hardware upgrade, the RNC is restarted using a "staircase" with three steps of escalation (Figure 8).

Step 1, or *program restart*, is the first system action taken to recover from a fatal state in a software application. This restart, which has little or no impact on the total service of the RNC node, is also the basic mechanism for upgrading specific parts of the software system. The upgrade is always made on a standby processor.

Step 2, or *processor restart*, is a hardware-oriented restart of the entire software application on a specific processor. This action is initiated when program restart is insufficient to recover from a fault situation or when major changes have been made in the application during system upgrade. This is also the regular way of taking a replaced hardware unit into operation.

Step 3, or *node restart*, which is an emergency alternative, is solely used in extraordinary fault situations. The entire node is reset and is thus not accessible for service during the initial phases of the restart. This step might also be used during extensive system upgrades to shorten the total time of upgrade but at the expense of service availability.

Software architecture

When developing the experimental WCDMA system, software architects used a high-level, object-oriented design and programming approach to evaluate time to market, quality, and ease of adaptation to future requirements. Because this approach proved to be highly successful it has also been applied to the development of the RNC. To keep complexity and lead times at bay, standard middleware packages have been included to support relational real-time databases, IP, HTML, CORBA, SS7 signaling, and so on.

Great strides have been made to ensure that the structure would be scalable. When defining the programming model, objects that frequently interact have been kept in the same processor. Furthermore, functions that cannot be managed by a single processor in a very large node have been designed to work in a distributed fashion, thereby allowing load to be spread over several processors.

Evolution

Thanks to a modular hardware and software architecture, the product characteristics of CPP and the RNC3810 can easily be modified by changing the proportion of installed boards or board placement. Likewise, the internal building blocks or components can evolve independently of one another.

Conclusion

The RNC3810 represents a milestone in the evolution of cellular switching nodes. The first member of a family of products built on a new architecture and technology platform, it now stands at the commencement of a long and successful evolution.

The RNC 3800 family of products has been designed to meet every possible customer need in terms of size, scalability, in-service performance, functionality and transmission environment.

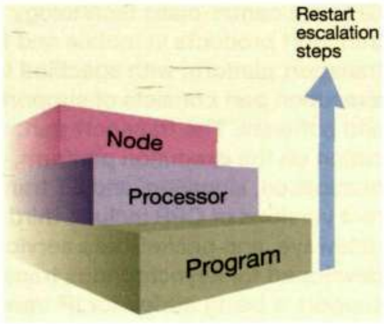


Figure 8
RNC restart escalation staircase.

REFERENCES

1 Reinius, J.: Cello—An ATM transport and control platform. Ericsson Review Vol. 76(1999):2, pp.48-55.

CPP—Cello packet platform

Lars-Örjan Kling, Åke Lindholm, Lars Marklund and Gunnar B. Nilsson

CPP is a carrier-class technology that has been positioned for access and transport products in mobile and fixed networks. It is an execution and transport platform with specified interfaces for application design. The execution part consists of support for the design of application hardware and software. The transport part, which can be seen as an internal application on the execution platform, consists of several protocols for communication, signaling, and ET transmission. Typical applications on current versions of CPP include third-generation nodes—RBSs, RNCs, media gateways, and packet-data service nodes/home agents. CPP was first developed for asynchronous transfer mode ATM and TDM transport. Now, support is being added for IP transport.

The authors describe the technical and customer benefits of adding IP support in CPP, walking the reader through the basic principles for IP services in CPP and the CPP IP architecture, which is very robust and scalable.

Why IP transport?

The world of telecommunications is characterized by change. Yesterday's dominant traffic type—voice—is being superceded by data traffic. In future telecommunications networks, voice will occupy only a small portion of bandwidth. This will put new demands on the networks, which will have to be packet-oriented and at the same time able to handle delay-sensitive traffic, such as voice and video (video conferencing). ATM technology has long been considered the solution to quality of service (QoS) in networks, but several areas of concern have since been identified in large and growing networks. These are

- scalability—the cost of the extra ATM layer makes it difficult for ATM vendors and operators to increase the link speed above 1 Gbit/s;
- administration and maintenance—in large ATM networks, the number of permanent virtual circuits (PVC) for router interconnections increases significantly, and thus the administration and maintenance of these PVCs becomes a major issue when networks need to be upgraded. A similar problem exists in pure IP networks, but the addition of ATM does not simplify matters. In this sense, ATM is an extra layer that greatly increases complexity; and
- cost—for example, the cost of having to support and maintain two kinds of network equipment.

These issues have become a driving force for introducing IP routing into telecommunications networks.

The Internet boom has been accompanied by increased investments in IP technology, and a lot of effort has gone into solving quality-of-service and routing administration issues. Multiprotocol label switching (MPLS), differential services (DiffServ, Box B), multiclass extension (MCE), and header compression have made it possible to replace ATM with pure IP networks. And CPP is well positioned to handle the new environment (as well as the transition to it). DiffServ and MCE will be implemented in the first IP release of CPP, and the architecture is ready to make use of MPLS.

BOX A, TERMS AND ABBREVIATIONS

API	Application program interface
ATM	Asynchronous transfer mode
BGP	Border gateway protocol
CPP	Cello packet platform
DiffServ	Differential services
E1/J1/T1	PDH transmission frame formats for 2 Mbit/s (E1) or 1.5 Mbit/s (J1/T1) transmission rates
ET	Exchange terminal
FIB	Forwarding information base
HA	Home agent
IP	Internet protocol
IPv4, IPv6	IP version 4, IP version 6
MCE	Multiclass extension
MGW	Media gateway
MPLS	Multiprotocol label switching
O&M	Operation and maintenance
OC3	Optical carrier 3 (155 Mbit/s)
OSPF	Open shortest path first
PBA	Printed board assembly
PDSN	Packet-data service node
PPP	Point-to-point protocol
PVC	Permanent virtual circuit
QoS	Quality of service
RBS	Radio base station
RIP	Routing information protocol
RNC	Radio network controller
RSVP-TE	Resource reservation protocol – traffic engineering
SCTP	Stream control transmission protocol
SIGTRAN	Signaling transport
SS7	Signaling system no. 7
STM-1	SDH transmission frame format for 155 Mbit/s
TCP	Transmission control protocol
TDM	Time-division multiplexing
UDP	User datagram protocol

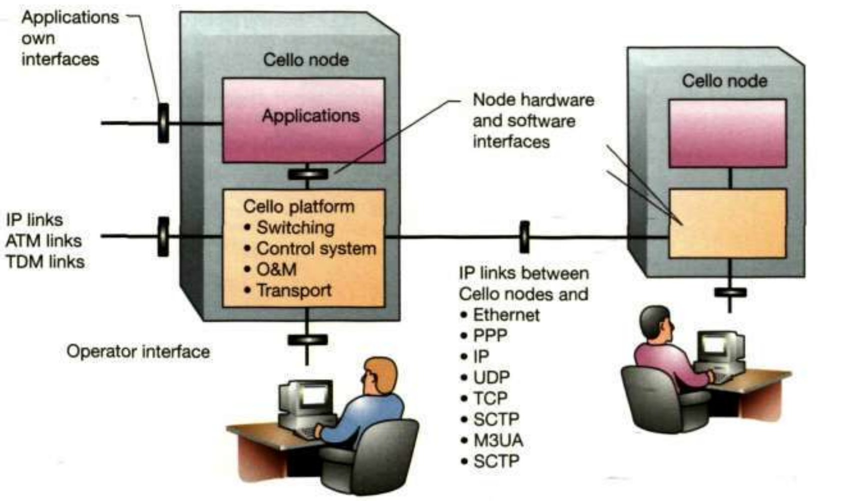


Figure 1
A network view of the Cello packet platform (CPP) including transport protocols.

Who are the customers?

The introduction of IP in telecommunications networks will commence in the core network in response to demands for large, high-speed networks. Datacom vendors are currently trying to grab as many market shares as they can. At present, these vendors can solely offer pure IP router products. Therefore, network operators can either opt to integrate telecommunications equipment and IP routers themselves or they can buy complete site solutions from a telecommunications vendor. Over time, the IP migration will extend further and further from the core network into the access network.

Ericsson knows that operators want vendors to come up with solutions and migration stories that are cost-effective and easy to maintain (Figure 2). The introduction of IP in

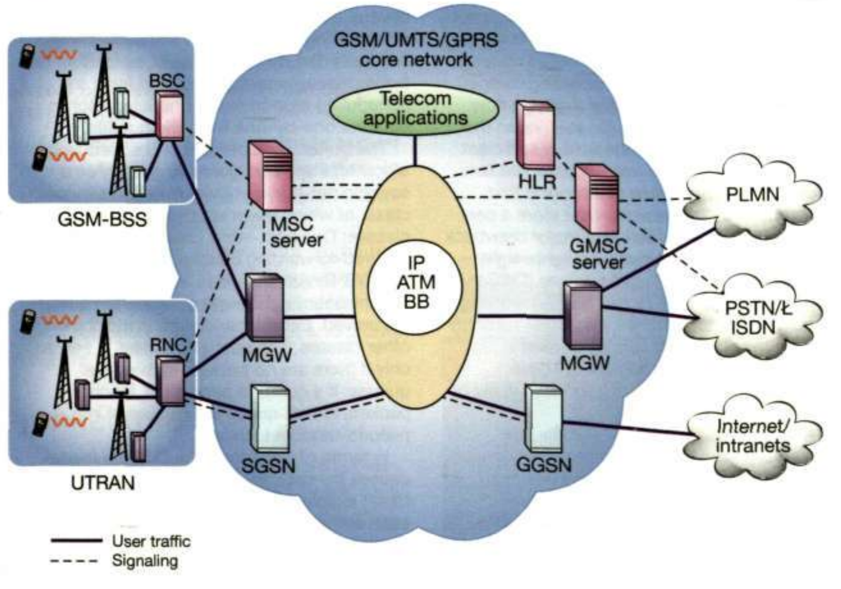


Figure 2
A telecommunications network with different types of backbone and access networks.

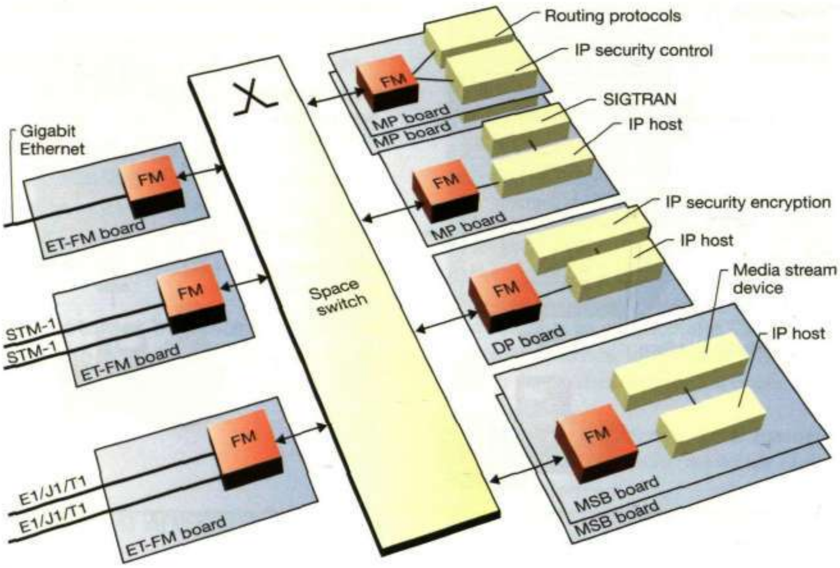


Figure 3
Embedded and distributed IP router.

BOX B, DIFFERENTIATED SERVICES

Traditional routers serve packets on a first-come-first-served basis. Since no differentiation is made with respect to the actual demands put on the delivery times of packet flows, this is often called best-effort service. In a best-effort network, packet flows that carry delay-sensitive voice receive the same service as packet flows that carry, say, a file transfer.

Different approaches have been proposed to solve this shortcoming. According to one approach, called the *IntServe* model, a flow must reserve resources in the routers on the path before packets can be sent. If sufficient network resources are not available, the flow is denied. An implication of IntServ is that every router in the network must store a per-flow state. This was seen as a major drawback and led to the introduction of a lightweight model called differentiated services (DiffServ). The DiffServ model is based on the following basic principles:

- A limited number of service classes is defined for specific purposes. These classes differ in terms of maximum delay or maximum drop probability.
- At the edge of a network, packets are marked in the header to reflect the service class to which the packet flow belongs.
- The routers in the network serve the packets according to its service class.

In principle, packet flows that belong to the same service class are merged into a common aggregated flow. The routers see these aggregated flows, but not individual packet flows.

A packet enters a classifier, where the service class mark is inspected. Depending on the mark, the packet is sent to one of several queues via an *enqueuer*. Depending on the filling in the queue, the packet might be discarded by the enqueuer according to a buffer management algorithm. Using a scheduling algorithm, a scheduler fetches packets one-by-one from the queues.

The buffer management and scheduling algorithms are configured to guarantee a specific per-hop-behavior for each service class, of which there are three standard classes: One best-effort class (BE) and two assured-forwarding classes (AF). Each assured-forwarding class is guaranteed a certain minimum bandwidth at which the queues are served. Excess bandwidth is distributed to other classes. The best-effort class is served only if there are no packets in the other queues. If a queue is almost full, incoming packets to that queue are discarded in a pseudo-random fashion.

In terms of resources, the DiffServ model causes the network to behave as if a few different logical networks were separated from each other. For example, one logical network could be used for voice traffic and another for file transfers.

telecommunications networks adds yet another transport protocol. Today operators are struggling with the migration from TDM to ATM. Later, when operators introduce IP, they will have to consider even more complex networks. Ericsson can offer telecommunications nodes with built-in transport capabilities for TDM, ATM and IP. Operators can thus capitalize on their installed base, especially in mobile networks. Likewise, Ericsson knows that network operators want a compact and cost-effective solution that has a consistent network-management interface. Ericsson will thus make pure IP routers obsolete in operator networks. CPP is the ideal choice of carrier-class technology for all telecommunications nodes in the access network, including the edge nodes to the core backbone.

Basic principles

Six basic principles for the IP services in CPP add value:

- embedded and distributed IP router by means of routing protocols in the main processor cluster and distributed forwarding on all device boards (Figure 3);
- fully distributed forwarding of IPv4 or IPv6 modules can be implemented in hardware or software. A hardware implementation can handle wire-speed transport (Figure 4);
- internal IP hosts can be located on a single printed board assembly (PBA) and connected to the IP router via the local forwarding module on the PBA (Figure 5). The IP host provides the application program interface (API);
- the internal hosts are IP end-systems that are visible and addressable from the external network and from within the node (Figure 5);
- a CPP node can house multiple virtual routers. In this case, each IP interface in the node is configured to belong to one specific virtual router; and
- CPP has built-in support for signaling system no. 7 (SS7) signaling gateway functionality by means of a complete SS7 stack for SIGTRAN, IP, ATM and TDM (Figure 6).

Architecture

The CPP IP architecture is composed of several subsystems that interwork with each other through well-defined interfaces (Figure 7). The IP forwarding subsystem provides fully distributed forwarding of IPv4

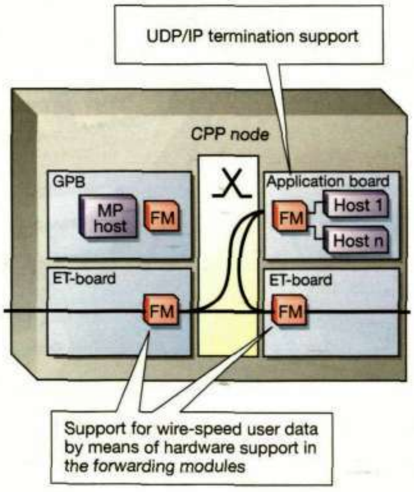


Figure 4
Fully distributed IP forwarding.

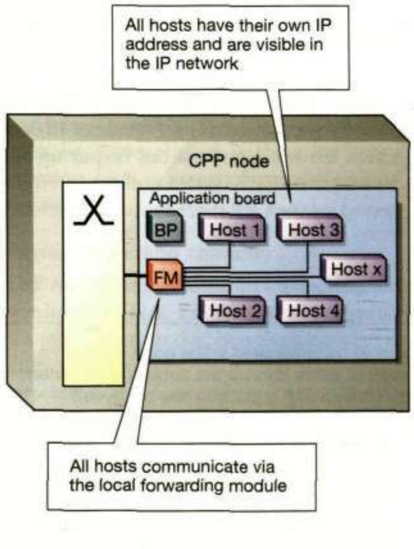


Figure 5
Multiple internal IP hosts.

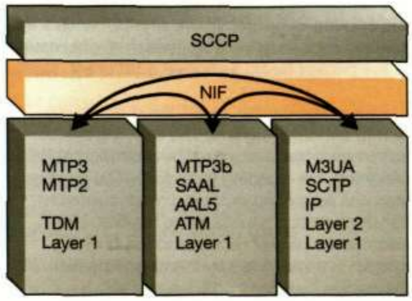


Figure 6
Built-in support for signaling gateway.

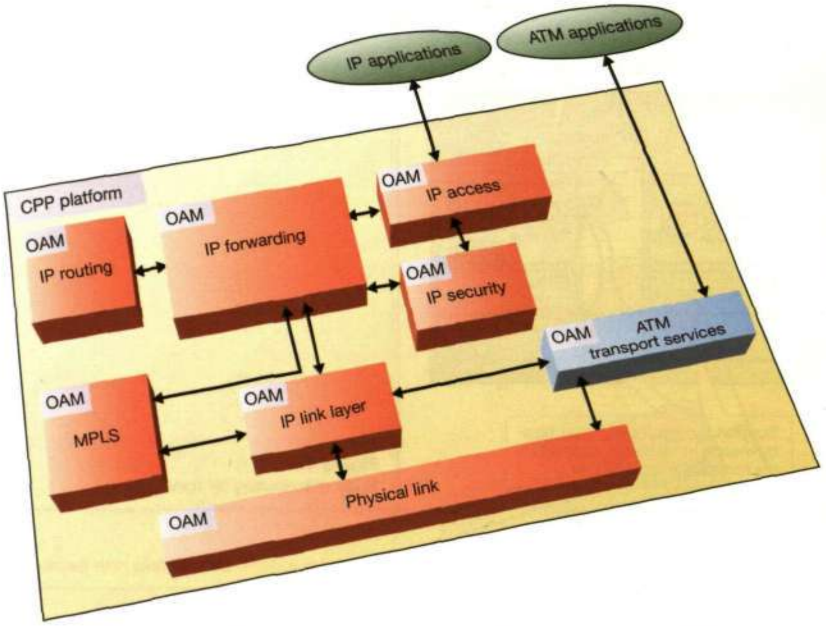


Figure 7
The CPP IP architecture.

and IPv6 packets. Forwarding modules (Box C) implemented in software or in dedicated hardware circuits can be put on any board in the CPP system and are interconnected by the CPP space switch. Hardware-based forwarding modules can achieve wire-speed forwarding. The IP forwarding subsystem also includes packet classification and filtering, including DiffServ queuing.

To forward an IP packet, the forwarding modules need a forwarding table that is calculated in the IP routing subsystem using static (configured) routes and dynamic routing protocols, such as the open shortest path first (OSPF), routing information protocol (RIP), and border gateway protocol (BGP). The routing protocol handlers monitor the network topology, and the routing table manager calculates a forwarding table at start-up and when the network topology changes. Forwarding information from the IP routing subsystem (Box D) is communicated to all forwarding modules in the CPP node using the dedicated forwarding information base (FIB) interface.

Telecommunications applications that use CPP IP transport services need to terminate and originate large amounts of IP traffic that comes from or is sent to the IP network. The IP access subsystem enables applications to use IP hosts in CPP. A user can access multiple IP hosts on every processor board in the system: each host is identified by a unique IP address. The hosts can handle IPv4 and IPv6, user datagram protocol (UDP), transmission control protocol (TCP), and stream control transmission protocol (SCTP) termination. If necessary, the hosts can be made robust by means of the moveable host concept (see also section on Robustness).

An automatic configuration mechanism simplifies host configuration. The IP access subsystem uses the IP forwarding subsystem to send and receive packets. For low-speed applications, CPP provides a distributed host mechanism whereby one IP host with one IP address can be distributed and used on multiple processor boards.

IP transport is also used for highly confidential traffic, such as control signaling and operation and maintenance (O&M) traffic. To guarantee the integrity of this traffic, the IP security subsystem provides tunnel- and transport-mode encryption and decryption of IP packets. The IP security encryption/decryption engines can be distributed on multiple processors or dedicated hardware circuits to yield greater capacity.

The CPP architecture has been prepared for the introduction of MPLS—CPP can serve as a label edge router or label switch router. However, modifications will need to be made to the forwarding modules, and ad-

BOX C, FORWARDING MODULES

Forwarding modules implement the functionality needed to forward or terminate an IP packet. They also provide the functionality for resolving dependencies on connected routers (the data part) and hosts (the control part).

Figure 8 shows the data part of a wire-speed forwarding module. Most of the functionality is implemented in dedicated hardware or network processors. In addition, an exception handler executes on a standard processor. The exception handler provides the functionality needed for dealing with some infrequent packet types. If, during any step in the forwarding process, an exception is identified, the packet is handed over to the exception handler.

Packets enter the forwarding module from an incoming link. Each packet is classified according to the type of service it requires. This classification might involve metering actual load. The type-of-service field in the packet might be changed as a result of the classification. The classification also serves as a sort of firewall, and might result in having the packet dropped.

The packet is next analyzed to find out if it should be terminated in the node. If so, a point of termination in the node is determined and the packet is put into one of several queues to the switch. Otherwise, the packet is further analyzed to determine

- the best next hop in the network; and
- various related parameters. One parameter contains the outgoing link and the point in the node where the forwarding module responsible for this link is located.

The packet is then put into one of the switch queues. After having passed the switch, the packet arrives at the forwarding module of the appointed outgoing link where it enters a second classification step, which might result in the packet being assigned a different class.

Finally, the packet enters the queuing system to the outgoing link. The queuing system has several queues served by a configurable, weighted, fair-queuing scheduler. The system also has advanced dropping mechanisms to handle overload situations on the link. The treatment a packet receives in the queuing system is determined by the class to which the packet has been assigned.

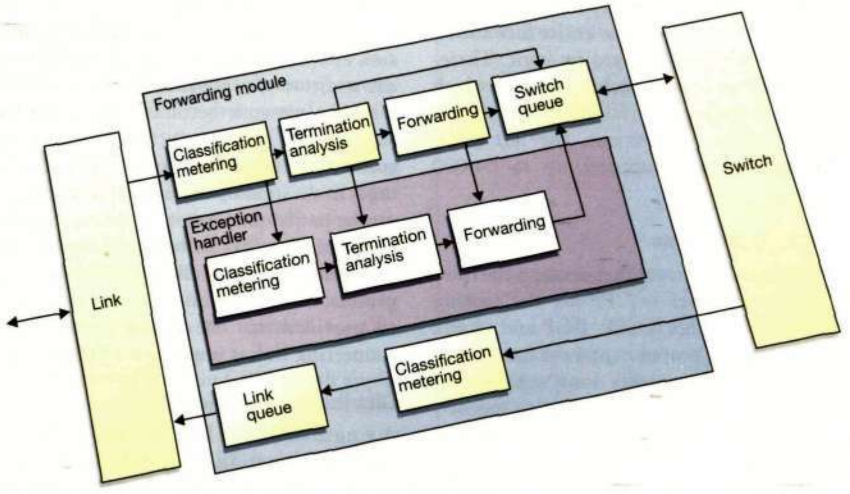


Figure 8
Internal architecture of the forwarding module.

ditional signaling protocols will be introduced, including the

- resource reservation protocol – traffic engineering (RSVP-TE); and
- CR-LDP.

Likewise, interworking will be introduced between IP routing and MPLS.

Together with the physical layer subsystem, the IP link layer subsystem provides access to the external network. For example, the IP link layer subsystem supports 10/100 Mbit Ethernet, Gigabit Ethernet, point-to-point protocol (PPP), IP over ATM, and frame relay. Likewise, the physical layer supports STM-1/OC3, E1/J1/T1 and Ethernet. The IP link layer functionality is connected to the IP forwarding subsystem using the generic link interface.

Scalability

CPP is uniquely scalable. It can be used in small applications (such as a small RBS) as well as large RNC and media gateway nodes. Indeed, virtually every aspect of scalability is covered by CPP: high-end and low-end nodes, payload capacity, processing capacity, number of routes, route updating capacity, number of physical links, link capacity, and cost.

In general, the IP functionality of CPP is composed of central functions and local functions. A central function solely exists in one instance for each virtual IP router, whereas a local function might exist in several instances. With few exceptions, the scaling of local functions is quite straightforward. By contrast, the scaling of a central function is an intricate matter since central functions have a single, consistent information base. Distribution is not precluded but must be accomplished without incurring excessive costs in terms of memory and capacity.

Payload handling capacity

CPP forwarding is based on interacting forwarding modules. Each forwarding module is a local function that handles a number of links and interfaces. Each processing entity (processor) and printed board assembly has its own forwarding and link module, which means that forwarding and link capacity are not adversely affected when new hardware is added. The only potential bottleneck is the intercommunication capacity of the forwarding module, which is based on the CPP space switch. At present, the capacity within the subrack is 16 Gbit/s, which is sufficient to support some 400,000 IP-borne voice calls. Moreover, multiple subracks can be interconnected to form very large nodes that constitute a single coherent system. As with forwarding, IP access mainly consists of local functions and scales smoothly when new hardware is added.

Forwarding information base

The number of routes to be handled directly affects the size of the forwarding information base (FIB). A very large FIB makes it unfeasible to house the entire forwarding table in fast path-forwarding logic. Therefore, CPP provides a caching scheme which forwards packets that fall outside the scope of the cache. At present, the fast path-forwarding path supports up to 64,000 routes.

Routing protocols

The central functions that demand the most processing power in CPP are the routing protocol handlers (OSPF, BGP and so on). A shortage of processing power for routing results in unacceptably long convergence time—that is, the time it takes to regenerate a routing table when the topology has changed.

Certain measures affecting network layer configuration can significantly reduce the processing load on the individual routers. These measures are hierarchical network topology, appropriate address allocation, route summarization, and area subdivision. CPP supports these measures as well as scalability at the node level.

Allocation of routing protocol handlers

Different routing protocol handlers (RPH) can be allocated to different processors. Every RPH is mastered by a single routing table manager. However, other scaling solutions are necessary when a single RPH requires more capacity than can be provided by one processor.

Adding a virtual router

Where the network is concerned, the addition of a virtual router is no different from the addition of a physical router. In either case, the network becomes more complex and the RPHs must work harder to keep track of the network topology. The advantages in the local node are that the interfaces can be partitioned between virtual routers, and functions that belong to different virtual routers can be allocated to different processors. In addition, costs can be avoided provided that the virtual-router interconnecting link is implemented efficiently.

Distributing a single RPH

A single RPH can be distributed to some extent by allocating certain sub-functions to different processors. The exact allocation of sub-functions varies according to the types of routing protocol in use. For OSPF, the shortest path calculation must be executed in one processor, whereas the handling of interfaces can be distributed.

Low-end scalability

For CPP to meet the requirements of radio base stations it must support low-end scalability. As a consequence, the CPP IP architecture has been designed to accommodate IP functionality on a single PBA and to eliminate the switch. Another feature is the FIB caching mechanism mentioned above. These features keep the cost of fast forwarding-path logic to a minimum even in relatively complex networks with multiple routes.

Robustness

The fully distributed IP-forwarding and IP-termination mechanisms in CPP are very robust. The failure of a board solely affects the forwarding and termination of IP packets on that board. All other boards continue forwarding and terminating as if nothing had happened, except that packets are not forwarded to the failed board. Should a network interface board or link fail, the routing protocols initiate link protection switching or automatic rerouting.

The robustness principles applied in CPP make it possible to define reliable programs that execute in the main processor cluster. These programs can be run on a standby processor should the main processor fail. The CPP IP functionality employs this fail-over concept for all central functions, such as the routing protocols, the central parts of IP ac-

BOX D, IP ROUTING

The Internet is made up of several autonomous systems (AS), each of which is operated by an Internet service provider (ISP) or an organization (Figure 9). In telecommunications, each IP-based radio access network or core network can be seen as an autonomous system. Interior gateway routing protocols, such as RIP, OSPF or IS-IS, are used inside autonomous systems to automatically create forwarding tables to be used by the routers.

OSPF is one of the most popular interior gateway protocols. This link state routing protocol discovers its neighboring routers and learns their network addresses by sending out 'Hello' packets on all its interfaces. The information it receives is stored in a topology database. OSPF then sends out a link state advertisement (LSA) packet indicating the status of its own links. The LSAs are 'flooded' by the other routers in order to reach every router in the AS. Based on this input, OSPF calculates the shortest path to all routers and builds an optimal forwarding table. Whenever a change in network topology takes place—that is, if a link or a router goes down—new LSAs are flooded through the autonomous system. OSPF in each router updates the topology database and recalculates the forwarding table.

Some autonomous systems can be very large and difficult to manage. Likewise, the LSA flooding can generate a heavy load in a large autonomous system. To reduce load and simplify management, OSPF allows autonomous systems to be divided into OSPF areas. Routers connected to one area need only have detailed knowledge about the topology of that area. All areas in an autonomous system are interconnected by the backbone area. The routers connected to more than one area are called area border routers.

cess, IP security, IP forwarding, and all O&M functions. Should the processor on which the routing protocols execute fail, then the routing protocols are simply transferred to the standby processor. In the interim, while the protocols are being transferred and a new forwarding table is being built, the forwarding of packets on all other boards continues uninterrupted using the most recent forwarding table.

In coming releases of CPP, the standby routing protocol will operate in a listening mode and maintain an updated standby routing table.

The internal hosts in CPP, which applications use for terminating and originating IP traffic, can be distributed on any processor board in a CPP node. Each host is identified by a unique IP address. Should a board fail, CPP automatically transfers the host (and IP address) to another board. At the same time, the forwarding table is recalculated and traffic destined for the host is forwarded to the new location.

If an application requires it to do so, CPP can configure a robust Ethernet IP interface that uses a primary and secondary physical port. Should the primary port fail, the secondary port is activated using the IP address that the primary port had.

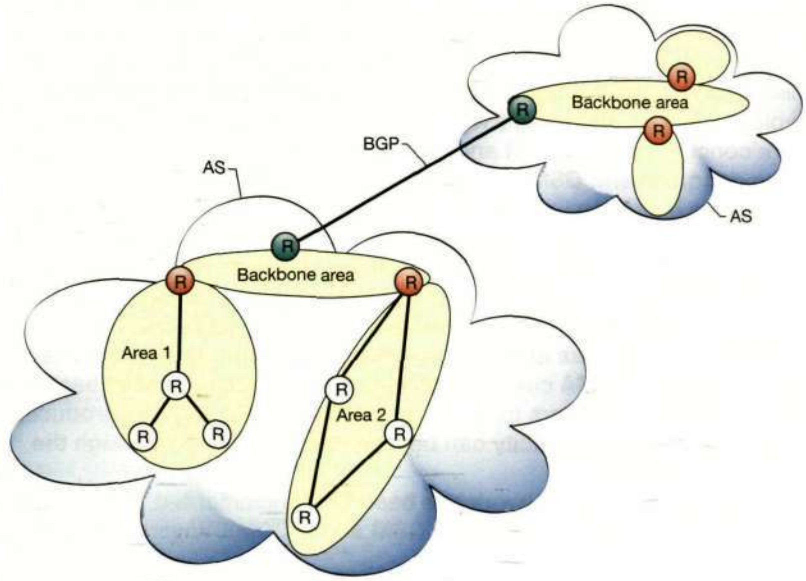


Figure 9
Example of section of the Internet consisting of two autonomous systems (AS), where one AS is divided into three OSPF areas. The border gateway protocol (BGP) is used to convey routing information between the two autonomous systems.

Conclusion

Ericsson knows that operators want vendors to offer solutions and migration stories that are cost-effective and easy to maintain. Consequently, Ericsson offers telecommunications nodes with built-in transport capabilities for TDM, ATM and now, IP. Ericsson also knows that network operators want a compact and cost-effective solution that has a consistent network-management interface. Accordingly, Ericsson will make pure IP routers obsolete in operator networks.

The CPP IP architecture is composed of several subsystems that interwork with each other through well-defined interfaces:

- The IP forwarding subsystem provides fully distributed forwarding of IPv4 and IPv6 packets.
- The IP routing subsystem calculates a forwarding table, which forwarding modules use to forward IP packets.
- The IP access subsystem enables applications to use IP hosts in CPP.
- An automatic configuration mechanism simplifies host configuration.
- The IP security subsystem guarantees the integrity of sensitive traffic.

The CPP architecture has been prepared for the introduction of MPLS—CPP can serve as a label edge router or label switch router. Interworking will be introduced between IP routing and MPLS. Together with the physical layer subsystem, the IP link layer subsystem provides access to the external network.

CPP is uniquely scalable. It can be used in small applications and large RNC and media gateway nodes. Virtually every aspect of scalability is covered by CPP: high-end and low-end nodes, pay-load capacity, processing capacity, number of routes, route updating capacity, number of physical links, link capacity, and cost.

The fully distributed IP-forwarding and IP-termination mechanisms in CPP are very robust. The robustness principles applied in CPP make it possible to define reliable programs that execute in the main processor cluster.

Ericsson seamless network

Roland Heickerö, Stefan Jelvin and Bodil Josefsson

The Ericsson seamless network is a concept that describes the path for GSM operators who are evolving their networks toward a third-generation environment in which GSM and WCDMA are regarded as one network. In this context—and throughout this article—GSM means all accesses deployed in the GSM band (GSM, GPRS and EDGE). The seamless network concept protects past and future investments in GSM by reusing network equipment: GSM is used for second- and third-generation services since it evolves with EDGE technology, and the existing core network evolves into a layered architecture that supports GSM and WCDMA.

Perhaps the most important driver of the seamless network concept is the fact that end-users will be able to have seamless third-generation services from GSM and WCDMA coverage. By allowing a balanced roll-out of WCDMA, the seamless network provides for flexible capital investment—after an initial WCDMA coverage phase, a parallel expansion of each technology enables operators to invest as needed. EDGE can be introduced rapidly, nationwide; capacity can be increased as needed through the addition of WCDMA.

When the seamless network has been fully implemented, it will make very efficient use of the combined GSM and WCDMA spectrum, treating the two as one.

Background

In 2000, to support the seamless GSM and WCDMA network, all subsequent standardization of GSM was transferred from ETSI to the 3GPP, to ensure a smooth harmonization of GSM and WCDMA. The outcome supports an evolved GSM core network with WCDMA and GSM/EDGE radio access.

There are currently more than 600 million GSM subscribers spread throughout 400 networks in 170 countries. Huge investments have been made to build these networks and it will be crucial for operators to continue to capitalize on them. Most operators in Europe have 20-30 MHz of GSM and 10-15 MHz of UMTS spectrum, and will need to find a way of optimizing the entire spectrum.

The seamless network

The Ericsson seamless network concept describes the path for GSM operators who are evolving their networks toward a third-generation environment in which GSM and WCDMA are regarded as one network. (Note: There is a similar evolution for CDMA, but that is outside the scope of this article.)

The transition from GSM to a third-generation environment entails adding more functionality and more value to the current GSM network and business model. This transition is not a series of revolutions but a smooth evolution whose parts add value to the whole (Figure 1).

The evolution begins with an upgrade of the GSM network with packet-data capabilities by adding GPRS, which introduces end-users to the "always-connected" experience.

The next step is to introduce third-generation services in existing and new spectrum. GPRS is enhanced with EDGE (Box B) to provide third-generation services in the GSM band; WCDMA gives third-generation services in new spectrum.

Network structure

To keep the discussion of a seamless network simple, we have divided the network into three areas (Figure 2):

- the GSM radio access network area;
- the WCDMA radio access network area; and
- a common area.

While GSM and WCDMA will not share a common radio access network any time in the near future, radio control features will allow them to work together as a common resource.

The common area in the seamless network consists of all the areas that are or will be shared between the GSM and WCDMA networks. This includes the core network with its GPRS packet backbone network, transmission, sites, handsets, the service network, customer administration system (CAS), and network management system.

Radio network integration

With the GSM/GPRS/EDGE and WCDMA network we have two third-

BOX A, TERMS AND ABBREVIATIONS

3GPP	Third-generation Partnership Project
AAL1	ATM adaptation layer 1
ATC	Adaptive traffic control
ATM	Asynchronous transfer mode
BSC	Base station controller
CAPEX	Capital expenditures
CAS	Customer administration system
CDMA	Code-division multiple access
EDGE	Enhanced data rates for global evolution
ETSI	European Telecommunications Standards Institute
FTP	File transfer protocol
GERAN	GSM/EDGE radio access network
GPRS	General packet radio service
GSM	Global system for mobile communication
GSN	GPRS support node
HLR	Home location register
HSCSD	High-speed circuit-switched data
HSS	Home subscriber server
IP	Internet protocol
OPEX	Operating expenditures
QoS	Quality of service
RAB	Radio access bearer
RBS	Radio base station
RNC	Radio network controller
SCS	Self-configuring system
STM	Synchronous transfer mode
TSG	Technical specifications group
UMTS	Universal mobile telecommunications system
UTRAN	UMTS terrestrial radio access network
WCDMA	Wideband CDMA

Figure 1
Evolution paths of second-generation standards.

generation-capable radio accesses in two different spectrums. To optimize efficiency and capacity in the combined spectrum, we must be able to manage the two resources as one, to create trunking efficiencies from inter-system handover, load sharing, and traffic diversion. Compared to separate operation of accesses, a combined network can yield trunking gains of up to 50%.

Table 1 shows a sample of the theoretical spectrum efficiency gains that can be achieved by combining GSM and WCDMA. (Note: Assuming 100% penetration of dual-mode handsets.)

In practice, the gains will be a result of the mixture of traffic (voice and data) and penetration of dual-mode handsets. Gains in efficiency will also increase with the requirement of high-bandwidth services.

To create adaptive traffic control (ATC), more intelligent communication is introduced between EDGE/GSM and WCDMA using an open interface between the base station controller (BSC) of GSM and the radio network controller (RNC) of WCDMA. In other words, Ericsson's solution for adaptive traffic control is a software function.

To take full advantage of each spectrum, the ATC functionality is complemented with self-configuring systems (SCS), which automate the settings of the most frequently changed cell and neighbor parameters for traffic control in the systems, using real-time performance monitoring and configuration data from the network.[1] This data includes real-time statistics and recordings from all active mobile terminals in the network on both the uplink and downlink.

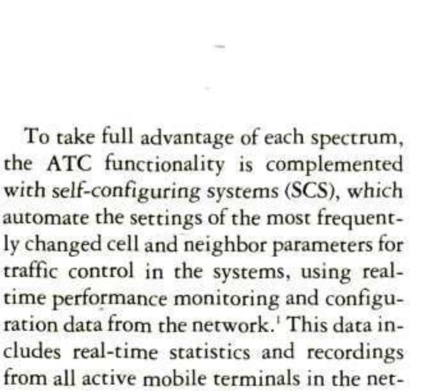


Figure 2
The Ericsson seamless network.

BOX B, EDGE

EDGE is a standardized set of improvements to the GSM radio interface that brings higher data rates and increased spectral efficiency for data services. With EDGE, the operator can serve three times as many subscribers than with GPRS, or provide triple the data rate for one end-user. EDGE provides the same type of third-generation services as WCDMA, but at lower data-transfer rates. Implementing EDGE is fast and cost-effective. EDGE uses the same channel structure, frequency planning, protocols and coverage as present-day GSM. Operators can thus derive more from the same physical resources.

Because the GSM frequency bands are a substantial part of an operator's total spectrum assets, it will become increasingly important to use GSM spectrum for third-generation services. The decision is not to choose between WCDMA and EDGE, but rather how to get the most from each.

Common infrastructure

Core network

The move to a third-generation environment in the core network results in a horizontally layered network that separates payload (voice and data), transport, session control, and applications or services into three distinct layers (networks) with open interfaces. This makes it possible to develop and expand the layers independently of one another. It also allows for the unification of transport technologies, such as IP, which brings telecommunications and data networks together. The common core network is actually an evolved GSM core network.

RBS sites

Since radio base station sites represent a substantial investment in the mobile network, operators have much to gain from sharing sites and site infrastructure.

For WCDMA, Ericsson has developed the RBS3000 series of base stations. An important input to the design of the RBS3000 was co-existence with RBS2000. WCDMA capability is introduced by adding new RBS3000 cabinets to existing sites and sharing transmission and antenna systems with GSM cabinets.

Ericsson has a track record of delivering superior RBS footprint. This also holds true for the two-cabinet solution for combined GSM and WCDMA sites. The built-in power supply and high power efficiency further reduce floor space requirements—that is, the installation does not require a separate power supply rack; similarly, it requires only a minimum of battery backup.

A common building practice is used for the GSM and WCDMA RBSs. This simplifies the installation and commissioning of WCDMA RBSs at existing GSM RBS sites. Therefore, the footprint is identical, as are fixing points and appearance. Furthermore, the same toolbox can be used for commissioning and integration, which reduces investments in equipment and the need for training personnel. Because the RBS cabinets can be mounted back-to-back or back-to-wall, operators can make the most of available space, especially in compact sites.

Transmission

Transmission in the radio access network will evolve as traffic grows and more bandwidth-hungry data services are introduced. The same transmission links can be shared between GSM and WCDMA RBSs, resulting in lower transmission costs.

To use available transmission resources in the most efficient way, Ericsson's RBS solutions support the use of fractional E1/T1, where fractions of the same physical E1/T1 can be used for STM, ATM and IP-based GSM and WCDMA base stations (Figure 3).

Another option is circuit emulation of GSM traffic using AAL1. This method of reducing transmission costs and enabling rapid network deployment is especially interesting for operators who can connect to an ATM network (Figure 4).

TABLE 1, GAIN IN SPECTRUM EFFICIENCY IN COMBINED GSM/WCDMA RADIO NETWORK (5 MHZ GSM/GPRS/EDGE AND 5 MHZ WCDMA).

	WCDMA capacity per cell (5 MHz)	GSM/GPRS/EDGE capacity per cell (8TRX) (5 MHz and 1/3 reuse)	WCDMA and GSM/GPRS/EDGE capacity separately	WCDMA and GSM/GPRS/EDGE capacity as one group	Spectrum efficiency (trunking) gains
Voice	49.7 Erl (60ch)	49.7 Erl (60 ch)	99.4 Erl	107.4 Erl	9%
UDI64	5.1 Erl (10 ch)	6.6 Erl (12 ch, ~4 slots per Erlang on average)	11.7	14.9 Erl	27%
64 kbit/s Web, e-mail	4.4 Erl (10 ch)	5.6 Erl (12 ch)	10.0 Erl	11.5 Erl	15%
UDI44	1.6 Erl (5 ch)	2.3 Erl (6 ch, ~8)	3.9 Erl	5.8 Erl	48%
144 kbit/s Web, e-mail	1.1 Erl (5 ch)	1.6 Erl (6 ch)	2.7 Erl	3.8 Erl	36%

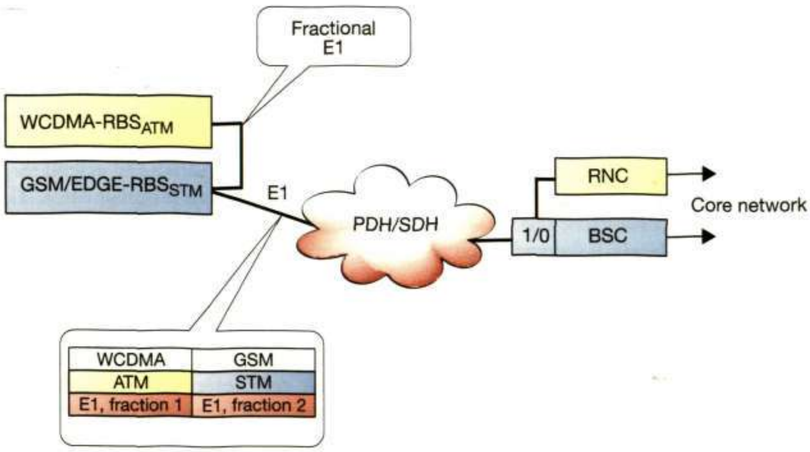


Figure 3
WCDMA on fractional E1.

Network management

Network management will evolve to support the "one network" evolution, and existing support systems will be integrated to give a unified environment for operation and maintenance. This will ease operation and is a prerequisite for optimizing adaptive traffic control.

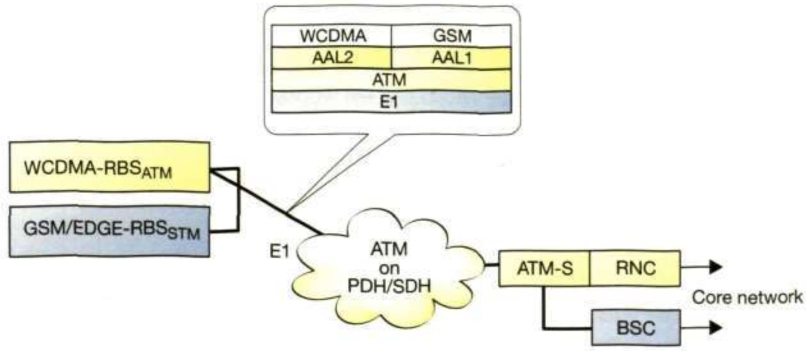
Handsets

Multimode GSM/GPRS/EDGE+WCDMA handsets are a prerequisite for the seamless network. Vendors have announced that they will begin supplying handsets in 2002—GSM/GPRS/WCDMA handsets for the European markets, and GSM/GPRS/EDGE handsets for the Americas. These two platforms, which are similar, feature color displays, cameras, joy sticks, enhanced keyboards, flash memory, and improved battery life. The next step, depending on market demand, will be to integrate the two platforms, creating a GSM/GPRS/EDGE+WCDMA platform.

Operator scenarios

Operators who already have a GSM network with nationwide coverage and who have already introduced GPRS want to use their investments in GSM when evolving it to the next generation of networks based on WCDMA. When rolling out WCDMA alongside GSM, operators face numerous challenges, such as regulatory requirements, cost optimization, and keeping the two access types aligned in terms of functionality. The seamless network must also be backward compatible and future-proof—that is, it must offer capacity and coverage fallback from GSM to WCDMA and vice versa. The load-sharing mechanism is thus dependent on how the operator builds and structures the common network.

Figure 4
GSM circuit emulation.



The path to a seamless network can be broken down into three steps (Figure 5). The arguments for providing a common network in each step may differ somewhat over time. Note: For some operators (mainly in the Americas), the steps will differ—they will deploy EDGE before WCDMA.

In many cases, the first implementation of WCDMA will be driven by regulatory requirements and mainly be concentrated to urban areas. Circuit-switched traffic will continue to dominate, which means that the load on the WCDMA radio network will generally be low.

While building coverage in their WCDMA networks, many operators will continue investing in GSM, to improve capacity and obtain coverage fallback for users who roam between the networks.

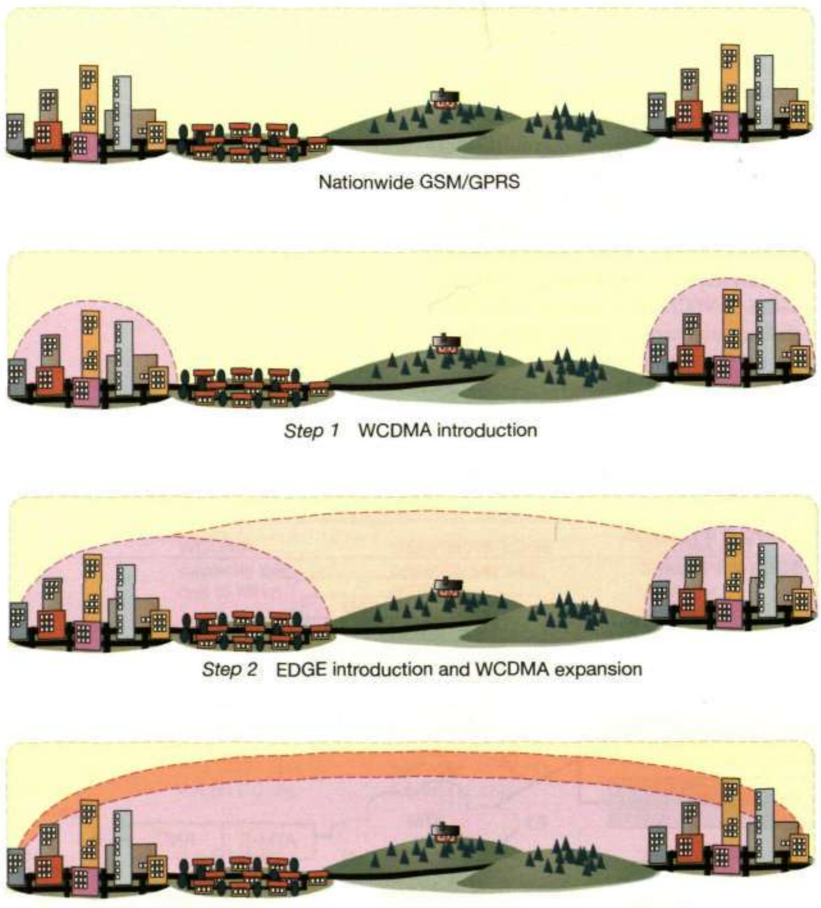
When numerous, attractive data services that require faster transfer rates have been introduced on a large scale, operators will be driven to increase capacity and extend their WCDMA networks to suburban areas. The operators can benefit from sharing traffic load between the GSM network and the WCDMA network. The main focus will be on quality assurance and transmission efficiency. Circuit-switched traffic will continue to dominate, while the load in some WCDMA cells (especially those in dense urban areas) will be heavy.

During the next phase of build-out, when WCDMA has become a mass-market technology, the focus will be on network optimization and beefing up capacity in all major areas. WCDMA coverage will also be extended into some rural areas.

Optimizing traffic between different access types and using all available spectrum in the most efficient way will be an essential aim of extending the network. The Ericsson seamless concept, with ATC and SCS, together with the combined network management system, will be supported in full.

Operators will derive great benefits from sharing sites and reusing power and transmission in the combined radio network. The majority of site hardware can be shared. While these benefits can be achieved earlier they are of prime importance when expanding capacity in urban areas and extending the WCDMA network outside urban areas.

Figure 5
Three phases of roll-out and expansion.



Load sharing and call cases

Load sharing

One of the main drivers for investing in the seamless network concept is the ability to steer traffic between the GSM and WCDMA radio networks. As mentioned above, this is handled by the ATC and SCS together with core network nodes, such as the HLR and GPRS support node (GSN). Basically, the operator has three ways of initiating load sharing:

- the sharing mechanism can be initiated by QoS requirements on services;
- it can be based on policy; or
- it can be triggered by type of subscription.

QoS-based service differentiation means that an application—for instance, streaming—moves to WCDMA when the GSM load is high. In the same way, an FTP file transfer on a background QoS is steered to GSM when the WCDMA load is moderately high. The capacity and coverage fallback is handled by the GSN.

With policy-based guidance, the operator defines parameters in the RNC and BSC, so that, for instance, all circuit-switched voice traffic and low-bandwidth data is managed in the GSM radio network and all high-bandwidth data is managed in the WCDMA radio network.

When load sharing is triggered by type of subscription, the operator assigns all "golden" customers (those with advanced third-generation services) to the best network—that is, the network that can provide the best service (lowest load, least congestion) at a certain point in time. This can be either the WCDMA or GSM network. The division of services between different types of subscription is set in the HLR and home subscriber server (HSS) nodes.

Call cases

The load-sharing function is also dependent on service type and QoS demands, handover possibilities, bandwidth, and the availability of GSM and WCDMA bearers. To steer traffic correctly between the GSM and WCDMA networks, each radio access bearer (RAB) must be aligned. For example, if a customer wants to run a Web-browsing service in the packet-switched domain while moving from one network to another, then each system must support an interactive or background bearer.

For a circuit-switched video telephony call, high-speed circuit-switched data (HSCSD) functionality is required in second-generation network environments, and a conversational bearer is required for circuit-switched multimedia in third-generation network environments. Likewise, for streaming applications on GPRS or EDGE, bearers that support streaming QoS are needed in second-generation environments; a streaming QoS bearer is required for packet data in WCDMA (Figure 6).

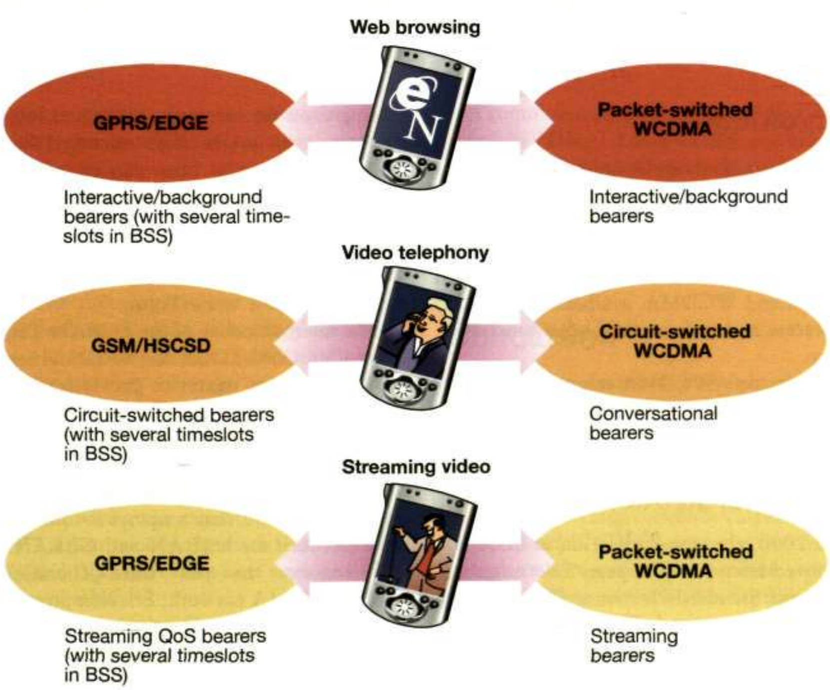


Figure 6
Examples of call cases and their relation to types of service, QoS class and bearers in GSM and WCDMA.

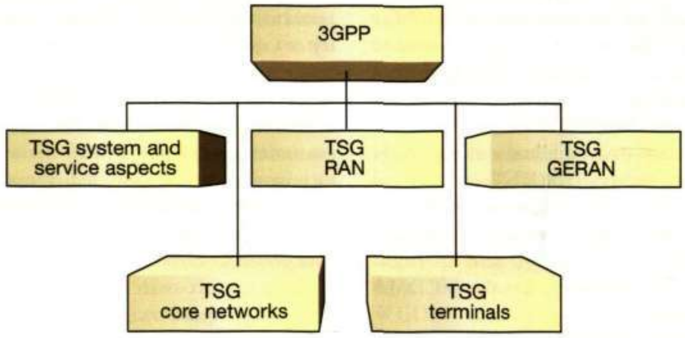


Figure 7
Standardization in 3GPP of GSM and WCDMA.

Evolution

At a high level, the seamless network concept relates to the evolution of the radio network, core network, and GSM and WCDMA network management. It also involves the handover mechanism between GSM and WCDMA, available radio access bearers, and multimode handset functionality.

Ericsson's WCDMA releases are coordinated with GSM releases to ensure that these areas are covered.

Standardization

In 2000, the standardization of GSM was moved from the European Telecommunications Standards Institute (ETSI) to the Third-generation Partnership Project (3GPP) to ensure the integrity of the GSM/WCDMA platform, thereby eliminating risks for incompatibility and inefficiency that might have occurred had the standardization been carried out by separate groups. A fifth technical specifications group (TSG) called the GSM/EDGE radio access network (GERAN) has been added to 3GPP to accommodate this work (Figure 7).

The main objective of the GERAN TSG is to align GSM/EDGE and WCDMA services, mainly as relates to providing conversational and streaming service classes. Best-effort and interactive service classes will also be supported.

These efforts will result in a GERAN system architecture that employs a common core network for UTRAN and GERAN.[2] To connect to the third-generation GSM/WCDMA network, Ericsson proposes to enhance the *Gb* interface to support a similar level of service as UTRAN. This mainly means support for the conversa-

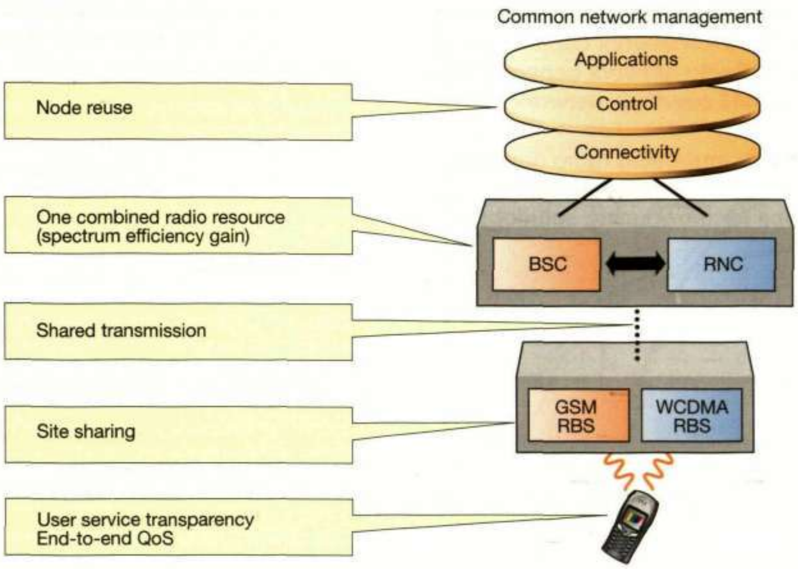


Figure 8
Benefits of the seamless network concept.

tional quality-of-service class, but with lower maximum bit rates.

Benefits

Many benefits can be derived from moving to a seamless network (Figure 8):

- The user receives the same services regardless of access technology—transparent user service. The focus is shifted from "technology coverage" to "applications coverage."
- Since the cost of base station sites is one of the operator's biggest, the ability to share sites between GSM and WCDMA will have great impact.
- Transmission can be shared between GSM and WCDMA from base stations to BSC and RNC.
- By regarding GSM and WCDMA radio accesses as a common radio resource, operators can obtain trunking gains.
- A common core network is used for GSM and WCDMA. Today's GSM and GPRS nodes can be reused to a large extent.
- A common network management system gives savings both in capital expenditures (CAPEX) and operating expenditures (OPEX).

Conclusion

The Ericsson seamless network concept supports the growth of today's services as well as the creation and growth of the Mobile Internet. Investments in GSM (past, present and future) are thus future-proof, and there is no conflict between investments in GSM and WCDMA since the two technologies evolve toward one network.

The seamless network adds flexibility in the deployment of third-generation services; it enhances system performance, and protects network investments by reusing resources.

REFERENCES

1 Gustås, P., Magnusson, P., Oom, J. and Storm, N.: Real-time performance monitoring and optimization of cellular systems. Ericsson Review Vol. 79(2002):1, pp 4-13.
2 Müller, F., Sorelius, J. and Turina, D.: Further evolution of the GSM/EDGE radio access network. Ericsson Review Vol. 78(2001):3, pp. 116-123.

Ericsson's GSM RAN capacity solutions

Peter Blom

High-capacity solutions are about building high-capacity networks in the most economical way, and therefore, GSM radio network capacity solutions are becoming increasingly important. Radio network capacity solutions can be divided into three categories: cell capacity, network capacity, and channel capacity.

The author discusses various solutions for improving radio network capacity in each of these areas. He also describes different implementations and recommends three general steps for introducing solutions in the network.

Capacity solutions for GSM radio access networks continue to be in focus. In the Americas, non-GSM technologies, such as TDMA and AMPS, occupy a substantial amount of spectrum. Likewise, competition makes it difficult for operators to acquire new spectrum. Therefore, operators who migrate to GSM are faced with the challenge of providing enough capacity for new services while maintaining capacity in legacy systems. The downturn in the world economy has also put constraints on operators, forcing them to maximize benefits from every investment. Since high-capacity solutions are about building high-capacity networks in the most economical way, GSM radio network capacity solutions are perhaps more important today than ever before.

Radio network capacity solutions can be divided into three solution categories:

- cell capacity solutions—these solutions consist of methods and features that permit more transceivers per cell;
- network capacity solutions—these solutions focus on adding different kinds of cells and on making the most of cell capacity by distributing traffic as efficiently as possible; and
- channel capacity solutions—these solutions center on ways of using the available throughput of the channels in the air in a more efficient manner, for example half-rate voice channels and GPRS.

Cell capacity

The one factor that has the greatest influence on cell capacity is frequency reuse. Cell capacity is thus determined by different methods and functions to enhance frequency reuse. Two common methods are

- multiple reuse pattern (MRP); and
- fractional load planning (FLP).

The multiple reuse pattern, which is based on baseband frequency hopping, yields the best results for networks composed mainly of filter combiners. The primary transceiver carries the broadcast control channel (BCCH) and must therefore have a relatively loose reuse pattern (explanation: a handset must listen to the information broadcast on the BCCH before it can make calls in a cell). But thanks to the frequency hopping gain, all remaining transceivers in the network can have a successively tighter reuse pattern. Compared to a non-hopping network, the MRP solution can more than double cell capacity. The drawbacks of MRP are that it requires

- considerable spectrum (greater than 7 MHz); and
- at least three transceivers per cell for good performance.

Fractional load planning is based on synthesized frequency hopping, which requires the use of hybrid combiners. In FLP, the gain from frequency hopping is not dependent on the number of transceivers in a cell, since each transceiver can hop on every frequency allocated to the cell. Notwithstanding, due to the characteristics of synthesized frequency hopping, the BCCH transceiver cannot hop frequencies. Ordinarily, to guarantee adequate voice quality for a non-hopping traffic channel (TCH), a frequency reuse of approximately 15-18 is needed. But by using the BCCH in an overlaid subcell feature, it is possible to plan BCCH frequency reuse as if the BCCH transceiver could hop frequencies, making a frequency reuse of approximately 11-12 feasible. The most common FLP methods in use are 1/3 and 1/1, and FLP can be implemented in frequency bands as narrow as 3 MHz.

BOX A, TERMS AND ABBREVIATIONS

AMPS	Advanced mobile phone system	HR	Half-rate
AMR	Adaptive multirate	IRC	Interference rejection combining
BCCH	Broadcast control channel	MAIO	Mobile allocation index offset
DTX	Discontinuous transmission	MRP	Multiple reuse pattern
EDGE	Enhanced data rates for global evolution	PCCCH	Packet-data common control channel
EFR	Enhanced full-rate	QoS	Quality of service
FAS	Frequency allocation support	RBS	Radio base station
FLP	Fractional load planning	TCH	Traffic channel
FR	Full-rate	TDMA	Time-division multiple access
GPRS	General packet radio service	TET	Traffic estimation tool
GPS	Global positioning system	WCDMA	Wideband code-division multiple access
GSM	Global system for mobile communication		

A third method, known as non-uniform frequency planning, can be a mix of MRP and FLP, or a totally free allocation of frequencies. Since this method is more complex, cell-planning and measurement tools are recommended, such as TEMS Cell Planner and frequency allocation support (FAS).

Ericsson's GSM system provides a host of features that minimize frequency reuse and maximize cell capacity. The most basic feature—frequency hopping—is unique to GSM and is the basis for all tight frequency reuse solutions. Since a cell can accommodate up to 16 different hopping or non-hopping frequency groups (channel groups), there is considerable inherent flexibility for adapting the network to different services with different requirements for radio quality. Radio base station (RBS) site synchronization and mobile allocation index offset (MAIO) management are provided to maximize the benefits of FLP. MAIO is the parameter that determines when a frequency is used in a cell that employs FLP. With proper MAIO planning it is possible to minimize or even eliminate interference between synchronized cells.

Similarly, quality-based dynamic power control and discontinuous transmission (DTX) are employed to minimize radio interference. The quality-based power control feature performs very well in networks that employ tight frequency reuse—it provides a substantial decrease in output power, especially compared to non-quality-based algorithms. Should quality-related problems persist, an intra-cell handover function finds a new channel in the cell on which the call can continue.

In addition, the dynamic overlaid/underlaid subcells feature divides the cell into two subcells with traffic management functionality between them. It also makes it possible to restrict the coverage of the overlaid subcell, thereby facilitating even tighter reuse in it. And finally, there is the adaptive multirate (AMR) voice codec for GSM full-rate channels. The feature uses several voice codecs and associated error protection (channel coding) to adapt to the quality of the radio environment. Compared to the full-rate (FR) and extended full-rate (EFR) voice codecs, this feature greatly improves robustness. In fact, compared to EFR, when AMR full-rate is used to its fullest extent—to tighten frequency reuse and to add more transceivers—the traffic capacity in the cell can be doubled.

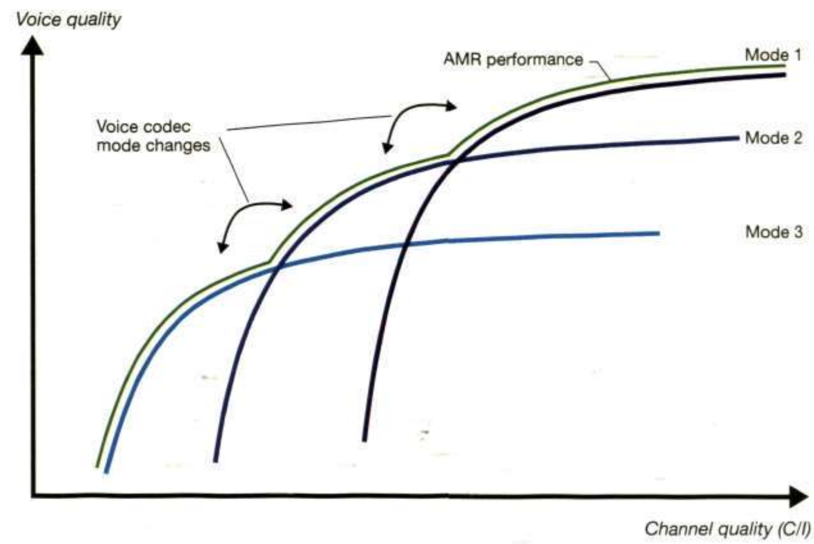


Figure 1
The principles of AMR. Voice quality is maximized by adapting—or by switching between several voice codecs—to the quality of the radio channel.

Network capacity

In addition to improving cell capacity, operators can introduce micro cells, since site acquisition for micro cells is usually easier and less expensive than when adding regular cells. To facilitate the deployment of micro cells, Ericsson provides the traffic estimation tool (TET), which enables operators to identify the optimal site location. And since it is possible to measure the amount of traffic a certain site location will carry, he can then calculate time to payback before the operator installs the base station.

A multiband network option is open to operators who have frequencies in a second frequency band. The amount of capacity that can be derived from implementing cells in a second frequency band depends on the amount of spectrum in the alternative frequency band. Even so, since the cell capacity solutions described above can be applied in any frequency band, the extra capacity is never negligible. To derive the optimum cost-benefit ratio, Ericsson recommends that the operator reuse all existing sites for RBSs that belong to the new frequency band.

Traffic management is an important issue in a network composed of cells of different

Figure 2
Network capacity: Efficient traffic-management features, such as Ericsson's multilayered hierarchical cell structure (HCS), should be employed when different kinds of cells and frequency bands are used.

sizes and frequency bands. Intelligent traffic distribution algorithms let cells cooperate and help one another to enhance network capacity to a degree that exceeds the sum of all individual cells.

With multilayered hierarchical cell structures—the most important traffic-handling function in Ericsson's GSM system—cells can be divided in up to eight layers, and traffic can be prioritized and distributed between these layers. There are also numerous add-on functions, such as

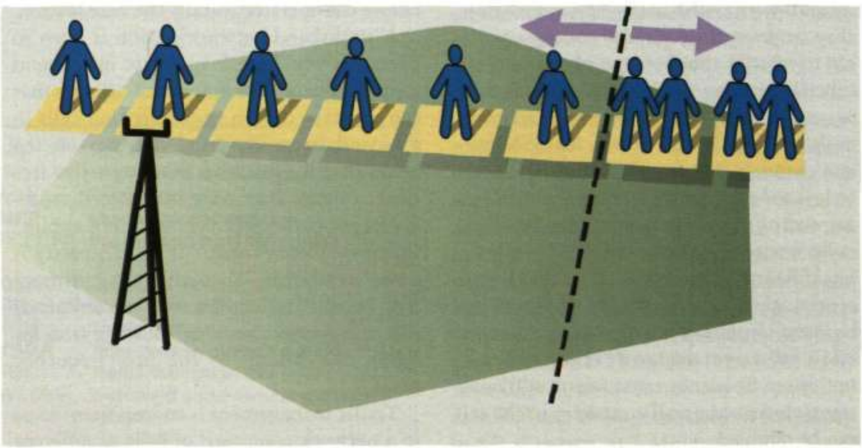
- *cell load sharing*, which distributes traffic within layers;
- *assignment to another cell*, which redirects traffic to other cells when congestion occurs during call setup; and
- *handling of fast-moving mobiles*, which moves calls to higher layers when there are too many handovers within a given interval. This function reduces the number of handovers, thereby increasing voice quality and reducing processor load.

Channel capacity

In the context of circuit-switched traffic, channel capacity is about half-rate voice channels and the way they are managed. Since the half-rate technique reduces the quality of voice, it has not been widely deployed. However, operators are now beginning to use this technique more and more, since it can be allocated on a dynamic basis (dynamic HR allocation) during traffic peaks.

In the context of data communications, GPRS is a channel capacity solution. It makes optimum use of channels and maximizes capacity by allowing several users to share the same channels. Ericsson's GPRS solution provides dedicated as well as on-demand packet-data channels. The solution also supports dedicated packet-data common control channels (PCCCH). During 2002, numerous improvements will be made available, including EDGE and

Figure 3
Dynamic half-rate allocation: To mitigate the impact on voice quality, half-rate techniques should only be used during traffic peaks.



quality-of-service-based (QoS) scheduling. EDGE technology, which can more than triple throughput per channel, enhances packet data capacity and facilitates a multitude of new services that require extra bandwidth. The QoS scheduling functions allow operators to differentiate their service offering to distinct user segments.

Capacity gains

Figure 4 compares the most common options for improving network capacity. As can be seen, most of the methods have the potential to double network capacity, and our recommendation is that the operator should implement or apply as many of these as possible. In particular, two solutions really stand out: AMR full-rate, and micro cells. Note: To derive the greatest gains from AMR full-rate, the penetration of AMR handsets must be high.

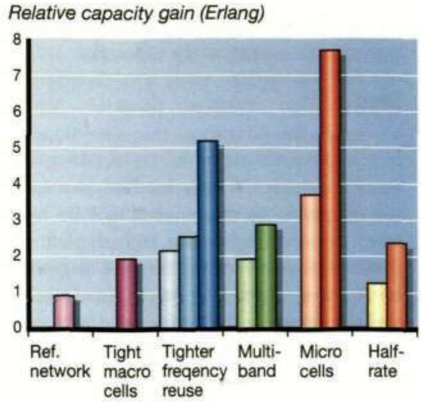


Figure 4
Reference network: Average network without frequency hopping.
Tight macro cells: Double the number of macro cells.
Tighter frequency reuse: (a) MRP network based on EFR; (b) FLP 1/1 network based on EFR; (c) FLP 1/1 network based on AMR.
Multiband: (a) 50% penetration of capable terminals; and (b) 100% penetration of capable terminals.
Micro cells: One micro cell every 200 m. (a) 2 TRX per micro cell; (b) 4 TRX per micro cell.
Half-rate: (a) 25% penetration of capable terminals; (b) 100% penetration of capable terminals.

Implementation aspects

Apart from gains in capacity, the two main parameters that an operator should consider when building a network are monetary cost and time—the actual cost of each solution is market-dependent, since the costs associated with cell sites (site acquisition, site preparation, rental costs) and transmission vary from market to market.

In every market it costs time and money to build sites. Accordingly, the greater the number of sites required, the higher the cost. When viewed in this light, we can conclude that tight frequency reuse offers the most expedient and cost-effective solution to improving capacity, since in many cases the operator needs only add transceivers to existing cabinets.

If the operator wants to maximize his use of existing sites then the second-best option is to deploy transceivers on other available frequency bands. In this case, the operator needs only add cabinets at sites where extra capacity is wanted.

A third option is to introduce micro cells—thanks to the small size of micro base stations, it is easier and less expensive to acquire sites for them.

A final option is to build more sites.

As mentioned above, the operator might also make use of half-rate channels, but since this option decreases voice quality, it should be allocated dynamically (dynamic HR allocation), and then mostly as a last-resort option when network expansion through other means cannot keep pace with growth in traffic. The half-rate technique can be deployed very quickly by activating various software features.

Recommended way of increasing capacity

Taking into account the potential for capacity and the implementation aspects for

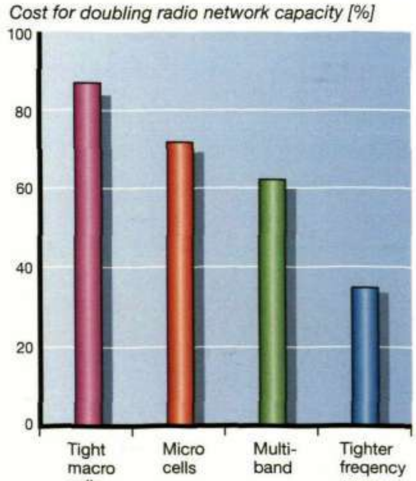


Figure 5
Cost comparison: The additional cost of doubling radio network capacity using various solutions relative to the original investment.

Figure 6
Micro base station RBS 2302.

each capacity-boosting solution, we can recommend a general order for introducing the different solutions in the network.

Step 1

Operators should activate frequency hopping and implement one of the tighter frequency reuse methods. Which method should be used depends on factors such as RBS hardware, amount of spectrum, and cell plan. Additional reuse-enhancing features might also be necessary depending on the reuse method and the degree of frequency use.

Operators should activate AMR. Although the initial gain in capacity is small, a significant improvement in quality can be achieved for users with AMR-capable terminals. As the penetration of AMR handsets increases it will allow more transceivers to be deployed in the network, gradually enhancing traffic capacity.

Step 2

Operators who have access to spectrum in other frequency bands should start deploying equipment to use those frequencies. Costs can be kept to a minimum if operators reuse existing sites.

Operators can add micro cells or indoor cells at traffic hot-spots, which include popular squares, conference centers, shopping malls and airports. The use of micro cells to cover hot-spots will offload the macro cells, and help operators to avoid the cost of having to split cells.

Operators can also use dynamic half-rate allocation to avoid congestion at peak traffic. This measure reduces the pressure on operators to build out the network, allowing them to build it out at a manageable pace.

Step 3

As traffic increases, the number of micro cells and indoor cells will also continue to grow. At some point, the micro cell layer will be almost contiguous. By adding more capacity to the micro cells and indoor cells, operators can achieve an extreme boost in capacity.

Future GSM RAN capacity solutions

GSM has been in commercial operation for some 10 years now, but there is still ample room for enhancing capacity. One area that shows great potential is the use of advanced interference-suppression algorithms. In the base station, we will see this type of algorithm introduced in the form of interference rejection combining (IRC). The interference-suppression performance of this feature far outperforms present-day receiver algorithms. Simulations show that IRC can give link gains of up to 5 dB in non-synchronized networks, and up to double that in synchronized networks. This can be translated into better voice quality and data throughput on the uplink. However, due to a lack of corresponding functionality in present-day handsets, the gain in capacity will be limited. To remedy this situation Ericsson is also researching interference suppression algorithms for terminals. Limited processing capacity, small size, and the importance of design (for example, only one antenna) impose harsh restrictions on algorithms. However, Ericsson has made some technological breakthroughs and expects to introduce powerful interference-suppression algorithms in handsets in the next few years.

Apart from boosting IRC performance, synchronized radio networks also enhance capacity. Synchronization is achieved by synchronizing all RBS sites to the GPS system. When all cells are synchronized to the same reference clock, interference planning will no longer be limited to cells located at the same site. Instead, operators will be able to determine which cells truly interfere with each other, no matter where they are located. Here, too, simulations show gains of up to 20% compared to a site-synchronized network.

AMR technology will also enhance the channel capacity solution for half-rate channels: AMR half-rate consists of a subset of the AMR voice codecs defined for full-rate channels. As with AMR full-rate, it adapts, by switching between voice codecs, to the quality of the radio environment. AMR half-rate improves voice quality compared to the present voice codec for half-rate channels, making the half-rate option a more attractive solution for increasing capacity.

Looking further ahead, better techniques for distributing traffic between GSM and WCDMA will be introduced until we can see a seamless GSM-WCDMA network, which will allow operators to use both technologies to their fullest, to maximize the end-user experience. It will also facilitate the introduction of new functionality in GSM. The introduction of adaptive antenna technology will yield even greater gains. Ericsson has tested adaptive antennas extensively to verify their performance. These tests show that adaptive antennas have the potential to more than double network capacity. In fact, it will be nearly doubled if adaptive antennas are installed at only 20-30% of the sites in a given area.

Conclusion

Radio network capacity solutions can be divided into three categories: cell capacity, network capacity, and channel capacity.

The one factor that has the greatest influence on cell capacity is frequency reuse. That is, cell capacity is determined by different frequency reuse methods and functions to enhance it. Two common methods are multiple reuse pattern and fractional load planning. A third frequency reuse method is non-uniform frequency planning.

To improve network capacity, operators can introduce micro cells, multiband operation (if additional spectrum is available), traffic management, and multilayered hierarchical cell structures.

Dynamic half-rate allocation technology enables operators to increase channel capacity for circuit-switched traffic during traffic peaks. Likewise, GPRS is a channel capacity solution for data communication—it makes optimum use of channels and maximizes capacity by allowing several users to share the same channels.

A recommended three-step approach to increasing capacity in the radio access network is as follows:

1. Activate frequency-hopping, employ tighter frequency reuse, and activate AMR.
2. If spectrum is available in a second band, deploy equipment in those frequencies. Add micro or indoor cells at hot-spots, and use dynamic half-rate allocation to reduce congestion during peak traffic.
3. Add more micro and indoor cells and increase the capacity in them.

Previous issues

No. 1, 2002

Real-time performance monitoring and optimization of cellular systems
IP technology in WCDMA/GSM core networks
The future of communication using SIP
Ericsson Mobile Operator WLAN solution

No. 4, 2001

Secure electronic transactions—The mobile phone evolution continues
On-demand mobile media—A rich service experience for mobile users
WAP 2.x architecture—Features, services and functions
RF multicarrier amplifier for third-generation systems
Ericsson's family of carrier-class technologies
Building a reliable, cost-effective and future-proof fiber optical access network

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

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

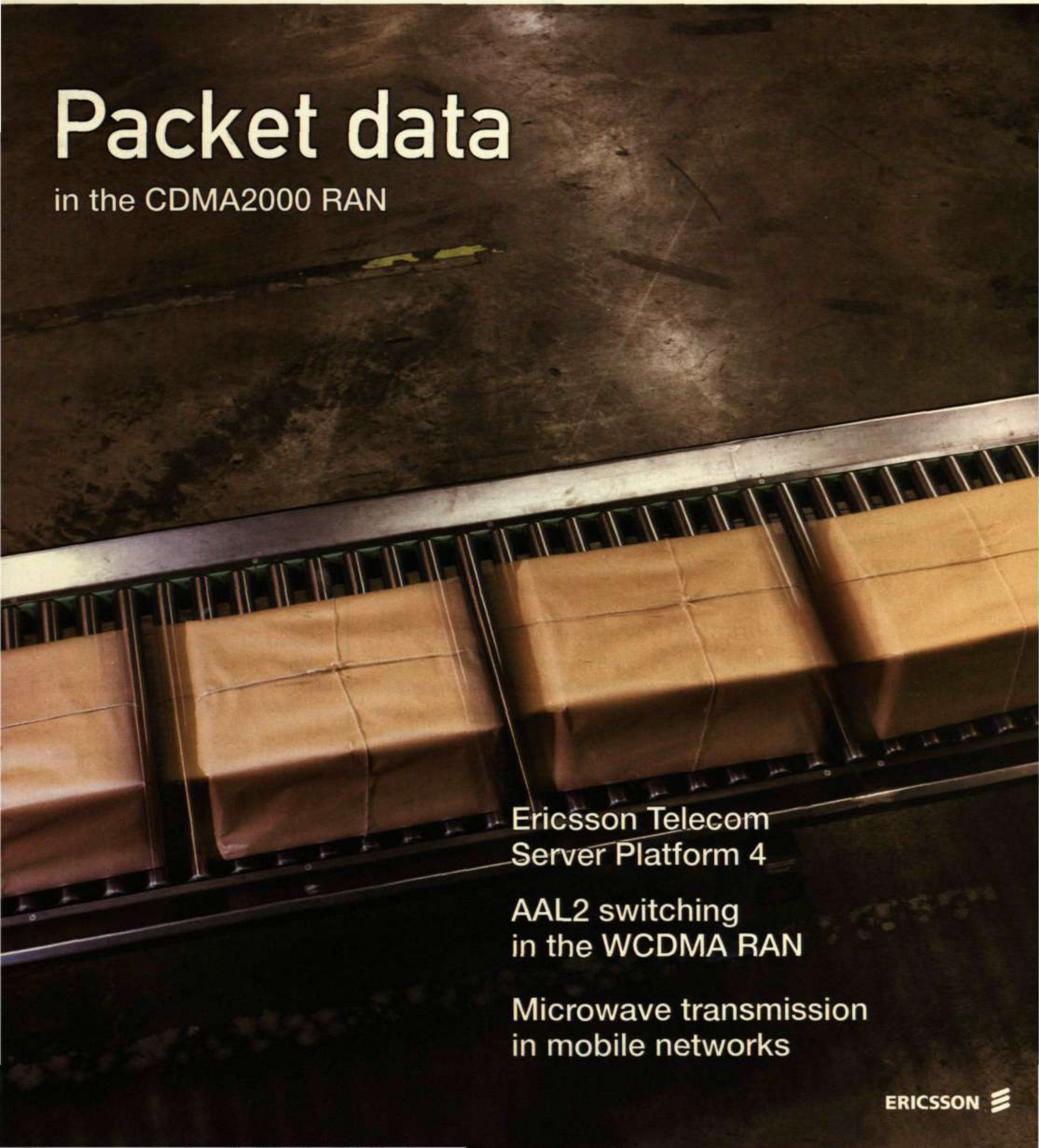
ISSN 0014-0171
Ljungföretagen, Örebro 2002

Ericsson

REVIEW

3/2002

THE TELECOMMUNICATIONS TECHNOLOGY JOURNAL



Packet data

in the CDMA2000 RAN

Ericsson Telecom Server Platform 4

AAL2 switching in the WCDMA RAN

Microwave transmission in mobile networks

ERICSSON

Cover: A key feature of CDMA2000 1X service is that the packet-data connection between the mobile station and the network is always on. Moreover, as the CDMA2000 air interface continues to evolve, the peak data rates will rise from today's153.6 kbit/s (1X Rev. 0) to 307.2 kbit/s (1X Rev. A) to 2.4 Mbit/s (1xEV-DO) to 3.1 Mbit/s (1xEV-DV), and finally, to All-IP service!

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 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:
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: 79, 2002

Contents

Contributors

In this issue

Aldo Bolle joined Ericsson in 1997, to work with product and system management of Ericsson's point-to-multipoint microwave system. From 1999 to 2001 he served as manager of the customer solution department. At present, he is a member of the strategic product management team at Ericsson Microwave Networks in Mölndal, Sweden. Before joining Ericsson, Aldo worked with SDH, ATM and network strategies at Telia Network Services. He holds an M.S. and Ph.D. in engineering physics from Chalmers University of Technology in Göteborg, Sweden.
aldo.bolle@emw.ericsson.se

Marco Coscia joined Ericsson in 1998 as a senior engineer supporting cdmaOne systems in San Diego. He then served as a manager in the project engineering group. In 2001, he joined Ericsson's C-RAN product management group as senior product manager. He is currently responsible for cdmaOne and CDMA2000 BSC products. His primary focus is on the BSC, BSC site solutions, and C-RAN transport solutions. Marco holds a B.S. in electrical engineering from the University of Connecticut and an MBA from Pepperdine University.
marco.coscia@ewu.ericsson.se

Victor Ferraro-Esparza joined Ericsson in 1998 as a systems engineer working with subscriber databases for GSM, GPRS and WCDMA mobile systems. In 1999, he worked with GSM number portability solutions. He was also appointed to represent Ericsson's interests in numbering, naming and addressing, coordinating activities relating to technology, business trends, products, standardization and regulatory efforts within various standardization bodies. In 2001, he became a product manager of numbering, naming and addressing issues with the responsibility of evolving the subscriber database business in all-IP networks and the NDVAS portfolio. In 2002, he became strategic product manager of the Ericsson AAA Server. Victor holds a graduate degree in telecommunication engineering and telecommunication systems from Universidad Politécnica Valencia, Spain.
victor.ferraro-esparza@ece.ericsson.se

Michael Gudmandsen is currently strategic product manager of the Ericsson Telecom Server Platform. Since joining Ericsson, in 1991, he has also worked with open platforms for OSS development, OSS (TMOS) and O&M-related product development, and as strategic product manager of AXE element management and I/O products. He graduated from Chalmers University of Technology in 1983.
michael.gudmandsen@uab.ericsson.se

Bo Karlander is presently senior specialist in the WCDMA RAN user plane and transport network. Since joining Ericsson, in 1967, he has worked with system analysis and traffic dimensioning issues for the first generation of computer-controlled telephone systems and with system design of data communications and multiservice switching systems for private and public networks built on packet switching, frame relay and ATM technologies. He was also involved in the initial development phase of CPP, on which Ericsson's WCDMA radio access network nodes are built. Since 1997, he has worked with requirement specifications for, and the system architecture of, Ericsson's radio access network system for WCDMA. Bo holds an M.S. in electrical engineering from the Royal Institute of Technology, Sweden.
Bo.karlander@era.ericsson.se

Kabir Kasargod is a product manager at Ericsson CDMA Systems in San Diego in charge of managing software feature requirements for Ericsson's first CDMA2000 software release. He joined Ericsson in 1999 as part of the Qualcomm acquisition. While at Ericsson, he has served in several positions including project engineer for Russian and Brazilian market deployments, senior engineer within the Latin America

technical development group, and product manager for the Ericsson cdmaOne BSC. Kabir holds a B.E. in electrical engineering from VJTI-University of Bombay, India, and an M.S. in electrical engineering from Rutgers University, New Brunswick, New Jersey, USA.
Kabir.kasargod@ewu.ericsson.se

Szilveszter Nádas joined Ericsson Research in 1998 to work at the Traffic Analysis and Network Performance Laboratory. Since 2000, the main focus of his work has been on modeling and analysis of the WCDMA radio access network. He holds an M.S. in electrical engineering from, and is also currently working toward a doctorate degree at, the Budapest University of Technology and Economics, Hungary.
szilveszter.Nadas@eth.ericsson.se

Andrea Nascimbene joined Ericsson, in 1991, to work with product marketing and system planning for private mobile radio. He later joined Ericsson Lab Italy and has been with the MINI-LINK BAS project since its start in 1998. He has worked as product manager and carried out extensive work in the area of standardization, serving as chairman in various ETSI and CEPT projects. He is currently line manager in charge of product and system management of MINI-LINK High Capacity terminals. He holds an M.S. in electronic and telecommunications engineering from the University of Pavia, Italy.
andrea.nascimbene@eri.ericsson.se

Kristofer Olsson is currently employed as a strategic product manager of the J-SCS at Ericsson's NDVAS product development unit. He joined Ericsson, in 1996, working as a product manager at Business Unit Japan for intelligent networks. He then took a position as strategic product manager within intelligent networks for GSM, and later, participated in developing the consultant program at Ericsson Mobile Systems. Kristofer graduated from the Linköping Institute of Technology with a degree in industrial engineering and management.
Kristofer.olsson@era.ericsson.se

Sandor Rácz joined Ericsson in 2000 and is currently a research fellow at the Ericsson Traffic Analysis and Network Performance Laboratory in Budapest, Hungary. His research interests include modeling and analysis of communication systems. Sandor holds an M.S. in electrical engineering with emphasis on telecommunications from the Budapest University of Technology and Economics, Hungary.
sandor.Racz@eth.ericsson.se

Jonas Renius, who joined Ericsson in 1969, is currently product manager of transport functions in the WCDMA radio access network. Since 1990, his chief area of focus has been on ATM. In 1995, for example, he helped define CPP, and he has since served as product manager of ATM products and of CPP and transport networks. Prior to 1990, he held numerous project and product manager positions within hardware and system design, primarily relating to PBX and broadband systems.
jonas.reinius@era.ericsson.se

Mike Sheppard joined Ericsson in 1997, initially managing the deployment of cdmaOne infrastructure as part of the BSS project management group. In 1998, he was placed in charge of all project deployments in North America. In April 2000, he was appointed release manager of cdmaOne software products. He recently served as senior product manager in charge of cdmaOne software products and CDMA2000 Release P2. Mike holds a B.S. in electrical engineering from Georgia Southern University and an MBA from Troy State University in Birmingham, Alabama.

Ericsson Review is also published on the Web. Visit us at http://www.ericsson.com/review

Editorial

Eric Peterson

This past summer, the Mobile Internet saved my bacon! Earlier in the spring, my uncle e-mailed me saying that he and his wife would be visiting Stockholm together with their son as part of an extended Baltic cruise. He asked if I would be home, and, if so, could we meet? He also gave me the name of the cruise line he would be traveling with and explained that I could monitor the Ports of Stockholm website (www.portsofstockholm.com) to learn when and where his vessel would call. I replied that I would indeed be home and would gladly meet him and play host for a day.

As the appointed day drew nigh, I again confirmed (my uncle had e-mail access on his ship) that I had been monitoring the websites and that I would be on hand when the vessel arrived. Unfortunately, neither my uncle nor I realized that TWO ships from the same cruise line would be calling on Stockholm on the same day and at the same hour. And as you have probably guessed, I went to greet the *other* ship.

I arrived early and anxiously watched as the ship was towed and moored, the gangway was lowered, and hundreds of eager tourists tumbled ashore and were herded toward the queue of sightseeing buses that stood ready to transport them to the city. Two hours later, however, after the last of the buses had departed and only crewmembers remained, I was kicking myself for not thinking to give my uncle the number to my mobile phone: so confident was I in the information I had, that it really hadn't seemed important. Now, of course, I felt downright stupid. The good news is that I get e-mail on my phone. When my uncle could see that I was not there to greet him, he returned to his cabin and e-mailed me the name of his ship and its berth. Moments later I was back inside my car racing to the other side of Stockholm to where my uncle, aunt and cousin were waiting. Thank goodness for the Mobile Internet!

And in case you wondered, we had a terrific visit.

Eric Peterson
Editor

Packet data in the Ericsson CDMA2000 radio access network

Kabir Kasargod, Mike Sheppard and Marco Coscia

Ericsson's CDMA2000 radio access network consists of a base station controller, radio base stations, a radio network manager and TEMS for CDMA2000. It uses open IOS/IS-2001-compliant interfaces to the mobile switching center (MSC) and packet-data service nodes (PDSN). A key feature of CDMA2000 1X packet-data service is that the packet-data connection between the mobile station and the network is always on.

This article discusses the CDMA2000 1X radio access network, the different services it supports, and how always-on, 144 kbit/s packet-data services are implemented. Concepts relating to 144 kbit/s packet-data service are presented from an end-user's perspective.

Introduction

In 1999, the International Telecommunication Union (ITU) approved an industry standard for third-generation wireless networks. This standard, called International Mobile Telecommunications-2000 (IMT-2000), is composed of three standards—commonly referred to as WCDMA, CDMA2000 and TD-SCDMA—based on code-division multiple access (CDMA) technology.

Within the CDMA2000 standard, several phases have been defined to support the ITU requirements for third-generation services. Figure 1 shows the evolution of the CDMA2000 standard. CDMA2000 1X, which is based on the IS-2000 standard (Revisions 0 and A), is

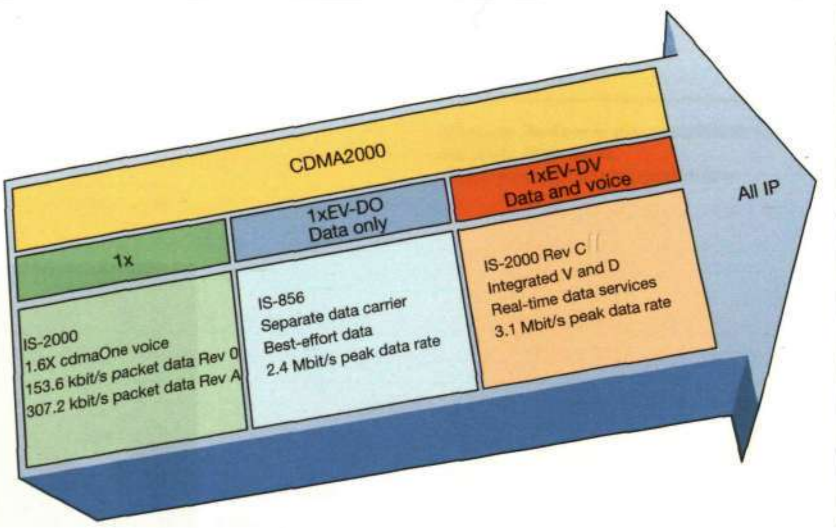
- backward compatible with cdmaOne;
- offers up to twice the voice capacity of cdmaOne systems;
- offers improved terminal battery life;
- supports always-on packet-data sessions; and
- provides data rates of up to 144 kbit/s—with peak over-the-air data rates (including overhead) of 163.2 kbit/s (IS-2000 Rev. 0) and 307.2 kbit/s (IS-2000 Rev. A).

Beyond CDMA2000 1X, the Third-generation Partnership Project 2 (3GPP2) proposes two (1xEV) standards: 1xEV-DO (data only) and 1xEV-DV (data and voice).

CDMA2000 1xEV-DO comprises a separate data carrier that provides best-effort packet-data service with a peak over-the-air data rate of 2.4 Mbit/s. CDMA2000 1xEV-DV provides integrated voice and data with real-time data services and a peak over-the-air data rate of 3.1 Mbit/s.

Ericsson has taken a leadership role in the standardization and commercialization of CDMA2000 services. Its CDMA2000 1X solution is based on true third-generation platforms that protect operator investments and provide a smooth migration path to CDMA2000 1xEV and beyond.

Figure 1
Evolution of the CDMA2000 air-interface standard.



Overview of the Ericsson CDMA2000 RAN

The Ericsson CDMA2000 radio access network (Figure 2) consists of

- a base station controller (BSC1120) with built-in packet control function (PCF);
- radio base stations (RBS1127/RBS1130);
- a radio network manager (RNM); and
- TEMS for CDMA2000, for planning and optimization of the radio network.

The BSC controls the RBSs, manages radio network resources, and provides user mobility. It also performs voice compression (vocoding), processes handovers, manages power control to ensure efficient use of network capacity, controls timing and synchronization within the radio access network, and provides interfaces to the RBSs, RNM and packet-data service nodes (PDSN).

The RBSs provide the radio resources and maintain radio links to mobile stations. The RBSs and BSC contain all necessary functions for their own management. The RNM supports operations at the CDMA2000 radio access network level.

TEMS for CDMA2000 enables operators to design integrated networks, test air interfaces, monitor performance, and optimize the design of CDMA2000 radio access and transport networks.

The Ericsson CDMA2000 radio access network uses open IOS/IS-2001-compliant interfaces to the mobile switching center (MSC) and PDSN. Ericsson has demonstrated the capabilities of the IOS interface through numerous inter-vendor deployments. This open-architecture philosophy provides flexibility and interoperability, and protects the investments of network operators who want to provide always-on, 144 kbit/s packet-data service.

Based on the IOS architecture, the BSC is connected to

- the core network (MSC) via an *A1/A2/A5* interface;
- an internal packet control function via an *A8/A9* interface; and
- the PDSN via an *A10/A11* interface.

The BSCs are interconnected within the radio access network via the *A3/A7* interface for inter-BSC soft handover. The RBSs are connected to the BSC via the *Abis* interface. The mobile station is connected to the RBS via the *Um* interface (the air interface) defined by the IS-2000 standard.

Inside the Ericsson CDMA2000 BSC, the PCF provides an interface to the PDSN via the RAN-to-PDSN interface—also known as the *R-P* or *A10/A11* interface. IS-2001 defines the *R-P* interface as two separate interfaces: the *A10* interface, which carries user data, and the *A11* interface, which carries signaling data.

The PCF is responsible for

- managing the packet-data states (active, dormant) of the mobile station;
- relaying packets between the mobile station and the PDSN;
- buffering data received from the PDSN for dormant mobile stations;
- supporting handovers; and
- PDSN selection.

A key feature of CDMA2000 1X packet-data service is that the packet-data connection between the mobile station and the network is always on (Figure 3).

Always-on service

To set up a packet-data call, a point-to-point protocol (PPP) session must first be established between the PCF and the PDSN. The first time a mobile station connects to the PDSN it establishes the connection via a packet-data call. Once the mobile station has made a PPP connection to the PDSN, it remains connected to the network. All subsequent data transmissions between the mobile station and the PDSN can be initiated by the mobile station or by the PDSN over the PPP connection.

Although the connection is always on, the CDMA2000 1X packet-data service does make provisions to preserve unused radio

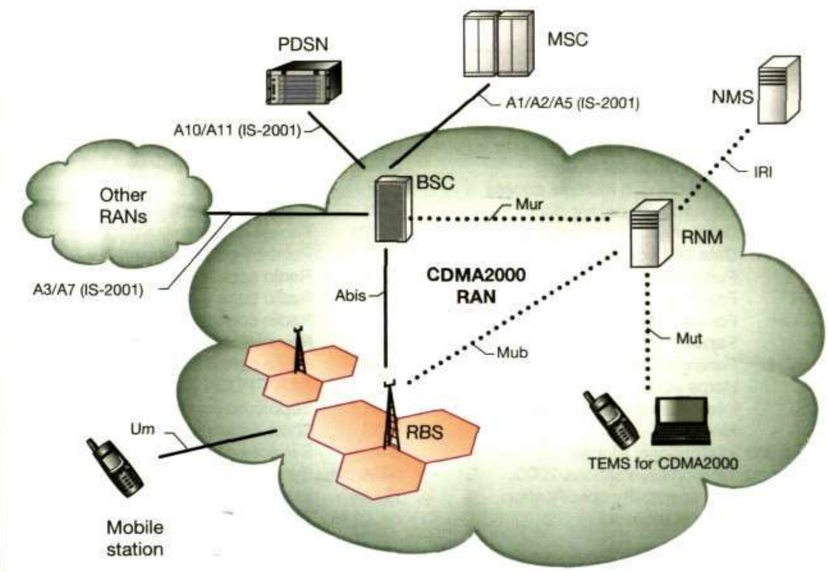
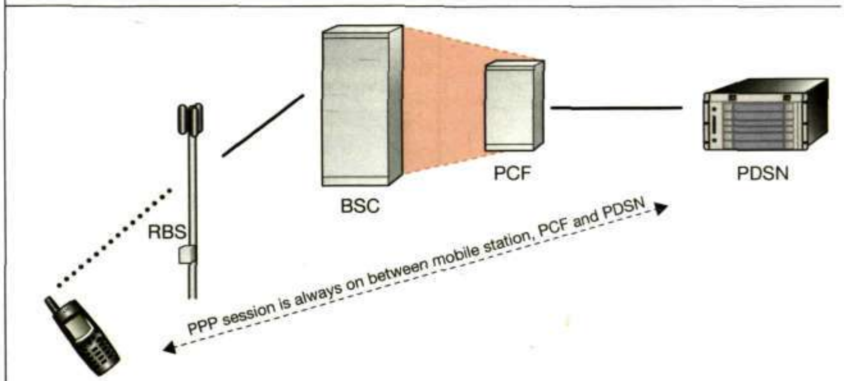


Figure 2
Overview of the radio network.

Figure 3
Always-on mobile station-to-PDSN connection.



BOX A, TERMS AND ABBREVIATIONS

3GPP2	Third-generation Partnership Project 2	ITU	International Telecommunication Union
AAL2	ATM adaptation layer type 2	MIB	Mobile IP
ATM	Asynchronous transfer mode	MSC	Mobile switching center
BSC	Base station controller	NMS	Network management system
CDMA	Code-division multiple access	O&M	Operation and maintenance
CPP	Connectivity packet platform	PCF	Packet control function
DO	Data-only	PCN	Packet core network
DV	Data and voice	PDSN	Packet-data service node
FCH	Fundamental channel	PPP	Point-to-point protocol
FER	Frame error rate	RAN	Radio access network
F-FCH	Forward FCH	RBS	Radio base station
FL	Forward link	RC3	Radio configuration 3
F-SCH	Forward SCH	RL	Reverse link
HA	Home agent	RNM	Radio network manager
IMSI	International mobile subscriber identity	R-P	RAN – PDSN (interface)
IMT-2000	International Mobile Telecommunications-2000	SCH	Supplemental channel
IOS	Interoperability specification	SMS	Short message service
IP	Internet protocol	SO	Service option
		STM	Synchronous transfer mode
		TEMS	Ericsson network optimization tools
		WCDMA	Wideband CDMA

link resources. That is, when the mobile station is neither sending nor receiving data and has been inactive for a certain period, the PCF tears down the radio link between the mobile station and the radio access network but maintains the connection (PPP session) between the mobile station and PDSN. This state is called dormancy.

Overview of the CDMA2000 1X packet-data service

Service options (SO) are used within the radio access network to identify different kinds of service, such as voice, circuit-switched data, packet data, and short message service (SMS). Table 1 shows the designated service options for IS-95 and IS-2000 packet-data service.

Ericsson provides 144 kbit/s packet-data service with over-the-air peak data rate (including overhead) of 163.2 kbit/s as specified in TIA/EIA/IS-707-A-1 service option 33 for IS-2000. When a packet-data call is set up, Ericsson's CDMA2000 radio access network uses radio configuration 3 (RC3) and a fundamental channel (FCH) to provide an initial packet-data rate of 9.6 kbit/s. To ensure backward compatibility with cdmaOne, the fundamental channel in CDMA2000 is similar to that in cdmaOne. The fundamental channel is mainly used for voice services, but in CDMA2000 it also supports low-data-rate packet services (9.6 kbit/s) and can be used to transmit signaling information.

For applications that require data rates exceeding 9.6 kbit/s, Ericsson's CDMA2000 radio access network uses an additional channel—the supplemental channel (SCH). Used in combination to transmit data, the fundamental and supplemental channels can provide 144 kbit/s packet-data service with over-the-air peak data rate (including overhead) of 163.2 kbit/s in the forward (to the mobile station) and reverse (from the mobile station) directions.

The Ericsson CDMA2000 radio access network uses the forward fundamental channel (F-FCH) and the forward supplemental channel (F-SCH) on the forward link, and the reverse fundamental channel (R-FCH) and reverse supplemental channel (R-SCH) on the reverse link. Table 2 shows the data rates of SO 33.

Delivering CDMA2000 packet data—an end-user's perspective

CDMA2000 packet-data call set-up

In the example that follows, a user—whom we call Janice—decides to catch up on some work using a data-capable mobile device to initiate a packet-data session and fetch her e-mail.

From the perspective of the radio access

TABLE 1, PACKET DATA SERVICE OPTIONS

Service option	Designated type of service	Max data rate	Associated standards
7	Rate set 1: IS-95 packet-data service	9.6 kbit/s	TIA/EIA/IS-657
15	Rate set 2: IS-95 packet-data service	14.4 kbit/s	TIA/EIA/IS-707
33	144 kbit/s packet-data service	163.2 kbit/s	TIAEIA/IS-707-A-1

153.6 + 9 .6 kbit/s indicates the combination of peak over-the-air data rates provided by the SCH (153.6 kbit/s) and FCH (9.6 kbit/s).

network, the same sequence of messages is used to set up a standard call from a mobile station as that for a packet-data call. There are essentially two requirements for establishing CDMA2000 packet-data calls from a mobile station:

- the allocation of radio resources; and
- the establishment of a PDSN link and PPP session (Figure 4).

After Janice dials the appropriate number and presses send, the mobile station sends (1) an *origination* message with the required packet-data service option (service option 33 for CDMA2000) to the BSC. The BSC, in turn, sends (2) a *connection management service request* message to the MSC, to authorize a radio traffic channel for the call. If Janice is an authorized user of the network, the MSC responds with (3) an *assignment request* message to the BSC, instructing it to allocate the resources needed for the call. Once the radio traffic channel has been established, the BSC sends (4) an *assignment complete* message to the MSC. Janice's mobile station is now authenticated and has the air interface resources it needs for the e-mail session.

The BSC can then generate and send (5) an *A9-setup-A8* message to the PCF, which finishes setting up the data session with the PDSN and responds to the BSC with (6) an *A9-connect-A8* message. If the PCF fails to set up the session, it sends an *A9-release-A8* message. To initiate set-up of an A10 connection to the PDSN, the BSC/PCF sends (7) an *A11 registration request* message to the selected PDSN, which validates the request and, if the request is acceptable, returns (8) a *registration reply* message to indicate acceptance. With the A10 connection in place, link layer and network layer frames pass over the connection in both directions. A point-to-point protocol (PPP) connection (9) is now in place between the mobile station and the PDSN, and Janice can begin communicating over the Internet—for example, by authenticating herself to an Internet service provider. Note: For packet-data calls, a circuit is never allocated between the MSC and BSC, since the data packets are routed directly from the BSC to the PDSN.

Data rate allocation and de-allocation

After the PPP session has been established between the mobile station and PDSN, if Janice has data to send or receive, Ericsson's CDMA2000 radio access network sets up a radio link for the packet-data call and sets up the fundamental and supplemental channels with appropriate data rates (Table 2). To better understand how a peak over-the-air data rate of up to 163.2 kbit/s is allocated to the end-user in the forward and reverse directions, we will now describe the events that trigger the allocation of the fundamental and supplemental channels, and the factors that determine the maximum allowable data rate associated with the supplemental channel.

TABLE 2, CDMA2000 1X PACKET-DATA RATES (SO33)

F-FCH/ R-FCH	F-SCH/ R-SCH	Peak over-the-air data rate (kbit/s)
9.6	0.0	9.6
9.6	19.2	28.8
9.6	38.4	48.0
9.6	76.8	86.4
9.6	153.6	163.2

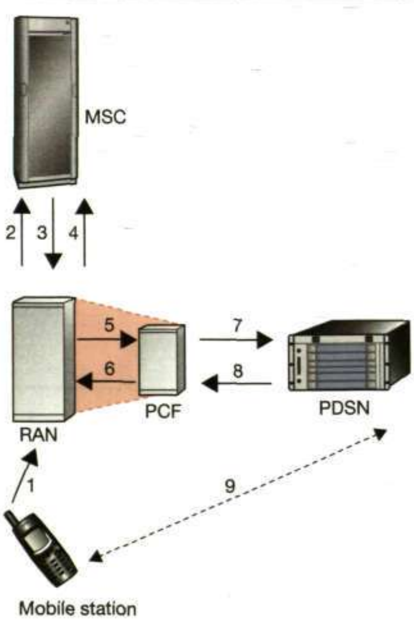


Figure 4
Set-up of a packet-data call.
1 Origination message
2 CM service request message
3 Assignment request
4 Assignment complete
5 A9 - setup - A8 message
6 A9 - connect - A8 message
7 Registration request
8 Registration reply
9 PPP session established

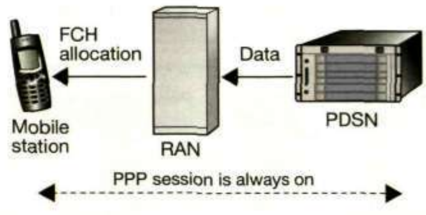


Figure 5
Allocation of the fundamental channel.

Accessing the FCH and SCH

When there is data to send, the mobile station or network initiates a packet-data session, and a fundamental channel is set up (Figure 5).

The radio access network checks the packet sizes and packet buffer sizes for packets to the mobile station, to determine whether or not it needs to set up the supplemental channel (Figure 6)—that is, it will not set up the supplemental channel unless packet size or packet buffer size exceeds a given threshold or unless the mobile station has requested a supplementary channel.

Factors that determine the maximum allowable SCH data rate

If the amount of data coming in to the radio access network from the PDSN justifies the use of a supplemental channel, the BSC sets up this channel at the highest data rate allowed or the highest data rate for which resources are available, whichever is less (Figure 6). If the operator permits the network to allocate, say, 16x SCH, and resources are available in the cell, then the BSC sets up a supplemental channel with peak over-the-air data rate (including overhead) of 153.6 kbit/s. A radio admission control algorithm, internal to the BSC, is responsible for allocating the radio resources and Walsh codes that are necessary for packet-data calls.

The maximum data rate of the supplemental channel is determined by the channel conditions of the user. A threshold for satisfactory channel condition is defined for each data rate (2x, 4x, 8x and 16x). The measured channel condition is compared to the required channel condition for each data rate. The highest data rate for which the reported channel condition exceeds the required value becomes the requested data rate for the supplemental channel. To initiate set-up of the supplemental channel, the requested data rate must be greater than 2x, otherwise set-up is aborted. The F-SCH is maintained as long as the channel conditions are satisfactory and as long as enough data is being transmitted.

De-allocation of the SCH

The supplemental channel is de-allocated when the volume of data being sent to the mobile station can be satisfied using the fundamental channel or if the channel conditions become unsatisfactory for the current rate. Unsatisfactory channel conditions for the chosen rate would lead to excessive frame error rates (FER), which wastes radio resources. A throughput-related metric is used to measure use and channel quality, and the supplemental channel is de-allocated if throughput falls below a given threshold (Figure 7).

Optimized sector selection for the F-SCH

Ericsson's implementation of the high-speed packet-data allocation of the F-SCH is optimized for high throughput and uses only one radio link to the best possible sector. Instead of using soft handover for the F-SCH, which requires additional and unnecessary radio resources, the best pilot is selected from the FCH active set (sector selection). Analysis of the F-SCH air link and radio link protocol has shown that sector selection works better than soft handover for the F-SCH. Note, however, that soft handover is always used when setting up the fundamental channel, since the signaling sent on the fundamental channel has more stringent channel quality requirements than the F-SCH. Moreover, because radio resources on the reverse link are not as sparse as on the forward link, soft handover is used to improve network reliability on the reverse-link supplemental channel (R–SCH).

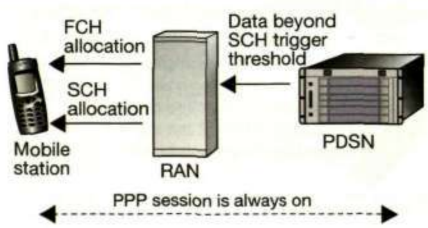


Figure 6
Allocation of the supplemental channel.

Mobility management

If Janice is on the go while transmitting and receiving data, the connection of her mobile station to the network will have to be managed as the mobile station moves from cell to cell—much the same as is done for circuit-switched services. Moreover, as the mobile station moves, the radio link and connection to the packet core network must be managed regardless of whether the mobile station is in an active or dormant state.

Types of handover

Ericsson's CDMA2000 radio access network supports different types of packet-data call handover (Figure 8). A handover can occur between two cells within the same BSC (intra-BSC) or between two cells belonging to different BSCs (inter-BSC). Handovers can be between cells using different frequencies (hard handover) or between cells that use the same frequency (soft handover). Handovers can also occur between sectors of the same cell and frequency (softer handover).

Packet-data handovers: impact on FCH, F-SCH, R-SCH and PPP session

Below we discuss the impact of packet-data handovers on the fundamental channel (forward and reverse), F-SCH, R-SCH and the PPP session established between the mobile station and the PDSN.

Many types of packet-data handover resemble those for voice service. The following discussion focuses on handover scenarios that differ from those for voice.

Intra-BSC F-SCH sector selection

When the F-SCH serving sector is no longer the best sector for the F-SCH, the BSC selects another sector to serve the F-SCH (sector selection).

- The serving BSC and PDSN remain the same.
- There is no change in the PPP session.
- The fundamental channel allocated to the packet-data call is undisturbed.
- The F-SCH allocated to the packet-data call is de-allocated from the current serving sector and re-allocated to a new serving sector.

Intra-BSC hard handover of the FCH and F-SCH/R-SCH

For hard handover of a packet-data call in the serving BSC (intra-BSC hard handover)

- there is no change in the PPP session;
- the F-SCH or R-SCH is de-allocated as a result of the hard handover in the BSC;
- the fundamental channel allocated to the packet-data call is de-allocated and a new fundamental channel is allocated to the call on the new CDMA channel; and
- a new F-SCH or R-SCH can be re-allocated by the BSC after the handover is complete.

Inter-BSC hard handover of the FCH

For hard handover between BSCs, the source and target BSCs may be connected to the same PDSN or to different PDSNs. If the BSCs are connected to different PDSNs, the

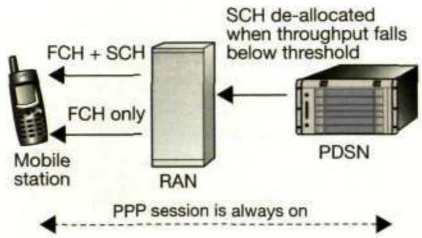
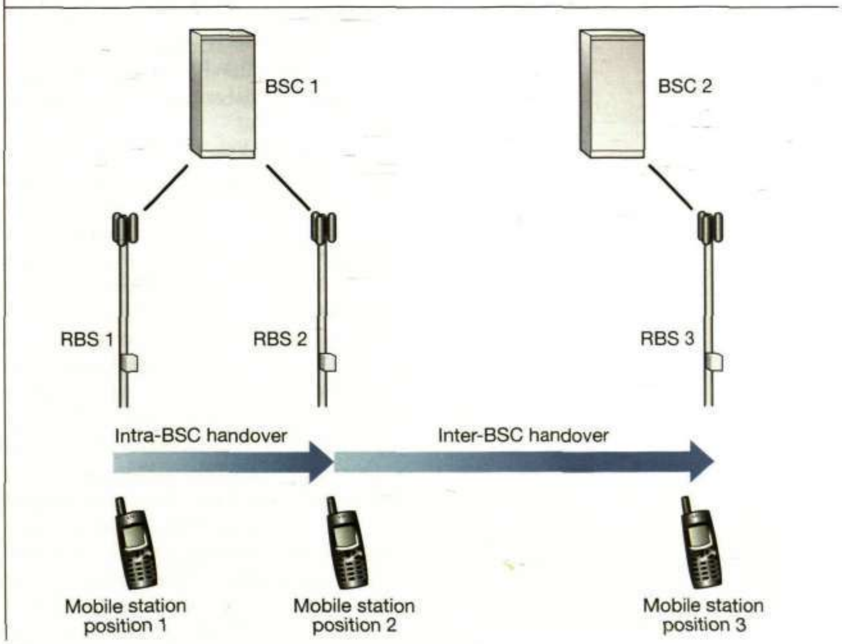


Figure 7
De-allocation of the supplemental channel.

Figure 8
Packet-data handovers.



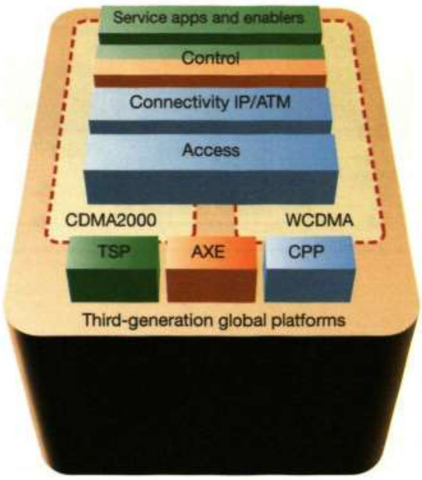


Figure 9
Use of global third-generation platforms.

mobile station can continue to use the same IP address, provided the mobile station supports the Mobile IP (MIP). When the mobile station moves between different PDSNs, the packets will be forwarded or tunneled by elements within the packet core network (PCN) to the PDSN that is currently serving the mobile station. Therefore, for inter-PDSN (with MIP) or intra-PDSN inter-BSC hard handover

- the mobile station continues to use the same IP address;
- the fundamental channel allocated to the packet-data call is de-allocated and a new fundamental channel is allocated;
- the F-SCH or R-SCH is de-allocated as a result of the inter-BSC hard handover;
- the target BSC becomes the new anchor for the call; and
- a new F-SCH or R-SCH can be re-allocated by the target BSC after handover is complete.

PDSN selection

Each PCF, which can be connected to multiple PDSNs, uses a PDSN selection algorithm (as specified in the interoperability specification) to balance the load handled by the PDSN when a mobile station moves from one PCF to another. The PCF is configured with the address of every PDSN that is connected to it. The PCF uses the international mobile subscriber identity (IMSI) of the mobile station to hash among the configured PDSN addresses and select a PDSN for the mobile station. This ensures that mobile stations are distributed over all PDSNs connected to a particular PCF. When a mobile station moves between PCFs, the PDSN selection algorithm ensures that the mobile station remains connected to the PDSN during handover to another PCF. This approach maintains the connection between the mobile station and PDSN and allows the mobile station to maintain its original IP address.

Dormancy

If Janice stops sending and receiving data, her mobile station will transition from an active to a dormant state. As explained above, dormancy is the state of a packet-data call in which the air link has been torn down due to inactivity. This is done to conserve radio link resources. The PPP session between the mobile station and PDSN is maintained (always on) throughout dormancy to ensure rapid reconnection when there is data to transmit.

Conditions for dormancy

The transition from active to dormant state can be initiated by either the mobile station or the BSC. A mobile station can initiate the transition by instructing the BSC to disconnect the traffic channel through a *release order* message. Similarly, if the inactivity timer at the BSC expires for a given packet-data session, the BSC can initiate a transition of the mobile station to a dormant state by issuing a *release order* message to disconnect the traffic channel for the mobile station.

Handovers in dormant mode

Handovers for a mobile station in dormant mode are handled by the network. In the dormant state, there are no fundamental or supplementary channels allocated to the user, but the PPP session between the mobile station and the PDSN is maintained through the PDSN selection algorithm as the user moves between PCFs.

Dormant-to-active state transition

The transition from a dormant to an active state can be initiated by either the mobile station or the BSC. The mobile station can initiate the transition by sending an *origination* message with the packet-data service option. Likewise, the BSC can initiate the transition of a mobile station from dormant to active state by paging it when the BSC receives data addressed to it from the PDSN. The PPP session does not need to be re-established since it is always on. In all other respects, this call is no different from a regular A8 connection set-up procedure. When there is data to send, the fundamental and supplemental channels are allocated by the network in the manner described above.

Ericsson's implementation

Ericsson is the only wireless vendor with a focused, well-aligned strategy for third-generation wireless networks. Ericsson uses the same third-generation platforms in its WCDMA and CDMA2000 solutions. As a consequence, Ericsson can focus its development resources on improved time to market and enhancements to the product line itself. Figure 9 highlights Ericsson's global platform strategy.

Ericsson's implementation of CDMA2000 1X is based on the connectivity packet platform (CPP, formerly called Cello packet platform) in the radio access network (Figure 10). CPP is used in the Ericsson BSC1120, RBS1127 and RBS1130 as well as in packet core network nodes, such as the PDSN and the home agent (HA). CPP is made up of hardware and software modules that use a cell switch to interconnect processors on different types of device, offering a flexible and scalable operating platform for network products. The physical infrastructure consists of a 19-inch subrack with a cell-switching capacity of 16 Gbit/s and contains clustered processors and device boards with scalable capacity and robustness.

CPP uses the ATM/IP transport system and supports STM-1, OC-3 and other standard physical interfaces. ATM adaptation layer type 2 (AAL2) functionality for signaling and multiplexing, as well as real-time IP functionality, are built into the platform. Likewise, the platform contains an accurate system clock function that can be synchronized to an external source or to any of the line interfaces. Ericsson's CPP is

- future-proof—this provides operators with an economical migration path to future-generation technologies, such as 1xEV;
- modular and flexible—modularity makes it easy to create nodes and products with various configurations and different capacity, reliability, and performance levels. CPP is very flexible in handling the challenges that will be brought upon by 1xEV, such as additional packet data and the mix of voice and data;
- robust—CPP uses a multiprocessor control system on a commercial processor and a real-time operating system with increased robustness for telecommunications-class availability. It uses ATM technology for node communications and switching, ensuring carrier-class quality throughout the radio access network; and
- ready for migration to all-IP systems—CPP is well prepared for easy migration to IP transport when the radio access network migrates to an all-IP network.

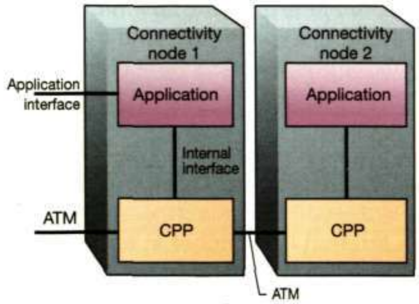


Figure 10
Ericsson's connectivity packet platform (CPP).

Conclusion

CDMA2000 is an approved third-generation mobile standard with support for high-speed, always-on, packet-data services. CDMA2000 1X is backward compatible with cdmaOne, provides increased voice capacity and supports packet-data services of 144 kbit/s (peak data rate of up to 163.2 kbit/s).

CDMA2000 1xEV-DO supports data-only services on a single CDMA carrier with best-effort packet data at a peak rate of 2.4 Mbit/s.

CDMA2000 1xEV-DV supports voice and real-time data services on a single CDMA carrier at a peak rate of 3.1 Mbit/s.

Ericsson's CDMA2000 radio access network solution, which is based on the company's global third-generation carrier-class technology (CPP) uses standardized, open interfaces to connect to the packet core network to deliver data services. Through always-on and dormancy features, Ericsson's CDMA2000 radio access network efficiently allocates high data rates to end-users while preserving valuable radio link resources. It has been designed with the future in mind, supporting CDMA2000 1X, today, and providing operators with investment protection and smooth migration to higher packet data rates when CDMA2000 1xEV is introduced.

TRADEMARKS

CDMA2000 is a trademark of the Telecommunications Industry Association (TIA).

cdmaOne is a registered trademark of the CDMA Development Group (CDG).

Ericsson Telecom Server Platform 4

Victor Ferraro-Esparza, Michael Gudmandsen and Kristofer Olsson

The marriage of telecommunications and the Internet puts new requirements on equipment. Customers have come to expect the same quality of service as they get from present-day telecommunications networks. At the same time, they expect new services in the multimedia and services domain.

We know from experience that modern telecommunications systems are extremely reliable and provide real-time responses. This level of reliability and response cannot currently be achieved using technologies in the Internet domain. On the other hand, the Internet is far richer in terms of content, where pictures (still and moving) are part of today's experience.

In an ideal world we would reuse technologies from the data communications industry and combine them with those from telecommunications systems. With the Telecom Server Platform 4, Ericsson has taken a giant step in this direction, combining its know-how and experience of telecommunications, reliable systems, and scalable systems with open technologies, such as Linux, CORBA, SNMP, LDAP and other standards, such as cPCI and *de facto* standards. The result is a carrier-class server which is always on, scalable, and open, and which adds value for Ericsson's customers.

Ericsson Review no. 1, 2002 described the Ericsson IP Multimedia solution and the role the Telecom Server Platform plays in this domain. The Telecom Server Platform also plays an important role in delivering services. In this article, the authors describe the Ericsson Telecom Server Platform 4 in greater detail. They also briefly describe the service network framework and give some examples of applications in the services domain.

Product overview

The Ericsson Telecom Server Platform 4 is more robust and fault-tolerant than any comparable server technology. It is extremely reliable with unique, linearly scalable capacity, and real-time characteristics, which means that transmission takes place with minimal and controlled delay.

The telecommunications-grade software is enabled by TelORB clusterware, which runs on top of DICOS and Linux. The software incorporates the very latest in signaling system no. 7 (SS7) and built-in node management, with support for the many protocols needed for interoperability between the Telecom Server Platform and operations support systems (OSS).

In its make-up, the Telecom Server Platform combines Ericsson's long tradition of designing robust and reliable hardware with commercially available components. The result is an open hardware architecture with excellent characteristics: scalable capacity, telecommunications-grade reliability, small footprint and minimal power consumption.

Hardware

The Telecom Server Platform uses commercially available hardware with ample capacity and dependable node performance. This lowers the cost of installation, operation, and maintenance.

The hardware fulfills all telecommunications requirements regarding power consumption, low electromagnetic radiation, reliable equipment practice, and reliable connectors. This ensures trouble-free operation. What is more, the modular design facilitates upgrading.

Some key features of the hardware are redundancy in all hardware components, "hot-swap" hardware replacement, a modular platform for maximum flexibility, standard

BOX A, TERMS AND ABBREVIATIONS

3GPP	Third-generation Partnership Project
AAA	Authentication, authorization and accounting
API	Application program interface
ASP	Application service provider
CAMEL	Customized applications for mobile network-enhanced logic
CAP	CAMEL application protocol
CM	Configuration management
CORBA	Common object request broker architecture
cPCI	Compact PCI
CPI	Customer product information
DICOS	Object-oriented operating system with excellent real-time characteristics. Developed by Ericsson.
E1/T1	PDH transmission frame formats for 2 Mbit/s (E1) alt. 1.5 Mbit/s (T1) transmission rates
FM	Fault management
FPGA	Field-programmable gate array
GEM	Generic Ericsson magazine
GESB	Gigabit Ethernet switch board
GUI	Graphical user interface
HTTP	Hypertext transfer protocol
IIOP	Internet inter-ORB protocol
I/O	Input/output
IPC	Inter-processor communication
IPv4	Internet protocol version 4
ISR	In-system reconfigureable FPGA
J2EE	Java 2 Enterprise Edition
LDAP	Lightweight directory access protocol
MAP	Mobile application part
MIP	Mobile IP
NAS	Network access server
O&M	Operation and maintenance
ORB	Object request broker
OSA	Open system architecture
OSS	Operations support system
PCI	Peripheral computer interconnect
PM	Performance management
RADIUS	Remote access dial-in user service
RAM	Random access memory
RMI	Remote method invocation
RPC	Remote procedure call
SCB	Support & connection board
SCS	Service capability server
SDRAM	Synchronous dynamic RAM
SIGTRAN	Signaling transport
SNF	Service network framework
SNMP	Simple network management protocol
SOAP	Simple object access protocol
SS7	Signaling system no. 7
SSL	Secure socket layer
TCP	Transmission control protocol
TMN	Telecommunications management network
UDDI	Universal description, discovery and integration
UDP	User datagram protocol
W3C	World Wide Web Consortium
WSDL	Web services description language
XML	Extensible markup language

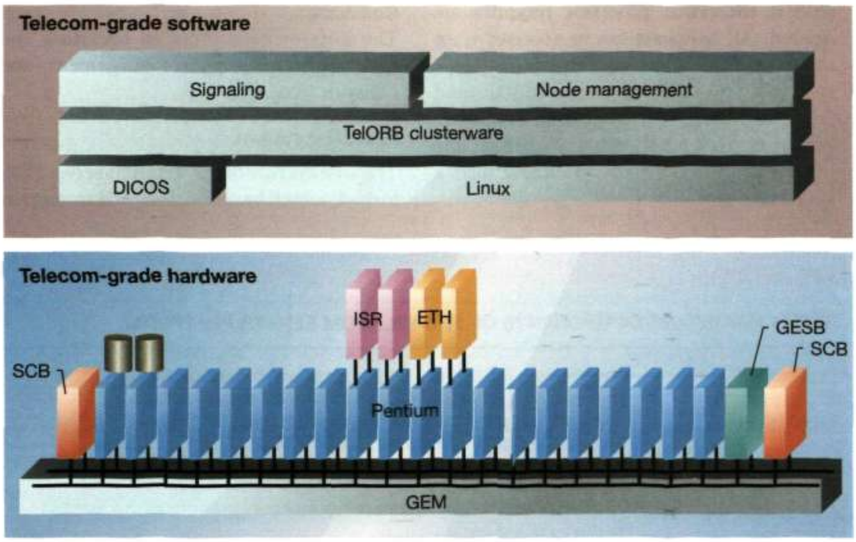


Figure 1
The Telecom Server Platform consists of telecom-grade hardware and software.

hardware interface boards, industry-standard components (creating a future-proof architecture), and optional geographical node redundancy for additional dependability.

The Telecom Server Platform 4 can be duplicated in a redundant standby node. That is, in case of failure, the load is automatically redistributed within the node or, in the case of geographical node redundancy, a remote standby node can take over the functions of the primary node.

The ability to hot-swap hardware allows operators to replace any component at any time without affecting system performance. In a market of converging technologies and standards, the use of industry-standard components creates an architecture that is future-proof, scalable, flexible and cost-effective.

The hardware platform includes processor modules (with or without peripherals), signaling processors, Ethernet switches, power supplies and fans. All processors are standard, commercial, off-the-shelf, single-board cPCI computers (Box B).

The unit is housed in an Ericsson cabinet (BYB 501). An expansion cabinet is avail-

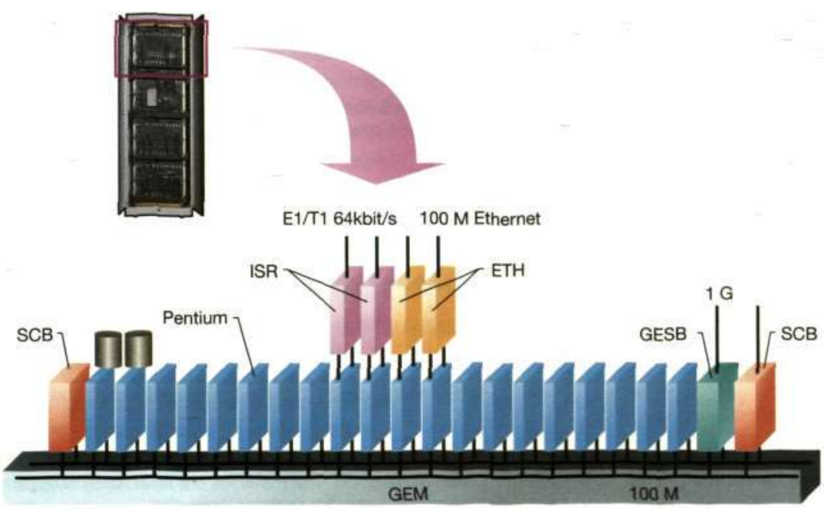


Figure 2
Each subrack consists of $n+1$ duplicated processors connected via duplicated 100 Mbit/s Ethernet. The subracks are interconnected via 1 Gbit/s Ethernet. Every component is at least duplicated.

able if additional processor modules are needed. All hardware can be accessed from the front panel, ensuring easy maintenance and replacement. To ensure high quality and efficient installation, all external cables are connected to a patch panel. Together, the main cabinet and expansion cabinet form a single node.

Software

The software consists of an operating system, clusterware, node management, and network signaling.

Operating system

The architecture of the Telecom Server Platform 4, which has been designed to support

BOX B, HARDWARE COMPONENTS OF THE TELECOM SERVER PLATFORM

Processors	Intel Pentium III 700 MHz cPCI boards with 1 GB SDRAM.
E1/T1/J1 ports	To execute scalable SS7/ITU-T/China/ANSI protocol stacks.
Ethernet ports	100 Mbit/s Ethernet connections used for external IP communication, such as the signaling transport (SIGTRAN) protocol.
Input/output devices	DVD: one drive for input purposes, including initial loading. Tape drive: one 20 GB drive, for example, for backups.
Hard drives	Up to 14 x 18.2 GB storage capacity for every executable unit needed for start-up and backup.
Patch panel	A jack frame to which the cables in the cabinet are terminated, and to which all external cabling is connected.
Alarm collector	Collects alarms from fans and supply voltage supervision at level-1 switches in the magazines.
GEM	Generic Ericsson magazine (GEM)—a standard subrack that provides a cost-effective solution with small footprint. Each GEM has a 100 Mbit/s Ethernet level-1 switch that connects all processor boards. A 1 Gbit/s Ethernet level-2 switch interconnects the magazines. The switches are duplicated for redundancy.
Cabinet	The Ericsson BYB 501 uses forced-air ventilation, allowing heat to dissipate through the bottom and out at the top of the cabinet. Ericsson expansion cabinets are available.
Power supply	Input to the cabinet is -48 volts.

Capacities	**Maxi**	**Midi**	**Mini**	**Micro**
E1/T1 connections	24	16	16	8
ITU No. 7/SS7 signaling channels	192(E1)/144(T1)	128(E1)/96(T1)	128(E1)/96(T1)	64
Ethernet ports	32	28	22	12
Processors	42	31	21	10
RAM data storage	42 GB	31 GB	21 GB	10 GB
DVD	1	1	1	0
Tape streamer (20GB)	1	1	1	0

Environmental requirements

Recommended minimum ceiling height:	210 cm
Relative humidity (min-max):	20-80%
Temperature (min-max):	+5° to 40°C*
Temperature (normal operation):	+20°C

* The tape streamer/tapes are a limiting factor when it comes to average temperature and humidity range. If the temperature reaches +35°C, the tape streamer tapes stretch, which leads to loss of data.

Agency approvals

The hardware has been designed to comply with NEBS level 3.
Seismic vibration, EN 300 019-2-3 and GR-63-CORE zone 4.
EMC - EN 300 386 class B,
Part 15, Subpart B, Class B/Federal Communications Commission (FCC) according to GR-1089-CORE.
Product safety - EN 609 50, IEC 609 50 and ANSI/UL 1950, third edition.

Design for environment

The Telecom Server Platform complies with Ericsson's policy to avoid the use of banned and restricted substances.

End-of-life treatment

Ericsson offers recycling services for old Ericsson products. The materials are taken care of by approved recycling companies in compliance with EU or other national legislation.

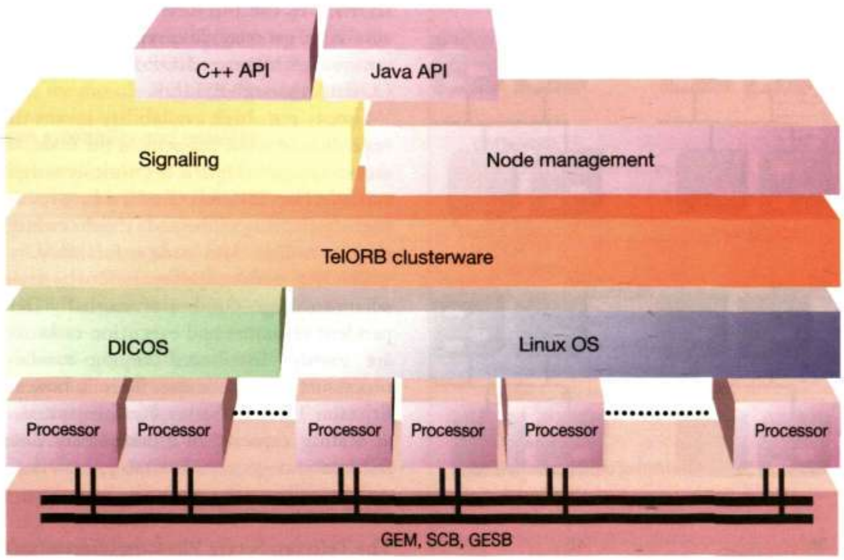


Figure 3
Software architecture of the Telecom Server platform. The TelORB clusterware also controls the Linux processors in the cluster.

different commercial hardware and multiple operating systems, provides the most appropriate executing environment for processes. The current version of the Telecom Server Platform uses the Linux operating system for UNIX-like performance, and DICOS for real-time, mission-critical tasks.

DICOS, which is based on queuing theory, offers soft real-time response. This kind of real-time performance is suitable for telecommunications and data communications applications, especially for database clusters.

TelORB clusterware

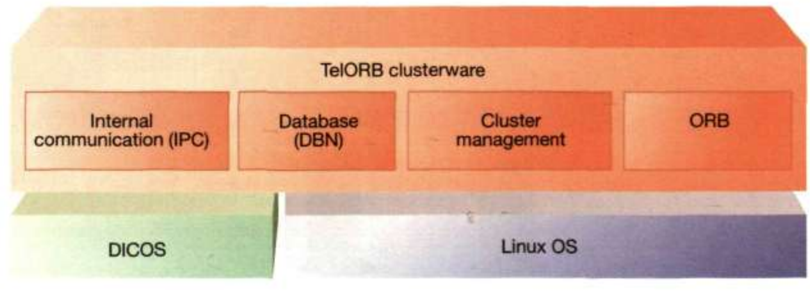
The communication layer of the TelORB clusterware connects the different processors to each other to enable inter-processor communication (IPC). Internally, the TelORB clusterware uses the lightweight protocol IPC to manage duplicated Ethernet backbones. When communicating to the external world, it uses the transmission control protocol (TCP), user datagram protocol (UDP), and IPv4.

A built-in object-oriented database provides persistent storage in RAM. The database is thus always held in primary memory and distributed over the processors. All data is replicated and stored on more than one processor. Should a subroutine or processor fail, the entire database remains available to the software.

The software management layer automatically configures the executing software to ensure that it runs efficiently on any of the processors available to the TelORB cluster. It provides support for software upgrades and monitors and manages software components to ensure high availability. It also supports binaries from third-party vendors. The node-management function includes an element manager that configures and manages every managed object of the TelORB clusterware and operating system.

An object request broker based on the common object request broker architecture (CORBA) has been incorporated to allow the TelORB system to communicate with other systems and the graphical user interfaces used to manage them. It supports

Figure 4
The architecture of the TelORB clusterware.



IIOP1.1 to the Internet inter-ORB protocol (IIOP) gateway, Java remote method invocation (RMI) over IIOP/IPC, and secure CORBA through SSLIOP.

Simply put, high availability means that regardless of what happens to the node, the services supplied by the network are not disturbed. The TelORB clusterware provides high availability to the node thanks to inter-node as well as intra-node redundancy.

TelORB enables applications to be divided into a large number of mutually independent resources and execution tasks that are evenly distributed among available processors in the cluster. This is how the Ericsson Telecom Server Platform succeeds at scaling capacity in a completely linear fashion.

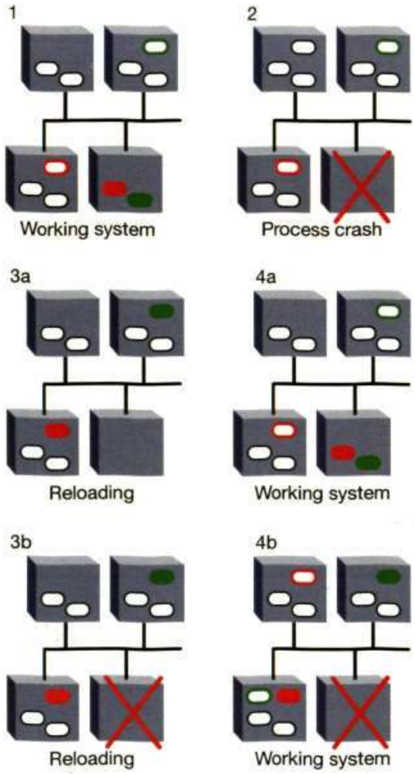


Figure 5
The TelORB clusterware handles hardware and software faults. All data is duplicated (1 & 2).
When a software fault occurs, the processor is restarted and the processes and data are restored (3a & 4a).
When a hardware fault occurs, the processor is taken out of service and new duplicates of the data are made in the cluster (3b & 4b).

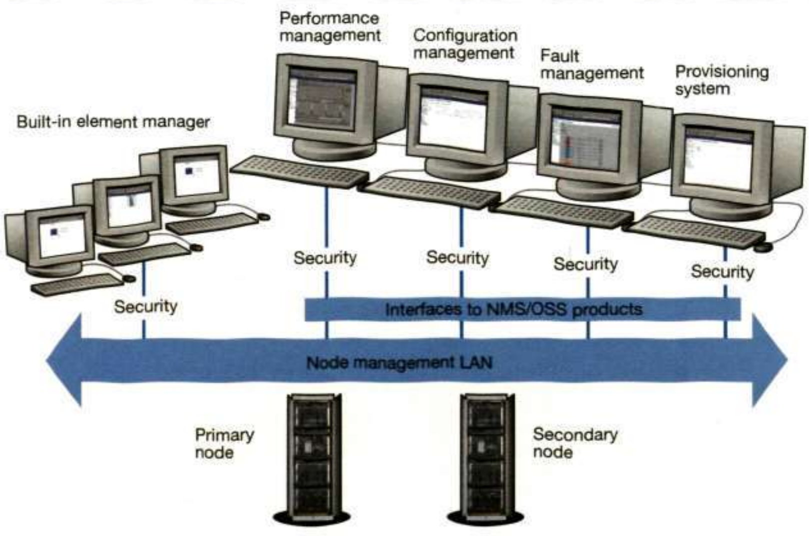
Node management

The Telecom Server Platform offers a node-management solution based on the telecommunications management network (TMN) model. To better align with operator requirements and industry trends, the node-management system also incorporates other standards, such as CORBA, lightweight directory access protocol (LDAP), hypertext transfer protocol (HTTP) and the simple network management protocol (SNMP).

The node-management system implements a manager-to-agent architecture that integrates the node into external network-management and customer-administration systems using standard protocols to communicate with external systems.

In the Telecom Server Platform 4, the node-management function incorporates fault management (FM), configuration management (CM), provisioning support, performance management (PM), logging service (and log querying), an XML exporter, an element manager, and license management.

Figure 6
Node management in the Telecom Server Platform.



Network signaling

The signaling solution of the Telecom Server Platform can handle SS7 and Internet protocols. The SS7 protocol is still the most commonly used protocol for data intercommunication in telecommunications networks. The SS7 signaling stack of the Telecom Server Platform 4 is the latest in SS7 technology and features a scalable implementation that can be distributed over the processors used for traffic handling. The design includes a front-end process for terminating the E1/T1 interfaces, and a back-end process for the upper layers of the stack.

Operator benefits

The Telecom Server Platform is the foundation for several revenue-generating applications for mobile and wire-line operators.

High availability and reliability

High availability is a must in today's modern networks. The Telecom Server Platform achieves high availability through the TelORB clusterware and highly reliable hardware. The main characteristics of the TelORB clusterware can be summarized as follows:

- always-on operation;
- automatic software recovery;
- data replication;
- overload control;
- software updates during operation;
- upgrade of operating system allowed during operation;
- online backup;
- hot-swap hardware replacement; and
- optional geographical node redundancy.

Scalability

Besides high network availability, operators also want scalable network nodes. The Telecom Server Platform has been designed with scalability in mind, ensuring that the initial investment suits customer needs, and that the system can grow with the business. Also, because it is built on commercial hardware, the Telecom Server Platform benefits from continuous enhancements in the processor industry. The modular architecture facilitates rapid expansion of capacity when and where needed. This can be achieved through predefined expansion paths between the configured cabinets.

Cost reduction

The Telecom Server Platform uses commercial, off-the-shelf hardware and software. Operators can thus employ best-in-class hardware and software, and benefit from economies of scale. Because it is scalable and supports co-location, the Telecom Server Platform allows operators to add new applications on a single node in accordance with market demand.

The Telecom Server Platform node management uses a Web-based graphical user interface (GUI) that enables operator staff to access all O&M tasks online—on- and off-site (Figure 7). This functionality reduces delays and travel costs, which, in turn, reduces the total cost of ownership.

Compared to other solutions on the market, the Telecom Server Platform is a total-

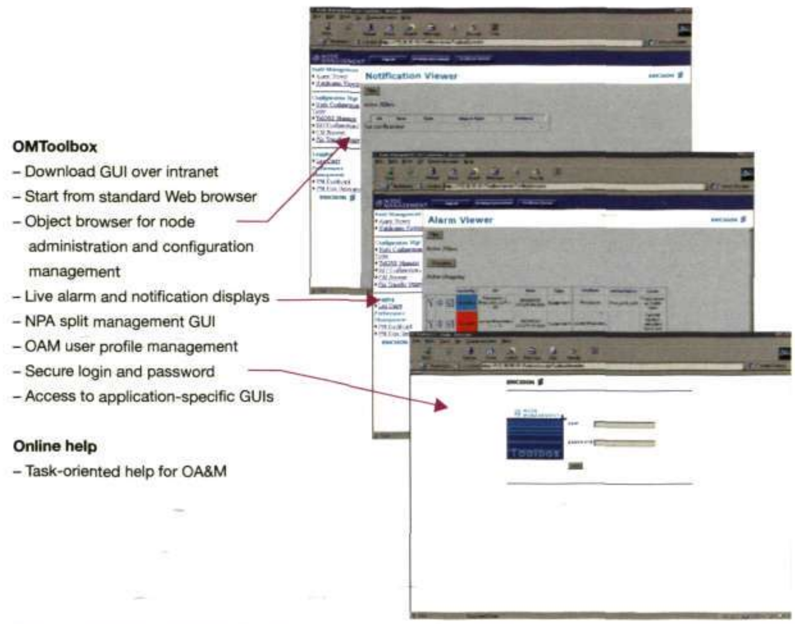
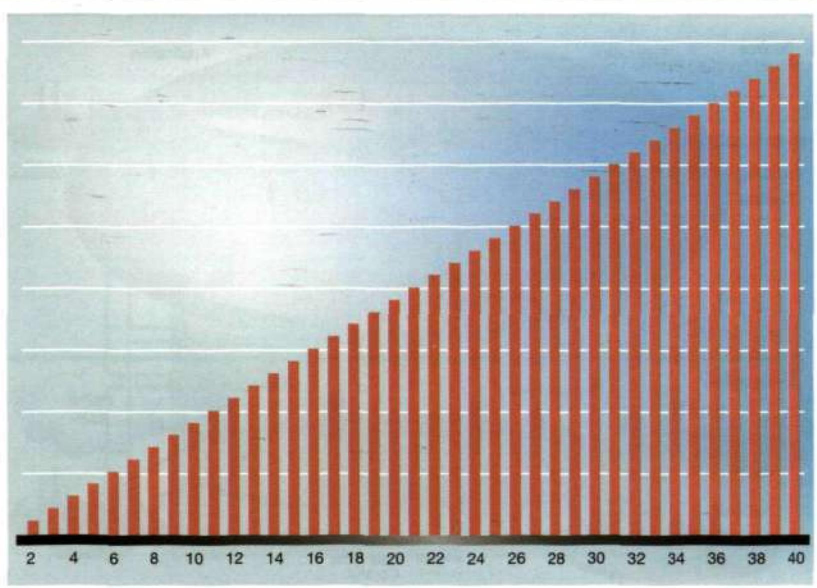


Figure 7
A built-in Web-based node management GUI enables staff to access all O&M tasks online.

Figure 8
The Telecom Server Platform scales linearly. Customers can thus increase capacity as traffic increases.



ly integrated solution in a very compact node with small footprint.

The extensive set of O&M and provisioning protocols makes it easy to integrate the Telecom Server Platform into existing O&M infrastructures.

Service network framework

The service layer represents the top layer in Ericsson's three-tier architectural model (Figure 9). The features and functions offered in the service layer pertain both to end-users and the operator of the network infrastructure. They are often offered in the form of an XML Web service interface.

Architectural and design decisions for products and solutions in the service layer are guided through the service network framework (SNF), which is Ericsson's textbook on how system development can be shifted toward horizontally layered systems. The SNF, which provides standards and guidelines for the system structure, capabilities, and interaction, is an architectural framework that consists of reusable designs for products and solutions in the service layer.

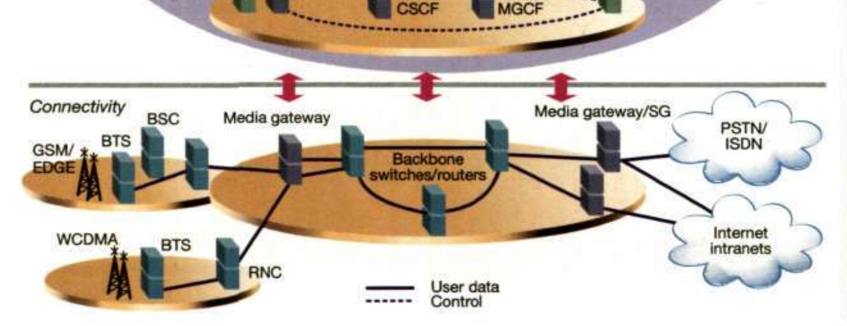
The SNF also forms the foundation for reuse. It has always been difficult to reuse objects from different sources, due to a lack of standards that define how objects should be constructed, what properties they should have, and how they should interact. The SNF provides these standards.

Similarly, the SNF provides specifications (or references to specifications) for common management, provisioning, charging, and shared services. Its service specifications are based on open-standard protocols. Apart from the implied need for IP connectivity, the SNF does not stipulate any specific operating environment.

In addition to the actual specifications, the SNF delivers an architectural framework that is designed to accommodate extensions and grow in the future, yielding a tool for expressing a common architectural direction for the service layer. The SNF is influenced by the following mainstream technologies:

- The application server—considerable Internet application development is conducted according to the Web application paradigm in which the application server deploys the application and provides the run-time and characteristics model.
- The directory server—enterprise and Internet computing domains tend more and more to use LDAP-capable directory servers as repositories of user and resource data. Many products and solutions in these domains interoperate easily with LDAP-capable directory servers.
- The distributed computing paradigm—the industry predominantly employs networks of computers to meet the challenges of scalability and availability. Distributed computing patterns, such as *n*-tier architectures, are used in combination with distributed computing technologies, such as CORBA and COM+, to provide system architectures that lend themselves to distribution over large networks of computers.
- The Java language and the Java 2 Enterprise Edition (J2EE) platform—in recent years we have seen a large gain in server-side Internet and enterprise developer momentum for the Java language with its associated server-side platform specification, J2EE.
- The Web services paradigm—there is a growing trend in distributed computing technology to expose traditional remote procedure call (RPC) services via universally accessible extensible markup language (XML) interfaces using methods standardized by the World Wide Web

Figure 9
Ericsson's horizontally layered network architecture.



Consortium (W3C) for universal description, discovery and integration (UDDI), the Web services description language (WSDL), and the simple object access protocol (SOAP). The widespread availability of Web services will greatly increase the variety of applications that can be created and deployed in a cost-effective way. The paradigm is simple, universally interoperable and pervasive.

Like the SNF, the operating environment of the Telecom Server Platform also supports and employs these technologies. The products and solutions that leverage the service network framework of reusable designs take a decisive step toward becoming highly scalable, manageable, standards-oriented, open and interoperable, secure and modular. By default, these services are provided in the operating environment of the Telecom Server Platform and are thus facilitated in products and services designed for the service layer.

Collections of products and solutions (systems) that use the technologies and provide the qualities defined above can be deployed to create services networks. Services that target operators and end-users are offered by the services of individual systems. This means that the size of a service network (in terms of the number of constituent systems) can vary and is related to the needs of the business it serves and the nature of the surrounding technical environment in which it is deployed.

Benefits to applications

Ericsson AAA Server

The Ericsson AAA Server provides the authentication, authorization, and accounting (AAA) functions that network operators and service providers need to provide Internet access. The key roles of the server are

- authentication—verification of the identity of an entity (user or application) prevents unauthorized actions and use of resources and services;
- authorization—end-users gain access to the network services and resources that match their profiles; and
- accounting—accounting data is collected and consolidated at the end of an IP session, thereby enabling the service provider to charge end-users on virtually any basis (connection time, amount of data transferred, services accessed, and so on). Billing records can be provided and distributed in customized and flexible formats.

The Ericsson AAA Server provides all the basic AAA functions defined in the RADIUS standard (RFC 2865, 2866). It also supports the DIAMETER protocol, which consists of the base protocol and certain specific DIAMETER applications, such as extensions for network access servers (NAS), Mobile-IP (MIP), strong security, and resource management.

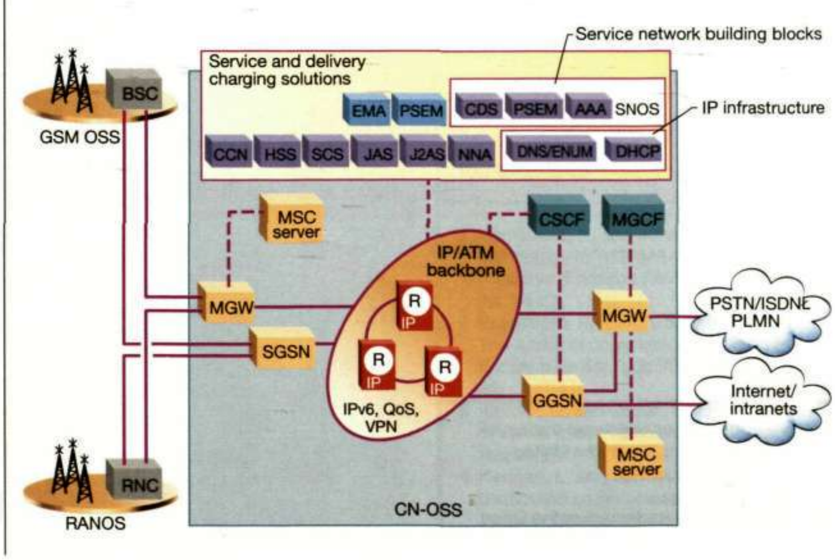
For services to work properly, the Ericsson AAA Server must be available at all times. This puts stringent requirements on the reliability of hardware and software. Lab trials based on a typical traffic model show that the Ericsson AAA Server easily scales to support from 250,000 to more than 5 million subscribers. Moreover, at 50% load it performs more than 2,000 transactions per second.

The inherent characteristics of the Telecom Server Platform are important for the Ericsson AAA Server: it enables the Ericsson AAA Server to scale from small to large systems, gives real-time responses to requests, and is always available.

Ericsson SCS

One enabler in the service network is the Ericsson Service Capability Server (SCS), which uses the capabilities of the Telecom Server Platform for real-time execution of services, open interfaces and scalability. The

Figure 10
Examples of applications that use the Telecom Server Platform in the core network.



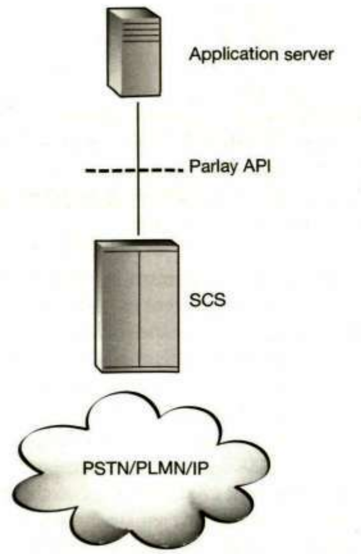


Figure 11
Application service providers (ASP) access the network via the service capability server (SCS) using the standardized Parlay interface.

aims of the SCS are to encourage the development of innovative services and shorten time to market for new applications. It provides open application program interfaces (API) for call control, user interaction and user status and location according to the Parlay/3GPP-OSA standards. Applications and application servers in the network can thus access network service resources regardless of platform and underlying network technology.

The Ericsson SCS is a Parlay gateway that gives various application services transparent access to network capabilities. It gives applications designed according to the Parlay API specification access to network resources in a controlled and secure way. It also functions as a firewall, ensuring that access to the telecommunications network is securely managed. This means that new players, such as application service providers (ASP) and virtual operators, can enter the market and offer services on top of the telecommunications network as a complement to the services of the network operator.

The Ericsson SCS map translates Parlay commands to different network protocols (such as CS1, CS1+, CAPv2, and MAP). In general, the Parlay API is independent of the underlying network. The vendor of the Parlay gateway decides which network protocols will be implemented.

One important role of Parlay is to hide the network complexity from the designer. Provided he complies with the Parlay specification, the designer should not be required to have in-depth knowledge of the underlying networks. Since Parlay is independent of the network and the programming language, the entire market for application development can instead concentrate on developing services. This means that the applications can be developed more quickly

TRADEMARKS

and at less cost. It also means that more innovative applications can be released to the market, helping operators to enrich their service offering, reduce churn, attract more subscribers and increase the use of airtime and revenues.

Ericsson supports application developers through its third-party program in Ericsson Mobility World. This support includes a website with a simulator, and the certification of applications. Where Parlay/OSA is concerned, Ericsson is currently partnering with Solomio, Appium and Wirenix, plus a number of associates.

A customer demonstration center strengthens and supports the sales process by showing that the Ericsson SCS and third-party applications actually work on the network. The customer demonstration center is also used to verify and test third-party applications.

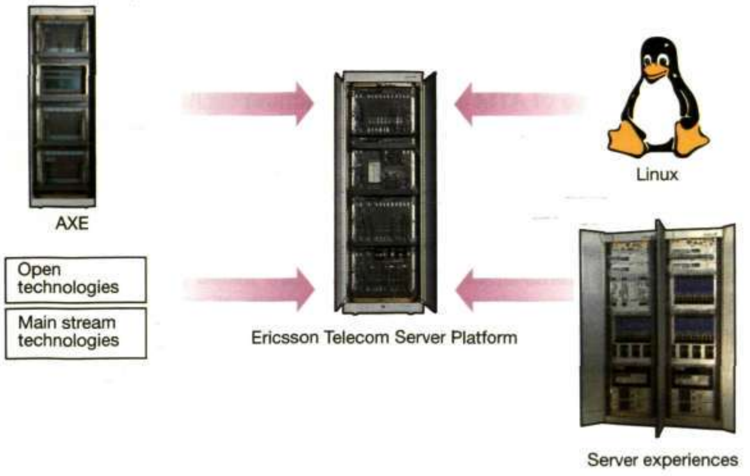


Figure 12
The Ericsson Telecom Server Platform combines the best of the telecommunications world with open Internet technologies.

Conclusion

Ericsson's Telecom Server Platform 4 joins the best of two worlds by integrating IT technologies into a telecommunications-grade server. Many of the applications that use the Telecom Server Platform benefit from its inherent characteristics.

Besides offering the excellent characteristics required for telecommunications systems, the Telecom Server Platform 4 incorporates the latest in open technologies, for example, Linux, which is an important enabler. Ericsson is working actively with Linux standardization bodies, such as the Open Source Development Lab, to make certain that a Linux standard is made available and distributed to the telecommunications industry. The addition of Ericsson's TelORB clusterware to the Linux software gives the system characteristics needed in a telecommunications environment.

REFERENCES

1 Hameleers, H. and Johansson, C.: IP technology in WCDMA/GSM core networks. Ericsson Review Vol. 79(2002):1, pp. 14-27.
2 Granström, P., Olson, S. and Peck, M.: The future of communication using SIP. Ericsson Review Vol. 79(2002):1, pp. 28-35.
3 Ahlforn; G. and Örnulf, E.: Ericsson's family of carrier-class technologies. Ericsson Review Vol. 78(2001):4, pp. 190-195.
4 Hennert, L. and Larruy, A.: TelORB—The distributed communications operating system. Ericsson Review Vol. 76(1999):3, pp. 156-167.

AAL2 switching in the WCDMA radio access network

Bo Karlander, Szilveszter Nádas, Sandor Rácz and Jonas Reinius

New switching technologies are needed in the access network of third-generation mobile networks to provide a cost-effective transmission of different kinds of services, such as AMR-coded voice and large bandwidth data. The use of ATM and AAL2 switching techniques in the traffic concentration nodes can significantly reduce the need for link capacity in the access network. The most important sources of these savings are the statistical fluctuation of the number of AAL2 connections, the fluctuation of the number of users at a base station due to mobility, and the granularity of the ATM virtual channel cell rate. Ericsson's offering of AAL2 switching technology enables operators to maximize these gains in the WCDMA radio access network. Indeed, in large networks, this technology has the potential to triple the capacity of the transmission link.

The authors describe the advantages of AAL2 switching in the WCDMA radio access network. They also include results from a study comparing bandwidth requirements when traffic between node B and the RNC is aggregated either at the ATM layer, by means of switching AAL2 paths using ATM cross-connect, or at the AAL2 layer, by means of dynamically switching AAL2 connections. The authors also describe the general architecture for terminating and switching AAL2 connections in Ericsson's WCDMA RAN node products.

Introduction

Third-generation mobile networks offer a wide variety of services, such as voice, circuit-switched data, and packet-switched data at bit rates ranging from a few kbit/s up to 384 kbit/s (eventually up to 2 Mbit/s). In this environment, highly adaptive transfer and switching methods are needed to deliver different kinds of services in a cost-effective and high-quality fashion. This applies to both the

- radio network layer, which is responsible for transferring data over the WCDMA air interface; and
- transport network layer, which is responsible for transferring data between the nodes (node B and RNC) of the radio access network.

As specified by the Third-generation Partnership Project (3GPP), the transport network layer for the WCDMA radio access network (RAN) is to use ATM transport network technology and protocols. ATM adaptation layer type 2 (AAL2) technology is used for the dominant part of data transfer.

The AAL2 protocol enables several user (radio network layer) connections to be multiplexed flexibly and efficiently on a common ATM virtual channel connection (VCC) between two nodes. The use of AAL2 switching in intermediate nodes has the potential to yield significant statistical multiplexing gains on transmission links that carry aggregated traffic for multiple nodes without loss of control over the quality of service of individual connections.

WCDMA RAN—transport network architecture and protocols

WCDMA RAN

Figure 1 gives a schematic view of a WCDMA network, which consists of user equipment (UE), the WCDMA terrestrial radio access network (WCDMA RAN), and the core network.

The WCDMA RAN handles all tasks that relate to radio access control, such as radio resource management and handover control. The core network, which is the backbone of WCDMA, connects the access network to external networks (PSTN, Internet). The user equipment (mobile terminal or station) is connected to radio base stations (node B) over the WCDMA air interface (*Iu*). During soft handover, one UE can communicate with several node Bs simultaneously.

According to the WCDMA RAN specifications drafted by the 3GPP, all radio network functions and protocols are separate from the functions and protocols in the transport network layer. The transport network layer provides data and signaling bear-

BOX A, TERMS AND ABBREVIATIONS

3GPP	Third-generation Partnership Project	MAC	Medium access control
AAL2	ATM adaptation layer type 2	Node B	Radio base station, RBS
AMR	Adaptive multirate (voice codec)	O&M	Operation and maintenance
ATM	Asynchronous transfer mode	PS 64	Packet-switched data at 64 kbit/s
CAC	Connection admission control (algorithm)	PS 384	Packet-switched data at 384 kbit/s
CBR	Constant bit rate	PSTN	Public switched telephone network
CCH	Common transport channel	Q.2630	ITU AAL2 signaling protocol
CID	Connection identifier	QoS	Quality of service
CN	Core network	RAB	Radio access bearer
CNA	Concentration node area	RAN	Radio access network
CPP	Connectivity packet platform (formerly called Cello packet platform)	RLC	Radio link control
CPS	Common-part sub-layer (packet)	RNC	Radio network controller
CS 64	Circuit-switched data at 64 kbit/s	RXI	CPP-based aggregation node
DCCH	Dedicated control channel	SDH	Synchronous digital hierarchy
DCH	Dedicated transport channel	SL1, 2, 3	Switching level one, two, three
DRNC	Drift RNC	SRNC	Serving RNC
E1	ETSI 2 Mbit/s line interface	STM	Synchronous transfer mode
GoS	Grade of service	TTI	Transmission time interval
Iu	Core network-to-radio network interface (3GPP)	UE	User equipment
Iub	Node B-RNC interface (3GPP)	UP	User plane
Iur	RNC-RNC interface (3GPP)	VC	Virtual channel
LAC	Link admission control	VCC	Virtual channel connection
		VCI	Virtual channel identifier
		VPI	Virtual path identifier
		WCDMA	Wideband code-division multiple access

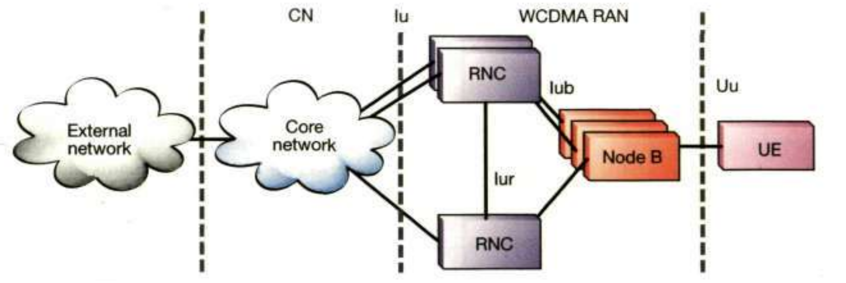


Figure 1
Schematic view of a WCDMA network.

ers for the radio network application protocols between RAN nodes, and includes transport network control-plane functions for establishing and releasing such bearers when instructed to do so by the radio network layer.

The initial WCDMA RAN specifications stipulate that the transport network layer must be based on ATM and AAL2 technology. However, Release 5 of the 3GPP specifications also includes the option of an IP-based transport network.

The focus of this article is on the design of ATM- and AAL2-based transport networks with particular emphasis on the *Iub* interface between the RNC and node B.

AAL2 at the *Iub* interface

Figure 2 shows the ATM and AAL2-based protocol stack at the *Iub* interface for transferring data streams on common transport channels (CCH) and dedicated transport channels (DCH) to the air interface.

The retransmission mechanism of the radio link control (RLC) protocol ensures reliable transmission of loss-sensitive traffic over the air interface. The RLC protocol is used by signaling radio bearers and by radio bearers for packet-switched data services, but not by radio bearers for circuit-switched services.

The medium access control (MAC) protocol forms sets of transport blocks in the air interface and schedules them according to the timing requirements of WCDMA. Each scheduled period, called a transmission time interval (TTI), is 10 ms in length or multiples thereof.

WCDMA radio connections, or radio access bearers (RAB), have bit rate values between 8 and 384 kbit/s. The size of the MAC transport block sets and length of the TTI are RAB-specific.

For data transfer over the *Iub* interface, the MAC transport block sets are encapsulated into *Iub* frames according to the *Iub* user-plane (UP) protocol for CCH or DCH data streams. Each *Iub* user-plane data stream needs a separate transport network connection between the RNC and node B. The transport network thus establishes one AAL2 connection for each data stream. In Figure 2, the AAL2 switch (optional) is used for building aggregating transport networks.

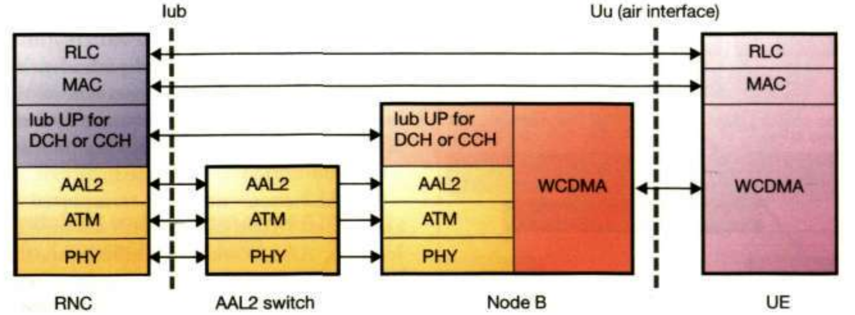


Figure 2
User-plane data transfer between the RNC and node B.

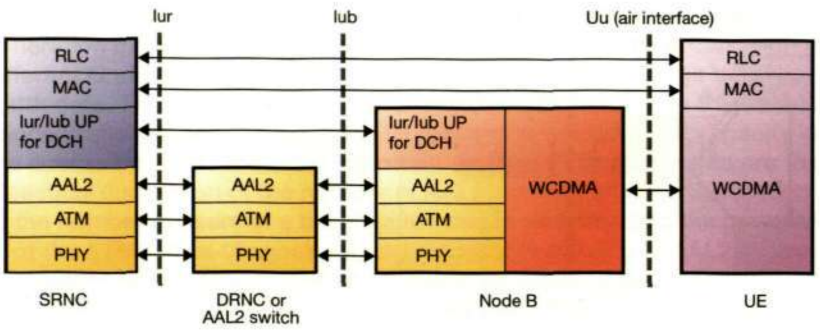


Figure 3
User-plane data transfer between the SRNC and node B.

AAL2 at the *Iur* interface

Mobile users sometimes move from radio cells controlled by a serving RNC (SRNC) to radio cells controlled by another RNC, designated drift RNC (DRNC). The *Iur* interface between the RNCs allows the SRNC to maintain contact with mobile users connected via the air interface to one or more cells controlled by another RNC. The user-plane protocol for the DCH data streams is established between the SRNC and node B. If the DRNC incorporates AAL2 switching, an AAL2 layer connection can be established from the SRNC to the node B—that is, without AAL2 connection termination in the DRNC—minimizing the transfer delays via the DRNC. This is particularly important for DCH data streams, which have strict timing requirements for soft handover. If an AAL2 switching network is built to interconnect multiple RNCs and node Bs, then every AAL2 connection for DCH data streams can be set up directly between the SRNC and node B without passing the DRNC. This configuration further reduces transmission costs. The AAL2 control plane (not shown in Figure 3) is terminated in every AAL2 switching node.

User frames are segmented and packed into AAL2 common-part sub-layer (CPS) packets, which are multiplexed into ATM cells (Figure 4). The AAL2 payload can vary in length (up to 45 bytes). The AAL2 header is 3 bytes in length. All ATM cells are 53 bytes in length, including a 5-byte header.

Thanks to AAL2 multiplexing, the AAL2 packets from several AAL2 connections can be transported on one ATM virtual channel connection (VCC). Each ATM cell on the VCC can carry AAL2 packets from different AAL2 connections. The connection identifier (CID) field in each AAL2 packet header identifies the AAL2 connection to which the packet belongs, much the same as the virtual path identifier (VPI) and virtual channel identifier (VCI) fields in the ATM cell header identify the ATM virtual channel connection.[1,2]

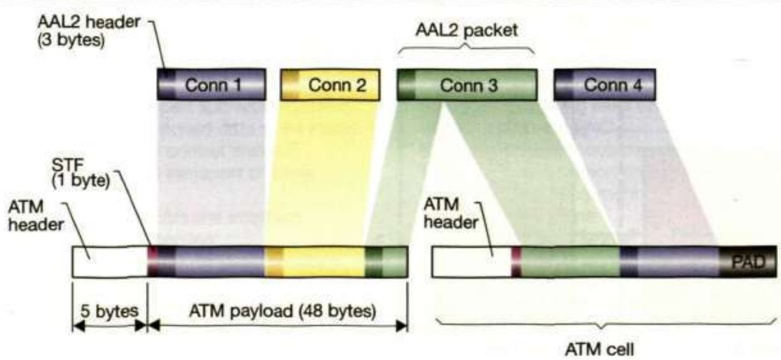
QoS requirements and admission control

The WCDMA RAN transport layer services must meet stringent quality of service (QoS) requirements. The most important measure of QoS performance in the *Iub* and *Iur* interfaces is maximum packet delay.[3]

To satisfy QoS requirements, AAL2 connection admission control (CAC) is executed before a new AAL2 connection is set up in the system. Connection admission decisions are based on the traffic descriptors and QoS requirements assigned to the connections. The AAL2 admission control procedure allocates bandwidth resources (from available virtual channel and path identifier resources) to AAL2 connections in the transport network. If the requisite amount of resources is not available to accommodate a new connection, it is rejected.

If AAL2 connections are transported in end-to-end virtual circuits with resource allocation, AAL2 connection administration control is only executed at the end points of the virtual circuits (node B and RNC). How-

Figure 4
Assembly of ATM cells.



ever, if the resources along the path of an AAL2 connection are not allocated end-to-end, the CAC decisions are replaced by or based on hop-by-hop link admission control (LAC) decisions at every AAL2 switch along the path.

Transport network functions

Node types

The Ericsson WCDMA RAN system is composed of three kinds of traffic-handling nodes:

- Different versions of node B are available for indoor and outdoor placement and to satisfy different needs for the air interface and transport network capacity. One node B can be configured to serve as a transport network hub using ATM cross-connect and AAL2 switching techniques to aggregate traffic to and from other node Bs.
- The radio network controller (RNC) is a modular multi-subrack node. It is available in various sizes to satisfy different capacity needs.
- The RXI is a transport network node that provides ATM cross-connect, AAL2 switching, and IP router services. This fault-tolerant node is a single-subrack design with numerous interface options and a high-capacity switch core.

The WCDMA RAN system also includes various other nodes not described in this article for operation and maintenance (O&M) support.

The system platform

All of Ericsson's WCDMA RAN nodes are based on the same carrier-class technology—the connectivity packet platform (CPP, formerly called Cello packet platform). Modular and robust in design, CPP is is characterized by a multiprocessor control system with multiple processor levels, and its use of cell-switching technology to internally interconnect all types of processor boards, external interface boards and application-specific boards. CPP also includes functionality for terminating and switching ATM and AAL2 connections and for terminating and routing IP traffic.[4]

Switching and termination of ATM and AAL2 connections in CPP

The CPP solution for terminating and switching ATM and AAL2 traffic is at the heart of Ericsson's ATM and AAL2 transport solutions. The node function that provides the through-connection of an ATM virtual channel connection is called the ATM VC cross-connect. Similarly, the node function that provides the through-connection of an AAL2 connection is denoted AAL2 switching.

The 3GPP does not specify how ATM layer connections are to be established and released in the WCDMA radio access network. Ordinarily, permanent virtual channel or virtual path connections are configured using network management actions. In the 3GPP specifications, the establishment and release of AAL2 connections is to be controlled dynamically and in real time by user requests—that is, by the functions of the radio network layer. The AAL2 connections in the WCDMA radio access network are controlled by Q.2630 signaling between the RNC and node B. This signaling can thus be used for setting up connections in a switching transport network between the node B and RNC. The network can be built up of multiple CPP nodes—for example, a tree structure of node Bs combined with pure transport aggregating nodes, such as the Ericsson RXI820 (Figure 5).

The internal cell-switching architecture of the node allows the terminating function of each ATM VCC to be distributed to the processor board or application-specific de-

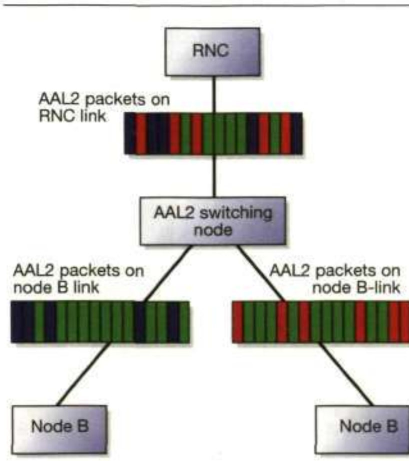


Figure 5
AAL2 switching network.

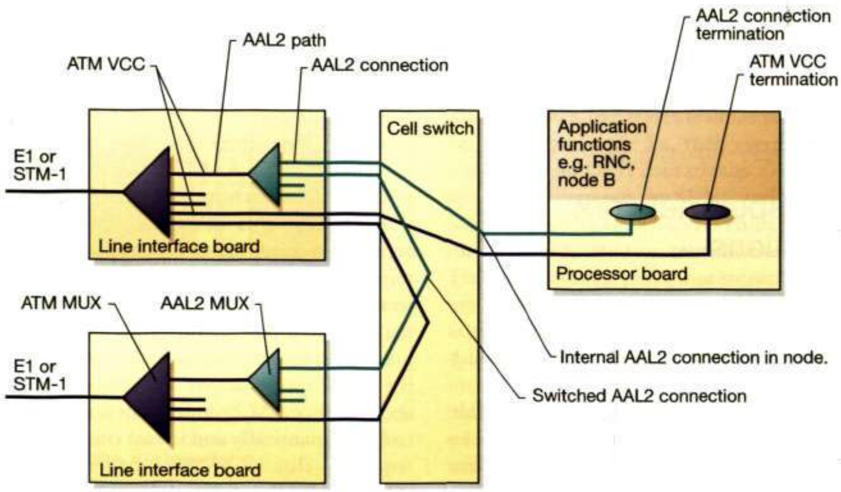


Figure 6
Switching and termination of ATM virtual channel and AAL2 connections in CPP nodes.

vice board on which the application function that uses the connection resides. The ATM line interface boards thus forward ATM cells in either direction between the external line interface and cell switch. The cell switch in CPP transfers ATM cells to and from connection-terminating boards in the node, or to and from external interface boards for cross-connected ATM VCCs.

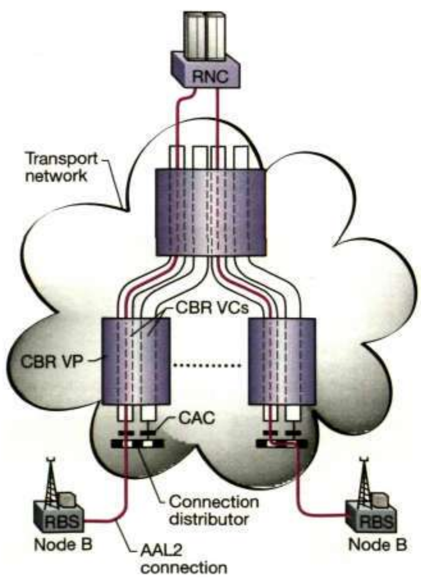


Figure 7
VC-CC model.

Each ATM VCC used by an AAL2 path is terminated by an AAL2 multiplexer function (located on the line interface board) that handles the AAL2 CPS layer, multiplexing and demultiplexing the AAL2 packets to and from ATM cells on the AAL2 path (an ATM VCC, see Figure 6).

For AAL2 connections that terminate within the node, each AAL2 packet is carried via the cell switch directly between the line interface board and the application board that terminates the AAL2 connection.

By contrast, for AAL2 connections that are switched through the node, each AAL2 packet is forwarded via the cell switch directly between the incoming and outgoing line interface boards.

AAL2 packets are switched by hardware on the line interface boards and via the high-capacity cell switch in the same way as ATM cells are switched. Therefore, every traffic-handling RAN node based on CPP technology can efficiently perform switching within the AAL2 layer.

WCDMA RAN transport network topologies and switching alternatives

Although it is possible to connect a node B to its RNC via a direct, physical point-to-point connection, more commonly an *Iub* connection is established via one or more intermediate aggregation nodes. Connections from multiple node Bs can, in this way, be multiplexed onto the same physical link interface to the RNC.

The basic methods of multiplexing and concentrating traffic in the WCDMA RAN transport network are

- physical-layer multiplexing, using SDH network technology—one STM-1 interface (155 Mbit/s) to the RNC can multiplex, for example, 63 E1 (2 Mbit/s) physical-layer connections to as many node Bs;
- ATM VP or virtual channel cross-connect, using ATM network technology—one STM-1 interface to the RNC can multiplex several virtual channels representing ATM layer connections to many node Bs. The number of VCCs is the same at the RNC interface as the aggregated number of VCCs at the node B interfaces; and
- AAL2 switching, using AAL2 network technology—one STM-1 interface to the RNC can multiplex AAL2 connections to multiple node Bs on a common group of AAL2 paths between the RNC and an intermediate node with AAL2 switching capability. Other AAL2 paths are established between the intermediate node and each node B. The AAL2 switching method is combined with ATM VC cross-connect in the intermediate node of end-to-end ATM VCCs between the RNC and each node B; for example, for signaling and O&M access.

These methods are typically combined—that is, AAL2 switching is introduced on top of ATM cross-connect, which is introduced on top of SDH multiplexing.

To accommodate AAL2 switching in intermediate nodes, resources must be allocated to process signaling for the set-up and release of AAL2 connection and for multiplexing and demultiplexing the AAL2 user-plane.

Compared to an intermediate node, which merely cross-connects AAL2 path VCCs, each AAL2 switching node introduces some delay during connection set-up and transport of the user-plane. The network designer must weigh these costs against possible savings in transmission capacity and network management, and decide to what extent AAL2 switching can be employed.

We have studied the gains in transmission capacity from AAL2 switching at different aggregation levels for *Iub* traffic, comparing two cases of network dimensioning:

- VC-CC—use of ATM VC cross-connect of AAL2 path VCCs in aggregation nodes (Figure 7).
- AAL2—dynamic switching of AAL2 connections between multiple downlink AAL2 paths and fewer uplink AAL2 paths in aggregation nodes (Figure 8).

All AAL2 paths are assumed to be configured as constant bit rate (CBR) virtual channels with a fixed, guaranteed capacity. The connection admission control function, which operates on each AAL2 path, needs to know the total capacity of the AAL2 path to calculate the number of different kinds of AAL2 connections that can be allowed on each AAL2 path while maintaining quality of service for every connection.

Benefits of AAL2 switching on the *Iub* interface

Evaluating the performance of AAL2 switching

To evaluate the performance of AAL2 switching, we used a fast and accurate CAC algorithm that takes into account the properties of traffic across the *Iub* interface.[5] The transport network layer grade-of-service (GoS) requirement of each RAB type is met if the blocking probability of its connection remains below a given threshold, for example, 0.3% for voice. Besides the offered traffic (measured in Erlang), the blocking prob-

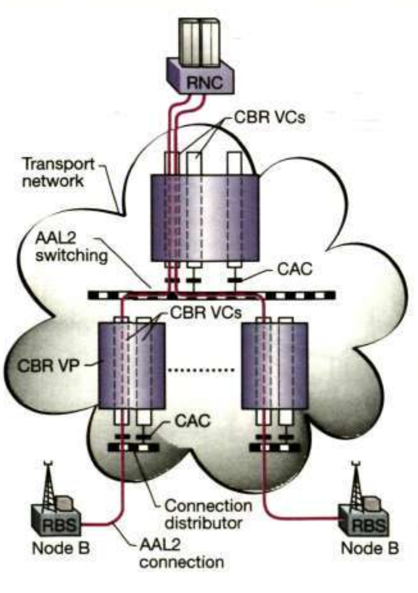


Figure 8
AAL2 switching model.

TABLE 1, CONSIDERED TRAFFIC MIX, USER LOAD IN MERLANG AND GOS REQUIREMENT

mErlang per user	Standard	Data-oriented	Contains PS 384	Target GoS (%)
Voice	20.25	17.00	25.00	0.3
CS 64	2.25	1.00	-	0.3
PS 64	3.30	5.00	1.00	0.7
PS 384	-	-	0.04	4.0

abilities also depend on the admission control algorithm, and on the number of available AAL2 paths (ATM VPs) between the two nodes.

The connections were generated randomly in a simulator, and the packet-level traffic descriptors were attributes of the generated connections. When a new connection was generated in the simulator, the CAC algorithms executed in the appropriate nodes in the access network, and we measured the blocking probabilities for each service.

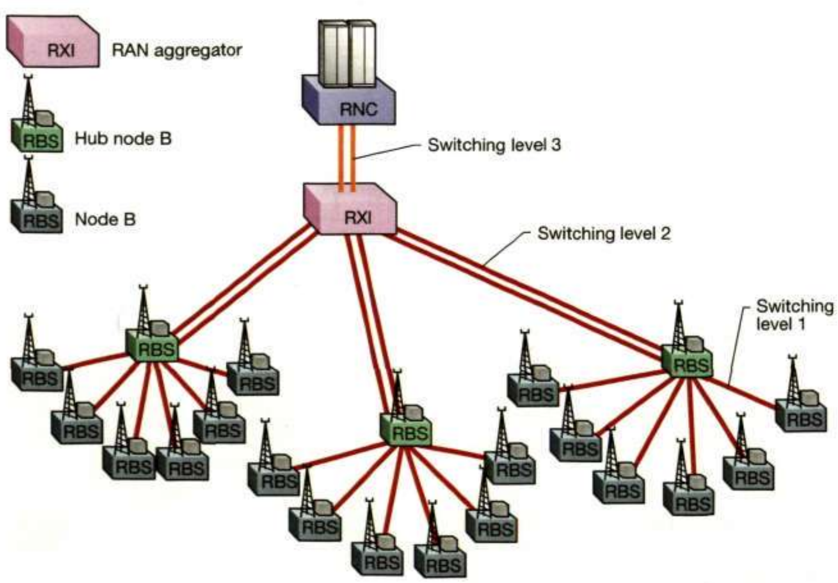
Traffic parameters

During the simulation, different RAB types were considered for the following services: AMR-coded voice, circuit-switched data at 64 kbit/s (CS 64), packet-switched data at 64 kbit/s (PS 64) and packet-switched data at 384 kbit/s (PS 384). The evaluation also took into account all common channels in a cell and a dedicated control channel (DCCH) connection for each RAB.

We ran simulations for different kinds of traffic mixes (Table 1). The results represent a standard mix of voice (12.2 kbit/s AMR-coded), CS 64 and PS 64. Since many operators do not intend to provide PS 384 bearers when network load is heavy, it was relevant to study traffic mixes that excluded this rate. The results from the other traffic mixes also favored AAL2 switching over VC cross-connect.

For each *Iub* interface, about 200 kbit/s was allocated for non-AAL2 control-plane and O&M signaling.

Figure 9
Tree topology of the WCDMA RAN.



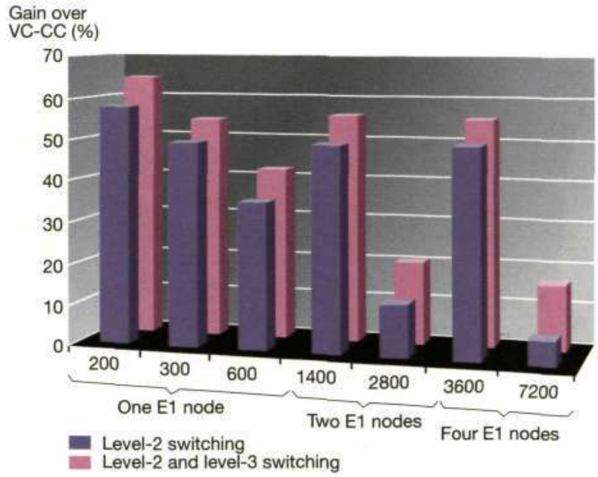


Figure 10
Reduced need for capacity for different number of users and switching layers in the 3x6 topology.

Note: While the gains from AAL2 switching are typical, the results apply to a specific traffic mix. The actual gains for other traffic volumes, mixes of voice and data, and other activity factors for voice and packet data might vary.

Statistical gain at different traffic loads

We obtained the numerical results by simulating the network at the connection level.

Low traffic volume

Compared to VC-CC switching, the greatest reduction in bandwidth using AAL2 switching can be expected when traffic is light—for example, when only a small part of the allocated bandwidth of the VCs is used.

Let us consider a tree network topology with three AAL2 switching nodes at switching level 2 (SL2), and six node Bs connected to each concentration node (Figure 9). We call this topology the 3x6 topology. Initially, the link capacity at the lowest level is one E1 per node B. The number of users is less than 600 per node B (600 is the maximum number, assuming a standard or data-oriented mix).

For VC-CC switching, the required bandwidth is 14 Mbit/s for a switching node in SL2, and 42 Mbit/s for a switching node in SL3.

We next increased the number of users served by a node B. Obviously, at some point, as the number of users grows, more capacity is needed to connect node B. Figure 10 shows how AAL2 switching reduces the need for capacity in the link compared to VC-CC switching at SL2 and SL3 for different numbers of users per node B using the standard traffic mix. When the number of users exceeds 600, two E1 access links are used. Likewise, when the number exceeds 2,800, four E1 access links are used.

The gain in capacity from using AAL2 switching is calculated as follows:

$$\text{Gain} = (C_{VC\text{-}CC} - C_{AAL2}) / C_{VC\text{-}CC}$$

where CVC-CC is the entire link capacity (aggregated capacity at the considered level) needed for VC-CC, and CAAL2 is the link capacity needed for AAL2 switching.

As can be seen, applying AAL2 switching in SL2 reduces the consumption of link capacity by 35-60%. The gain in capacity obtained from AAL2 switching is especially high (around 60%) when the use of SL1 links is low. The gain is even greater (up to 65%) at SL3.

Thanks to the large gain in capacity from AAL2 switching, the operator can install less capacity at SL2 and SL3 when first introducing service, and later upgrade the link capacity between switching nodes.

Fully loaded system

Operators generally try to avoid loading the network to its limit. Nevertheless, they do need to determine the capacity limit and in-

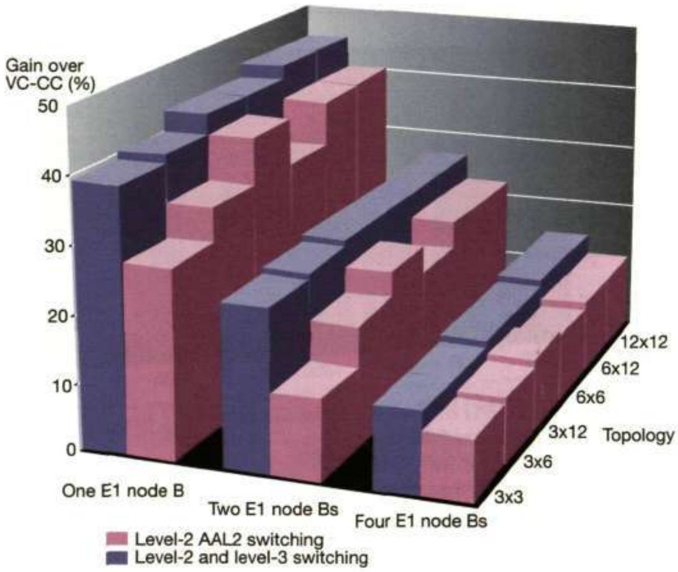


Figure 11
Fully loaded system. Gain in level-2 and level-3 link capacity from AAL2 switching.

vestigate traffic cases close to this limit. QoS problems are more likely to occur when load is heavy. Figure 11 shows the gain in capacity for the different network sizes with up to twelve switching nodes in the second level, twelve node Bs connected to each switching node (12x12 topology), and a fully loaded system.

Compared to VC-CC switching, the gain in capacity increases as more node Bs are added per SL2 switching node. Due to a homogenous distribution of traffic among node Bs, the gain from AAL2 switching at SL2 is the same for configurations with the same number of node Bs per switching node (for example, 3x12, 6x12 and 12x12). By

TABLE 2, USER DISTRIBUTION THROUGHOUT THE DAY

Users per node B	CNA1	CNA2	CNA3
Morning	1,400	2,800	2,800
Mid-day	2,800	1,400	2,800

applying AAL2 switching at L2 and L3, the gain increases slightly as a function of the number of switching nodes.

The 3x6 topology used two E1s per node B (Figure 11): An 18 percent gain in capacity was obtained when AAL2 switching was applied at level 2, and 25%, when applied at levels 2 and 3.

Gain from changing traffic distribution

In general, the distribution of traffic among node Bs is not homogenous. A concentrating switching node aggregates the traffic of several node Bs. Therefore each switching node has an associated serving area, which we call the concentration node area (CNA).

Let us assume that a CNA covers an office area and a residential area. Applying VC-CC switching, we attempt to establish as many VC connections between node Bs and the RNC as are needed to serve the sum of the maximum traffic of each node B. In this case, it does not matter that peak traffic occurs at different hours in the office and residential areas.

However, AAL2 switching allows the concentrated link capacity to be dimensioned for the sum of the actual traffic in the CNA.

Consider a 3x6 topology, in which the node Bs are connected by two E1 links. The maximum number of users in the switching node areas is 2,800 per node B assuming a standard mix of traffic. Let us also assume that due to user mobility, the number of users per node B changes over time (Table 2). Compared to VC-CC switching, the gain from AAL2 switching is 26% on SL1 and 32% on SL2.

TABLE 3, COMPARISON OF SWITCHING ALTERNATIVES

	VC-CC	AAL2
Statistical multiplexing capacity reduction	No	High (7 to 65%)
Reduction if traffic distribution changes	No	Yes (26 to 32%)
QoS guarantee	Yes, with CAC for each AAL2 path configuration	Yes, with CAC for each connection setup
Number of VCs minimized	No	Yes

Conclusion

The performance of AAL2 switching is superior to that of VC-CC switching. In several network scenarios that use AAL2 switching, the increased cost of processing AAL2 signaling is more than compensated for by significantly increased efficiency in the transmission network, and reduced costs of transmission.

The savings are greater when there is little traffic in the network—precisely when the network operator needs most to keep costs down. AAL2 switching also considerably reduces the need for link capacity in a fully loaded network with traffic concentration.

By applying AAL2 switching, operators can significantly reduce the need for link capacity since the network can adapt to changes in traffic levels.

REFERENCES

1 ITU-T Rec. I.363.2, "B-ISDN ATM adaptation layer type 2 specification"
2 ITU-T Rec. Q.2630.1 and Q.2630.2, "AAL Type 2 Signaling Protocol (Capability Set 1 and 2)"
3 3GPP TR 25.853, "Delay Budget within the Access Stratum"
4 Kling, L.-Ö., Lindholm, Å., Marklund, L. and Nilsson, G. B.: CPP—Cello packet platform. Ericsson Review Vol 79(2002): 2, pp. 68-75.
5 Sz. Malomsoky, S. Rácz and Sz. Nádas, "Connection Admission Control in UMTS Radio Access Networks", accepted for publication to Computer Communications, Elsevier Science, 2002.

Microwave transmission in mobile networks

Aldo Bolle and Andrea Nascimbene

Microwave links became an enormous success with the roll-out of second-generation mobile networks. With close to 500,000 units delivered to date, the Ericsson family of MINI-LINK microwave products has an important role in mobile operator networks. Now, the advent of third-generation mobile networks is starting a new wave of deployment characterized by cost-effective and flexible roll-out, and short site-to-site distance. Moreover, we are seeing a shift in focus from plain point-to-point bit transport to a network view with optimized site solutions.

The authors address the launch of Ericsson's microwave solution for transmission in current second-generation and imminent third-generation mobile networks, showing how combined use of the point-to-multipoint and point-to-point technologies provides the most cost-effective and spectrum-efficient solution.

The inherent reliability and cost-effectiveness of microwave technology have been given a dominant role in connecting mobile radio base stations (RBS). The roll-out of packet-data and third-generation mobile networks fundamentally changes the traffic demands on transmission systems. Consequently, new microwave transmission techniques and solutions are required.

With the continuous growth of mobile subscribers and mobile data communication, operators need enhanced microwave transmission systems. In particular, enhanced features are needed to handle changing traffic patterns efficiently, to offer increased capacity, and to make optimum use of radio spectrum.

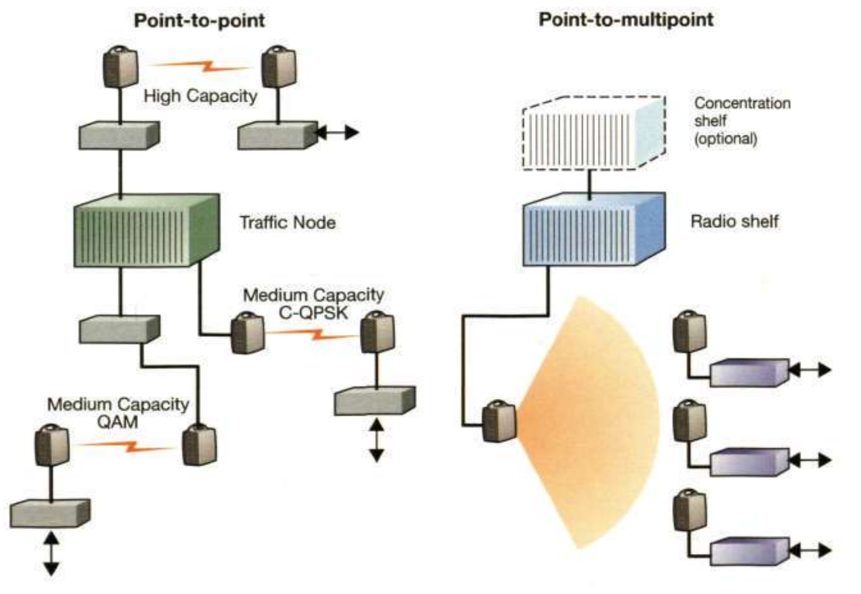
To achieve profitability, operators must have flexibility and be able to respond quickly to dynamic market conditions. These requirements make microwave, with its ease of implementation, ideal for access transmission.

Today, 60% of all second-generation RBSs are connected via microwave technology. As voice and data traffic increases in mobile networks, PDH-based point-to-point microwave solutions can be complemented with ATM-based point-to-multipoint solutions and SDH equipment to create a unified, fully integrated and cost-effective transmission solution that gives operators the best network control and most profitable operation.

MINI-LINK portfolio

The MINI-LINK portfolio includes solutions for point-to-point as well as for point-to-multipoint operation. Terminals and smart nodes (Figure 1) are used for implementing the building blocks in a network.

Figure 1
The MINI-LINK portfolio.



MINI-LINK point-to-point

Ericsson's microwave point-to-point portfolio consists of MINI-LINK Medium Capacity and High Capacity terminals, and the MINI-LINK Traffic Node (Figures 2-3). Depending on the range and capacity to be implemented, the MINI-LINK portfolio offers frequencies ranging from 7 to 38 GHz, for hop lengths of several tens of kilometers to just a few kilometers, and transmission capacities of up to 155 Mbit/s. Constant envelop offset – quadrature phase-shift keying (C-QPSK) and quadrature amplitude modulation (QAM) schemes are available for the terminal configurations. The MINI-LINK Traffic Node, which is a smart node for point-to-point operation, has been optimized for the aggregation nodes in the network, thus providing the ideal capacity and functionality to solve transmission needs. It complements the terminals with the additional features needed to provide a complete and efficient site and network solution.

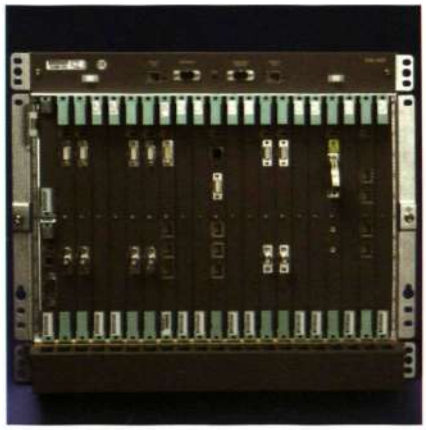


Figure 2
MINI-LINK Traffic Node.

BOX A, TERMS AND ABBREVIATIONS

AAL2	ATM adaptation layer type 2	QAM	Quadrature amplitude modulation
ATM	Asynchronous transfer mode	RAN	Radio access network
CPP	Connectivity packet platform	RBS	Radio base station
C-QPSK	Constant envelop offset – quadrature phase-shift keying	SDH	Synchronous digital hierarchy
E1/E2/E3	ETSI digital multiplexing stage	SNMP	Simple network management protocol
IP	Internet protocol	STM-1	Synchronous transport module level 1
LAN	Local area network	T1/T2	ANSI digital multiplexing stages
MIB	Management information base	VC	Virtual container
OC-3	ANSI digital multiplexing stage	xDSL	Digital subscriber line
PDH	Plesiochronous digital hierarchy		

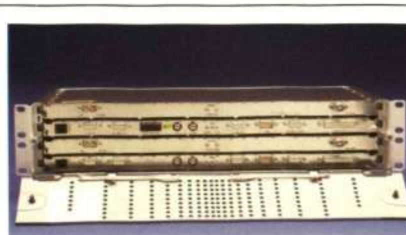


Figure 3
MINI-LINK Medium Capacity terminal (left) and MINI-LINK High Capacity terminal (right).

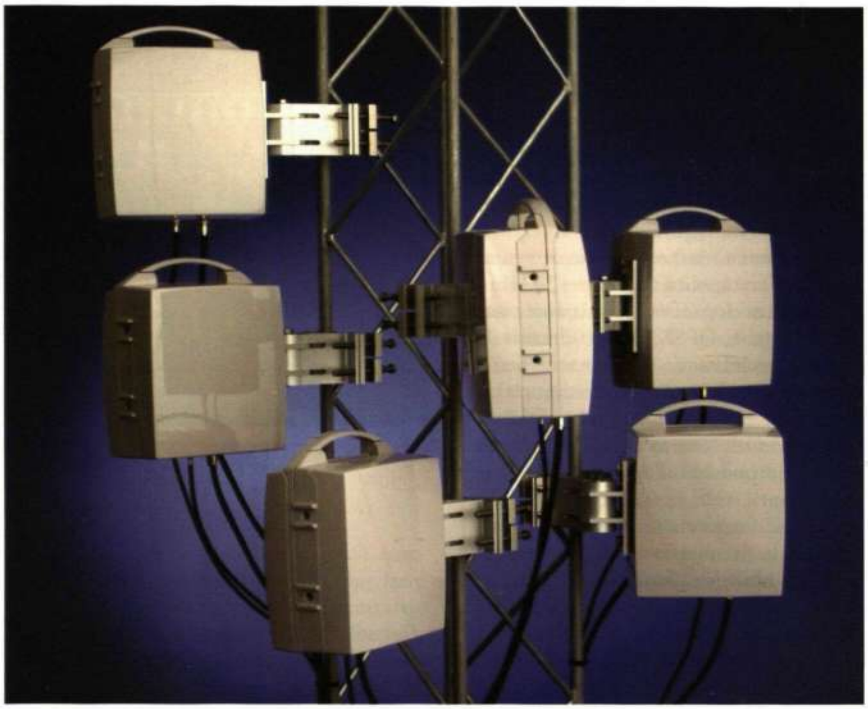


Figure 4
MINI-LINK outdoor radios.

MINI-LINK point-to-multipoint

The MINI-LINK point-to-multipoint system (Figure 5) provides 37.5 Mbit/s data transfer per sector. Each sector can be 90° in the standard solution or 180°/360° in the "launch" solution, in accordance with the required capacity and RBS density ratio. The capacity within a sector can be fixed or dynamically allocated to each terminal, allowing, in the latter case, reallocation of capacity within a few milliseconds. The system is thus very suitable for data traffic, both for business access and backhaul in mobile systems. It uses ATM to guarantee different classes of service. E1, ATM (over E1/T1, E3/T3 or STM-1/OC-3) and Ethernet interfaces are available. The system operates on frequencies from 24 to 31 GHz and uses the C-QPSK modulation scheme.

Figure 5
MINI-LINK BAS radio shelf.

Management system

The third building block in the portfolio is the MINI-LINK Manager (Figure 6), which enables operators to manage a complete MINI-LINK microwave transmission network from a single screen. Network element management provides functionality for managing faults, performance, configurations and security. Together with local craft terminals (LCT) and the element-management functionality embedded in the network elements, the MINI-LINK Manager gives operators the tools they need for efficient and cost-effective operation of a MINI-LINK network.

MINI-LINK Manager has several export interfaces for easy integration into other network-management systems. It can be incorporated into a total management solution for mobile systems, either as part of a complete solution provided by Ericsson or as an integration with an existing management system.

MINI-LINK features

Bandwidth aggregation

The point-to-point and point-to-multipoint smart nodes are hub solutions developed to support a large number of sites and future increases in capacity. Being scalable, the smart node enables the aggregation of traffic bandwidth that originates from a large number of end-nodes. At Medium Capacity aggregation nodes, the bandwidth is aggregated into a medium-capacity interface (maximum 34 Mbit/s) Similarly, at High Capacity aggregation nodes, the bandwidth is aggregated into a high-capacity interface (STM-1 or greater). Traffic from the aggregation nodes can be further transmitted either on microwave or optical links.

In a point-to-multipoint system, the air interface is shared among multiple access terminals. The shared media enables multiplexing gains over the air, provided a packet-based infrastructure is employed. MINI-LINK point-to-multipoint is based on ATM end-to-end, which enables multiplexing gains and efficient usage of the bandwidth when second- and third-generation traffic is handled in the aggregation nodes.

Use of spectrum

Spectrum is a sparse resource. Besides the continuous development of radios in newly allocated frequency bands, some important new features have been introduced in the MINI-LINK portfolio to deal with future shortages of spectrum. To allow the operator to increase transmission capacity within an existing frequency spectrum, higher-order modulation methods (based on 16 and 128 QAM) have been introduced in the MINI-LINK point-to-point portfolio. These new features give the operator additional flexibility in balancing spectrum and power efficiency in the network.

Point-to-multipoint systems (Figure 7) make efficient use of spectrum by

- allocating capacity per ATM cell (ATM granularity gain) instead of on a 2 Mbit/s-basis;
- ATM multiplexing in conjunction with fast dynamic capacity allocation. The network can be "oversubscribed" in terms of number of registered users while still maintaining QoS; and
- delivering unused access capacity to other services, such as wireless LAN access points or business access users, based on the diversity gain of the daily traffic profile (daily profile gain), since the busy hours for residential users generally differ from those of business users.

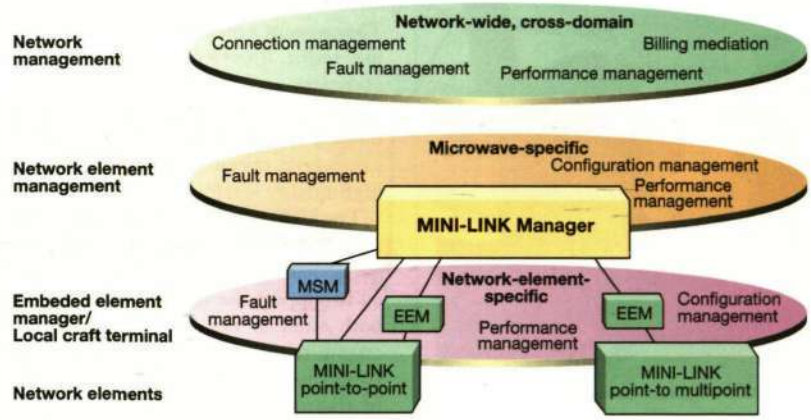


Figure 6
MINI-LINK Manager—its role in network management.

Figure 7
Aggregation gain. The diagram shows the aggregated link capacity required by multiple base stations per base station. The red line indicates aggregating link capacities. The yellow line represents peak load capacities, and the blue line, average traffic loads. The aggregation gain increases as the number of base stations connected in the same sector increases.

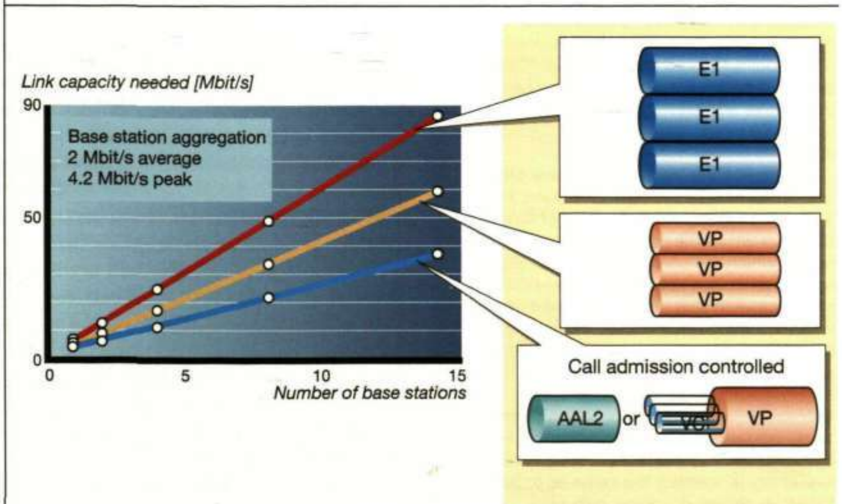


Figure 8
Installing MINI-LINK.

BOX B, PLANNING MICROWAVE TRANSMISSION, AVAILABILITY AND QUALITY

Traditionally, operators have deployed mobile backhaul networks using a combination of point-to-point microwave and leased lines. A determinant when choosing between microwave and leased lines is the individual operator's needs in terms of network control and transmission quality.

Typical leased-line contracts have often guaranteed availability figures around 98.7%, which corresponds to a potential of four or five days downtime per year. Microwave networks (which are often used to relink the entire connection between the end-RBSs and the switch site) are dimensioned for 99.95% availability or better, which corresponds to four hours or less of downtime per year.

In conclusion, the availability of a microwave network is very much a planning issue. By selecting high-quality products in combination with proper network planning, availability is normally the same as or better than that of fiber or copper networks.

Protection

The operator's most important asset is end-user traffic. If service delivery is not reliable, end-users will change service providers. High-quality equipment that is complemented with additional protection mechanisms gives operators a means of delivering high-quality services.

The MINI-LINK products are protected against equipment failure and radio propagation anomalies. All hardware is duplicated to support the configurations on one or both sides of the radio connection. The transmitting equipment can be configured to operate in hot standby or working standby transmission mode.

The MINI-LINK Traffic Node adds yet another level of protection—network or ring protection. This functionality enables the operator to build reliable ring structures based on any microwave capacity up to 155 Mbit/s. These protection mechanisms work at the E1/T1 level, protecting every or pin-pointed E1/T1s within the total payload.

The Traffic Node solution also includes line-protection mechanisms without the duplication of hardware. Instead, the E1/T1s to be protected are routed into two separate ports on the same interface board.

Ease of installation and visual impact

Speed of installation is always a business consideration, especially during the roll-out of third-generation networks in Europe. Microwave is less costly and time-consuming to deploy than copper leased lines. The MINI-LINK portfolio has been optimized for simple installation with a compact, easy-to-carry outdoor unit (Figure 8). The single-cable interface between the indoor and outdoor unit, and the single-bolt alignment fixture are well known. The point-to-multipoint system is even less complex and therefore faster to install, since only one end of the link has to be installed. In addition, new base stations or interfaces can be added to the backhaul network configuration, literally in a matter of minutes, minimizing maintenance and upgrade costs.

The point-to-multipoint hub needs only one antenna (and a single cable between outdoor and indoor equipment) per sector, regardless of the number of connected RBSs. This strongly minimizes the visual impact, especially in cities and towns where antenna pollution is an important issue. Moreover, fewer antennas means fewer sites (simpler site acquisition) and reduced installation time and cost.

Data communication networks

The Ericsson network solution for transporting operation and maintenance (O&M) information from equipment to the management center is based on IP communication over Ethernet, with a distributed management information base (MIB) architecture.

MINI-LINK provides efficient in-band data communication between end-nodes and aggregation nodes. The MIB is physically located in each network element. Using the simple network management protocol (SNMP), operators can access O&M information in the MIB remotely from a network management system. They can also access the information locally, on site, by means of the local craft terminal. The terminal software can be upgraded remotely from a central location, or locally using a laptop connected to the terminal.

Each MINI-LINK Traffic Node and High Capacity terminal holds its own IP router for extending the data communication network throughout the transmission network, and transporting O&M information on other equipment via external service channels.

Combined solutions for the mobile transport RAN

In dense areas, point-to-multipoint has clear advantages over point-to-point transmission. As a simple rule of thumb, point-to-multipoint becomes an interesting option when four or five RBSs can be seen from one location. However, the two technologies are, and will be, used in combination. Point-to-point microwave, which is typically deployed in areas with fewer RBSs, can be combined with point-to-multipoint to overcome distances or interference.

The combination of Ericsson's point-to-point and point-to-multipoint product families results in the most cost-optimized and spectrum-efficient solution for second- and third-generation networks (Figure 9).

E1/T1 aggregation via point-to-point links is typically suitable in small hubs where the number of directions (or connected RBSs) is limited and spectrum is not an issue (the required bandwidth is very likely to be a portion of that required to deploy the large hub).

ATM aggregation, typically via point-to-multipoint, is more suitable in large hubs where the number of directions (or connected RBSs) is great and spectrum efficiency is a must (since it determines the size of the frequency blocks required).

The hubs are connected to each other, to the switch site, or both, via point-to-point systems in accordance with the required range, capacity and available spectrum.

E1/T1 multiplexing nodes

The E1/T1 multiplexing node is the current solution for present-day second-generation networks. In all likelihood, it will also be the most efficient solution for operators who plan to add third-generation services in environments where second-generation traffic will continue to dominate. This is also the typical solution for operators who want to reuse as much of the existing network as possible (by exploiting spare capacity on the microwave links or on STM-1/OC-3 rings). This aggregation strategy might also be justified by the price structure for leased lines. The benefits of a network based on E1/T1 multiplexing nodes are low initial investments and secure upgrade with minimum disturbance to existing traffic.

Figure 9 exemplifies how a combination of point-to-multipoint and point-to-point links can efficiently serve the Medium Capacity and High Capacity aggregation nodes.

The Low Capacity and Medium Capacity aggregation nodes typically handle from two to four radio base stations—that is, from two to four directions. These nodes are generally deployed where RBS density is low and the RBS-to-RBS distance is great.

In the southbound direction (Figure 9), the Medium Capacity aggregation nodes interconnect the end RBSs through MINI-LINK point-to-point; in the northbound direction, the connection can be made via MINI-LINK point-to-multipoint (Low Capacity ATM aggregation nodes) or point-to-point terminals (Medium Capacity E1/T1 aggregation nodes), depending on capacity, protection and range requirements.

Ordinarily, the High Capacity aggregation nodes are located in suburban or urban areas where RBS density is high. During operation, error-free transport over microwave links is guaranteed by large fading margins and forward error correction mechanisms, which make microwave links highly suitable for ATM and IP transport. In these sites point-to-multipoint is likely to connect the end RBSs. Those RBSs that are outside the point-to-multipoint coverage range are connected through point-to-point links.

When E1/T1 traffic is aggregated, the MINI-LINK point-to-multipoint system

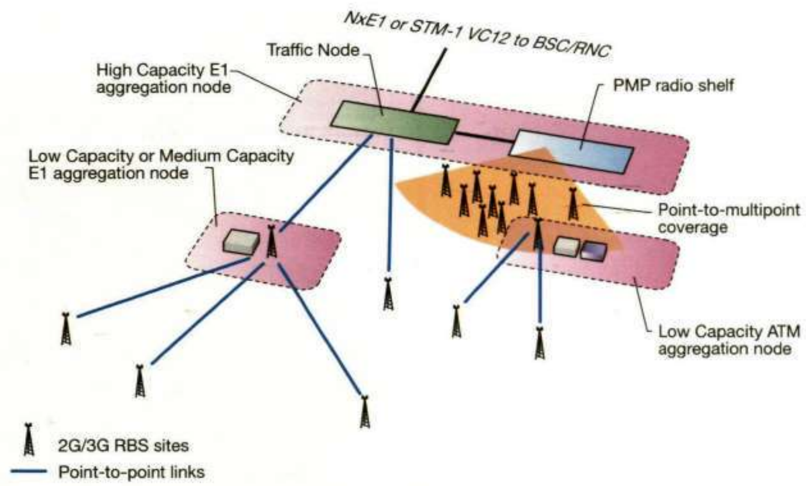


Figure 9
Example of site solutions based on E1 multiplexing and the combining of point-to-multipoint and point-to-point.

for second-generation traffic connects to the MINI-LINK Traffic Node through *n*xE1, which, in turn, provides a single STM-1 VC12 interface to the switch site.

The main drawbacks of the E1/T1 multiplexing solution can be limited expansion and greater long-term cost of operating the network.

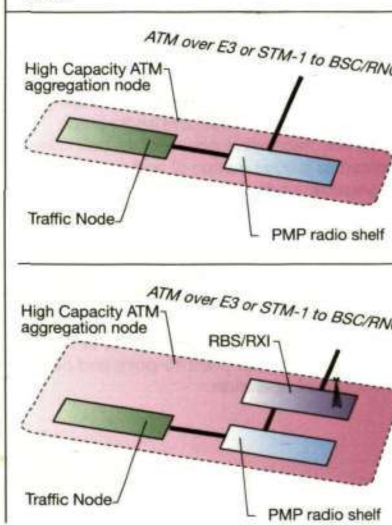
ATM aggregating nodes

When third-generation traffic dominates over second-generation traffic, ATM aggregating nodes can be used to provide the most cost-effective network solution. Networks based on ATM aggregating nodes are likely to be typical for greenfield operators and for incumbent operators who want to overlay the existing network or to replace existing leased-line connections.

The MINI-LINK point-to-multipoint hub provides port aggregation, aggregating traffic from point-to-multipoint and point-to-point terminals. It also provides a very efficient and cost-effective solution for cellular backhaul applications. It can also aggregate traffic from leased lines and xDSL lines. In the northbound direction, a single ATM-over-STM-1 VC4 interface provides a very clean and cost-effective solution that optimizes backbone capacity, switch site complexity and cost.

This solution can also be used in combination with Ericsson's RBS and RXI products, providing a complete Ericsson mobile and transport network (Figure 10). In

Figure 10
Top: Example of High Capacity aggregation node handling ATM.
Bottom: Example of combined CPP and MINI-LINK point-to-multipoint (PMP) radio shelf.



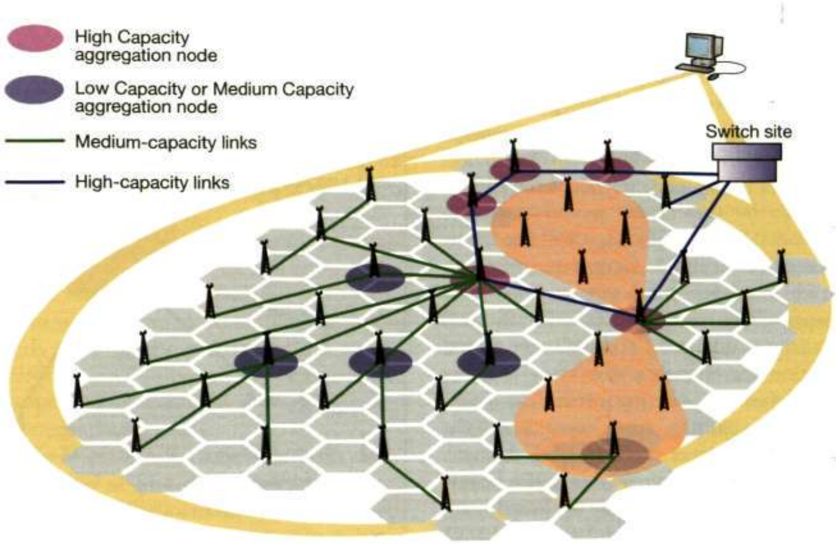


Figure 11
Network architecture.

addition to the benefits of ATM aggregation, the solution brings optimized statistical multiplexing gain, thanks to the AAL2 switching functionality of the connectivity packet platform (CPP, formerly called Cello packet platform). Because ATM virtual-path multiplexing and port aggregation are performed in the MINI-LINK point-to-multipoint radio shelf, the AAL2 functionality is achieved while optimizing costs (no increase in number of boards).

BOX C, NETWORK ARCHITECTURE

The use of short-haul microwave radio has evolved from scattered cable replacements to the forming of complete microwave-based transmission networks. The requirements put on the products have shifted from optimization of the terminal or hop level to optimization of the network level. In a microwave network, one can define logical nodes (or physical sites) with distinct characteristics. The logical building blocks are the end-node and aggregation node. Any microwave network can be implemented as a combination of end-nodes and aggregation nodes (Figure 11).

To address the network aspects, Ericsson's products are optimized for the different types of network node. Therefore, the MINI-LINK portfolio comprises compact, cost-effective access terminals and smart nodes that feature advanced traffic routing and multiplexing. The MINI-LINK portfolio includes access terminals and smart nodes for point-to-point and point-to-multipoint operation.

Typical building blocks of a microwave network

End-node
The end-node is the smallest building block. By definition, it supports transmission in only one direction. In most cases, the capacity of the end-node ranges from 2x2 up to 34 Mbit/s. Ordinarily no redundancy is required at end-node sites and therefore the normal microwave configuration is 1+0. Point-to-point and point-to-multipoint end-nodes are foreseen. The end-node should support traffic interfaces ranging from multiple E1/T1s to Ethernet. Ideally, in a point-to-multipoint system, the end-node will provide an ATM interface for third-generation backhaul, to take better advantage of the shared air interface.

Low Capacity and Medium Capacity aggregation nodes
The Low Capacity and Medium Capacity aggregation nodes have a northbound microwave link that carries traffic up to 34 Mbit/s. In the southbound direction these nodes have a limited number of subtended end-nodes.

Ericsson's solution to the Medium Capacity aggregation node has been to design smart, cost-effective Traffic Nodes that can aggregate all traffic from the southbound links into another microwave link in the northbound direction. The solution supports protected and non-protected configurations of the Medium Capacity aggregation node. The solution also supports dropping and insertion of local traffic.

High Capacity aggregation node
The High Capacity aggregation node has a northbound transmission link with a traffic capacity of 155 Mbit/s or greater. The northbound media can be either optical or microwave. The topology in the northbound direction can be ring or point-to-point. Since the High Capacity aggregation node supports a considerable amount of traffic, it is assumed that most of the sites will aggregate a substantial number of southbound links. Some end-nodes are directly connected to the High Capacity aggregation node and some are connected through a Medium Capacity aggregation node. Point-to-point, point-to-multipoint and E1/T1 and ATM aggregating sites are supported. The Ericsson solution to the High Capacity aggregation node can be designed to be very compact and cost-effective, as part of an all-microwave solution that supports 155 Mbit/s traffic capacity.

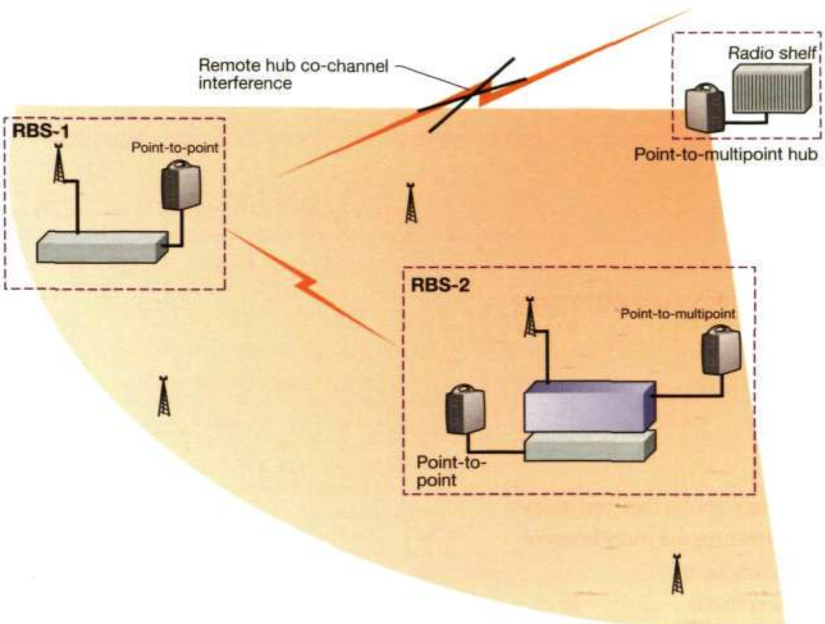


Figure 12
Operators can reuse frequencies by combining point-to-multipoint and point-to-point systems.

Conclusion

The key issues for efficient support of the mobile network infrastructure are:

- A complete portfolio of point-to-point solutions (any frequency, any capacity, PDH/SDH), explicitly designed for a smart network-oriented approach.
- An ATM-based point-to-multipoint solution that provides a suitable combination of high coverage and high capacity.
- A combination of these technologies to provide the most cost-effective and spectrum-efficient microwave solution.
- An integrated management system for the entire portfolio.
- Proven reliability and large production capability for secure roll-out.

Uniquely, the Ericsson MINI-LINK portfolio can meet all of these requirements.

BOX D, IMPROVING SPECTRUM EFFICIENCY BY COMBINING POINT-TO-MULTIPOINT AND POINT-TO-POINT SYSTEMS

One fundamental issue in microwave network planning is the efficient use of the frequency spectrum. National authorities and international committees regulate the availability of spectrum. Point-to-point links typically require a license per link, whereas licenses for point-to-multipoint systems are issued as regional or national block allowances. In many cases, operators prefer block licenses since they allow faster planning and deployment of the links.

In point-to-multipoint cellular deployments, a few locations inside the multipoint sector can experience interference from neighboring hubs. However, this effect can be minimized by avoiding reuse of frequencies in neighboring sectors or by combining point-to-multipoint with point-to-point technologies.

In Figure 12, the RBS-1 location is assumed to be affected by co-channel interference from a remote point-to-multipoint hub if connected to the local hub through a point-to-multipoint terminal. If the RBS-1 is instead connected to the RBS-2 location by means of a point-to-point link, the antenna angular discrimination improves the carrier-to-interference ratio and guarantees error-free operation. It is worth noting that the point-to-point link can reuse part of the same point-to-multipoint spectrum, allowing for a very spectrum-efficient solution. Thanks to the combined MINI-LINK point-to-point and point-to-multipoint solution, only a single 28 MHz link is required for the complete network deployment (excluding the spectrum for the northbound connections).

Previous issues

No. 2, 2002

No. 1, 2002

No. 4, 2001

No. 3, 2001

No. 2, 2001

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

ISSN 0014-0171
Ljungföretagen, Örebro 2002