Managing the complexity evolution

The data challenge – quantified by massive information volumes, demanding user expectations, constant connectivity and the concept of more than 50 billion connected devices – creates sustainable business opportunities if complexity, volume and individuality can be managed.

LTE Rel-10 will be finalized at the end of this year, supporting operators in their bid to offer new services demanded by subscribers without risking existing investments. The migration to all-IP means new demands on the voice interconnect between operator networks.

Today, the telecommunications industry uses several business models to serve multiple subscriber types. Twenty years ago we had one business model with one subscriber type; in just a couple of years, we may have as many models as we have subscribers, making business complex and costly for operators. Ericsson’s approach to cross-functional Business Support Systems (BSS) will allow operators to manage the complexity and turn it into revenue.

Device and infrastructure hardware considerations also factor in the complexity equation. Scalable parallel computing will improve hardware efficiency and help reduce costs.

Network evolution, almost revolution, exposes communications systems to new threats. Evolved network architectures that operators and users alike can rely on for integrity and identity protection are a foundation for continuous growth for everyone.

“Information and content have a symbiotic relationship to connectivity. The higher the connectivity speed, the greater the availability of information and content; the greater the availability of information and content, the greater the demand for connectivity.”

Håkan Eriksson
Group CTO and President of Ericsson Silicon Valley
Security in the Evolved Packet System

Wireless telecommunications systems must live up to user and network service provider expectations regarding trust and privacy. Besides the obvious need for authentication and encryption (to enable reliable charging and to hinder eavesdropping), new architectures require more sophisticated protection mechanisms. The authors provide an overview of the security architecture and security features of the Evolved Packet System for LTE and non-3GPP accesses.

Evolution of the voice interconnect

During the past decade, network operators have transformed their existing wireline and wireless core networks from Time Division Multiplex (TDM) or Asynchronous Transfer Mode (ATM) transport technology to IP transport. The evolution of the voice interconnect between operator networks to IP is another natural step towards an all-IP architecture. This transition will result in more efficient bandwidth usage, better speech quality, reduced costs, and a clear migration path to a full multimedia service interconnection.

Scalable parallelism using dataflow programming

Dataflow programming was invented to address the issue of parallel computing. Dataflow models are well suited to describe many forms of computation, particularly in the area of digital signal processing. This article presents case studies about an MPEG-4 video decoder and a multi-standard OFDM receiver. Hardware and software have been synthesized from the dataflow models, and experimental results are presented.

Next generation LTE, LTE-Advanced

Next generation LTE, LTE-Advanced or LTE Rel-10 is the next step in radio access technology. Whatever the name – next generation LTE, LTE-Advanced or LTE Rel-10 – the next step in LTE evolution allows operators to introduce new technologies without putting existing investments at risk.

Business Support Systems

The BSS helps an enterprise or organization to secure revenue and quality while supporting many business functions, including marketing, product offerings, sales, contracting, and delivery of goods and services. A well designed BSS helps an enterprise stay ahead of the competition by providing the flexibility to adapt to a constantly changing marketplace. This article explores Ericsson’s approach to this complex area, discussing the transition from today’s array of silo-like, integration-heavy, multi-vendor environments to the unified, cross-functional and easy-to-use solutions of the future.

Ericsson Business Communication Suite

Ericsson’s Business Communications Suite (BCS) is a set of applications that an operator can install to offer enhanced communication services to companies in the small and medium enterprise (SME) segment. With BCS, operators can extend their portfolios – and consequently revenue streams – with a branded end user experience spanning multiple devices into the rapidly growing Unified Communications (UC) market. Here we present an overview of the Ericsson BCS package in the very important enterprise communications segment.
Security is a fundamental building block of wireless telecommunications systems. It is also a process – new threats are discovered over time, forcing communication systems to evolve.

**Wireless telecommunications systems must live up to user and network service provider expectations regarding trust and privacy. Besides the obvious need for authentication and encryption (to enable reliable charging, and to hinder eavesdropping), new architectures require more sophisticated protection mechanisms.**

The authors provide an overview of the security architecture and security features of the Evolved Packet System for LTE and non-3GPP accesses.

**Background**

The fundamental needs for authentication and encryption were addressed in the design of the Global System for Mobile Communication (GSM), which mitigated problems in earlier wireless telecommunications systems and helped to make GSM a widely successful system.

The design of the Universal Mobile Telecommunications System (UMTS) retained the good security features of GSM and introduced new ones, including:

- Public review of encryption algorithms by the security community;
- The MILENAGE algorithm set, which was specified by 3rd Generation Partnership Project (3GPP) as an example authentication algorithm to be used by network service providers who did not want to design their own algorithm;
- 128-bit encryption key length (increased from 64 bits);
- Mutual authentication and mandatory integrity protection of signaling between wireless terminals and the network – this feature was added to protect against a false base station; and
- Encryption from the terminal to a node beyond the base station.

In 2004, 3GPP started work on the next-generation radio technology, called Long Term Evolution (LTE). The main drivers of this work were the needs to increase capacity and throughput, and to decrease latency. Work was also started on the Evolved Packet Core (EPC), with the aim of simplifying the core network, and to integrate non-3GPP access technologies with the EPC.

Work on EPS security began in 2005. It was based on the strong security from UMTS, but has in fact improved security even further.

**Long Term Evolution**

The trust model in LTE (Figure 1) is similar to that of UMTS. It can roughly be described as a secure core network while radio access nodes and interfaces between the core network and the radio access nodes are vulnerable to attack.

The system architecture for LTE is flatter than that of UMTS, having no node that corresponds to the radio network controller (RNC) in UMTS. Therefore, the user equipment (UE) user plane security must either be terminated in the LTE base station (eNB) or in a core network node. For reasons of efficiency, it has been terminated in the eNB. However, because eNBs and backhaul links might be deployed in locations that are vulnerable to attacks, some new security mechanisms have been added.

Security over the LTE air interface is provided through strong crypto-
graphic techniques. The backhaul link from the eNB to the core network makes use of internet key exchange (IKE) and the IP security protocol (IPsec) when cryptographic protection is needed. Strong cryptographic techniques provide end-to-end protection for signaling between the core network and UE. Therefore, the main location where user traffic is threatened by exposure is in the eNB.

Moreover, to minimize susceptibility to attacks, the eNB needs to provide a secure environment that supports the execution of sensitive operations, such as the encryption or decryption of user data, and the storage of sensitive data like keys for securing UE communication, long-term cryptographic secrets and vital configuration data. Likewise, the use of sensitive data must be confined to this secure environment.

Even with the above security measures in place, one must consider attacks on an eNB, because, if successful, they could give attackers full control of the eNB and its signaling to UEs and other nodes. To limit the effect of a successful attack on one eNB, attackers must not be able to intercept or manipulate user and signaling plane traffic that traverses another eNB — for example, after handover.

**User authentication, key agreement and key generation**

The subscriber-authentication function in LTE/3GPP Evolved Packet System (EPS) is based on the UMTS authentication and key agreement (UMTS AKA) protocol. It provides mutual authentication between the UE and core network, ensuring robust charging and guaranteeing that no fraudulent entities can pose as a valid network node. Note that GSM Subscriber Identity Modules (SIMs) are not allowed in LTE because they do not provide adequate security.

EPS AKA provides a root key from which a key hierarchy is derived. The keys in this hierarchy are used to protect signaling and user plane traffic between the UE and network. The key hierarchy is derived using cryptographic functions. For example, if $key_1$ and $key_2$ (used in two different eNBs) are keys derived from $key_0$, a mobility management entity (MME), an attacker who gets hold of, say, $key_2$, still cannot deduce $key_0$, or $key_1$, which is on a higher layer in the key hierarchy. Furthermore, keys are bound to where, how and for which purpose they are used. This ensures, for example, that keys used for one access network cannot be used in another access network, and that the same key is not used for multiple purposes or with different algorithms. Because GSM does not have this feature, attackers who can break one algorithm in GSM can also compromise the offered security when other algorithms use the same key.

Further, the key hierarchy and bindings also make it possible to routinely and efficiently change the keys used between a UE and eNBs (for example, during handover) without changing the root key or the keys used to protect signaling between the UE and core network.

**Signaling and user plane security**

For radio-specific signaling, LTE provides integrity, replay protection, and encryption between the UE and eNB. IKE/IPsec can protect the backhaul signaling between the eNB and MME. In addition, LTE-specific protocols provide end-to-end protection of signaling between the MME and UE.

For user-plane traffic, IKE/IPsec can similarly protect the backhaul from the eNB to the serving gateway (S-GW). Support for integrity, replay protection, and encryption is mandatory in the eNB. The user-plane traffic between the UE and eNB is only protected by encryption as integrity protection would result in expensive bandwidth overhead. Notwithstanding, it is not possible to intelligently inject traffic on behalf of another user: attackers are essentially blind in the sense that any traffic they try to inject would almost certainly decrypt to garbage.

**Handover in LTE**

When handover occurs between two eNBs, the source eNB needs to transfer security parameters to the target eNB (Figure 2). At the same time, there might be a need to

- restore security should the source
eNB have been compromised (forward security); or
keep previous traffic secure should the target eNB have been compromised (backward security).
In either case, the core network can provide the target eNB with a new key, unknown in the source eNB, to be used after handover. An attacker who has compromised one of the eNBs and obtained its key will not know which key will be (or has been) used in the other eNB. The UE, on the other hand, has all the information it needs to deduce the correct keys.

A simpler procedure also used in LTE, which ensures only backward security, is to have the source eNB derive a new key from the current key via a cryptographic function. Only the derived key is transferred to the target eNB.

**Handover to legacy systems**

When a UE moves between LTE and other 3GPP radio access technologies, the security context may also be transferred in much the same way as when a UE moves between GSM/EDGE Radio Access Network (GERAN) and Universal Terrestrial Radio Access Network (UTRAN). LTE also includes the caching of security contexts. This saves on the number of times a subscriber must be authenticated when a UE rapidly moves back and forth between LTE and UTRAN.

**Non-3GPP access**

3GPP Rel-6-enabled access through interworking Wireless Local Area Network (WLAN) technology gives users internet access via a (U)SIM-based subscription. To enable the use of (U)SIM cards, the AKA protocol is carried by the extensible authentication protocol (EAP) and IEEE 802.1X. Knowing that legacy WLAN technology (such as IEEE 802.11b) has sub-optimal security, 3GPP has also allowed user traffic to be tunneled across the access network using IKE/IPsec.

EPS/3GPP Rel-8 takes this concept one step further, enabling end users to use common security and mobility protocols based on Internet Engineering Task Force (IETF) specifications to access the EPC over basically any non-3GPP wireless or wireline access technology. Rel-8 also defines optimized handover
between LTE and high-rate packet data (HRPD) access developed by 3GPP2.

The disparity between the security solutions offered for the different access technologies was an immediate challenge. Wireline accesses (xDSL), for example, employ a security model that relies on the physical security of the wire, thus omitting user-specific credentials and cryptographic protection.

A heterogeneous patchwork of security solutions needed to be avoided, as did a “one-size-fits-all” approach, since this might overprotect accesses with good security and lead to sub-optimal performance. Instead, a common framework has been introduced with a simple security classification for different accesses.

**Principles**

The common denominator for any non-3GPP access to EPC is the use of a USIM card. EAP AKA-based mutual authentication is always performed between a UE (USIM) and the authentication, authorization and accounting (AAA) server. The AAA server fetches credentials from the home subscriber server (HSS). EAP AKA authentication also provides cryptographic keys for data integrity and encryption between the UE and network at the access layer, at the Internet Protocol (IP) layer, or both. The EAP AKA protocol has been extended to support keys that are bound to the identity of the access network. This limits the risk of key misuse, as discussed above.

Next, a given non-3GPP access can be treated either as a trusted non-3GPP access or as an untrusted non-3GPP access. An access is trusted if it can provide all necessary security itself. Untrusted accesses need IPsec tunneling (similar to the Rel-6 inter-working WLAN solution). The issue of trust is not solely a matter of access technology, however; network service provider A, for example, might trust a given access network, whereas network service provider B might not. Figure 3 gives an overview of the architecture for non-3GPP access to EPC.

The access-level authentication or security association is optional for untrusted access. Since the access is not trusted, it is not clear whether the security provided by the access would add anything. Instead, mandatory EAP AKA authentication provides an IKE/IPsec security association between the evolved Packet Data Gateway (ePDG) and UE, protecting all traffic across the entire access network.

EAP AKA is used for trusted access to create an access-level security association between the UE and the non-3GPP access network. However, this access-level authentication is optional when mobility is based on the dual-stack mobile IPv6 protocol (DSMIPv6). This is because DSMIPv6 always uses EAP AKA authentication between the UE and MIP home agent (the packet data network gateway, PDN-GW), which adequately fulfills the authentication needs.

Because the security procedures for trusted and untrusted accesses differ, the UE needs to know the “trust value” of the access. This can be made available via the authentication signaling. If no signaling is received, the UE inspects a configuration file on the USIM to determine the trust value. If the UE does not find the access network identity there either, it reverts to a default and assumes the access is untrusted.

**Box B**

**Key sizes and algorithms in LTE.**

At present, the integrity protection and encryption algorithms use 128-bit keys. However, the system is prepared to use algorithms with 256-bit keys. LTE uses the following encryption algorithms:

1. 128-EEA1 based on the SNOW 3G algorithm. It is identical to UEA2 as specified for UMTS.
2. 128-EEA2 based on the advanced encryption standard (AES) in counter mode.

Likewise, LTE uses the following integrity protection algorithms:

1. 128-EIA1 based on SNOW 3G. This algorithm is identical to UIA2 in UMTS.
2. 128-EIA2 based on Advanced Encryption Standard (AES) in cipher-based message authentication code (MAC) modes.

HRPD is one non-3GPP access that has been treated in a special way in 3GPP standardization in order to fulfill strict performance requirements for mobility between LTE and HRPD.

A delay-optimized handover between LTE and HRPD could be handled in the same manner as handover between LTE and GERAN/UTRAN — for example, by transferring the security parameters in use. However, since HRPD in CDMA is not part of the 3GPP family of accesses, the mapping of security parameters between HRPD and LTE is not straightforward.

Another option is to perform new security procedures every time the UE enters a new access. But if the UE in question is only able to operate one radio at a time, this option suspends all traffic while the UE performs security procedures in the target access.

One straightforward approach to reduce the delays is that before attaching to the LTE access, the UE performs HRPD access attachment procedures (including security procedures) directly over the HRPD radio access. The UE can then attach to LTE over LTE radio access and continue using LTE. In this case the security context in HRPD access is prepared and cached and a later handover from LTE to HRPD can be performed more efficiently.

Another alternative is for the UE to perform attach-and-security signaling with the target HRPD access while it is still active in LTE access. In this case the LTE access network transparently forwards HRPD-specific signaling (including EAP-AKA) between the UE and HRPD access network (Figure 4). The LTE network needs not be aware of HRPD-specific messages or parameters (and vice versa). Also in this case the target HRPD access network is prepared when the UE executes a handover to HRPD. This alternative is however more complex and requires additional functionality in the LTE and HRPD access networks.

**Summary and future outlook**

EPS security has adopted UMTS security as a baseline, drawing on the successful concepts of UMTS to build an even more secure and flexible solution. The most prominent components are:

- the EAP AKA procedure with a key hierarchy, in which keys are bound to their use;
- prepared for stronger 256-bit cryptography;
- a new key-updating mechanism for intra-LTE handovers;
- backhaul security;
- resistance to attacks on eNBs; and
- integration of security for non-3GPP accesses.

These features address attacks from a false eNB, and confine the consequences of a compromised eNB to itself. The use of cryptographic algorithms means that signaling and user traffic is strongly protected over the air interface and backhaul.

Future 3GPP security work will focus on LTE Advanced. But because no big changes are anticipated with respect to core LTE functionality, the basic security mechanisms will continue to prevail. From a security point of view, the main new development will be relay nodes, and work in this area is already underway.

Other developments will include the introduction of home base stations,
machine-to-machine communication, local IP access, and selected IP traffic offload. Home base stations require new security solutions for authentication with the core network and authorization of end-user access to the radio cell. Home base stations will also be deployed in environments where it is easy to launch attacks on the physical hardware. Accordingly, the network must be able to detect such tampering. The first versions of related standards have already been released.

Local IP access will allow users to access local residential or corporate networks from a 3GPP device. Selected IP traffic offload deals with achieving a more optimal traffic path for user internet traffic which is not intended to reach services in the operator’s core network.

In addition there is work ongoing in 3GPP as well as in the Broadband Forum (BBF) for providing a more optimized interworking of BBF fixed access networks with the EPC. This work may result in additional security work to be done in 3GPP and BBF.
Bengt Sahlin joined Ericsson Finland in 2001. He has worked in the area of security ever since, and is currently Section Manager for the Network Security Section at Ericsson Research NomadicLab. Bengt holds an M.Sc. in computer science from Helsinki University of Technology, and is currently Ericsson’s standardization coordinator for security in 3GPP (SA3 working group).

Karl Norrman joined Ericsson Research in Kista, Sweden in 2001. His main areas of interest are security protocols and architectures. He holds an M.Sc. in computer science from Stockholm University in Sweden, and is currently Ericsson’s standardization coordinator for security in 3GPP (SA3 working group).

Mats Näsland worked for Ericsson since 1999, when he joined the then newly formed Communication Security Lab, now Research Area Security. His research interests cover most aspects of fixed and mobile network security, and in 2009 he received Ericsson’s Inventor of the Year award. Mats holds a Ph.D. in computer science from the Royal Institute of Technology (KTH) in Stockholm, Sweden.

Rolf Blom is an Expert in Mobile Communications Security at Ericsson Research in Kista, Sweden. He first joined Ericsson in 1984 and worked until 1995, mainly with the development of crypto and crypto-based products for defense communications. After a period of work abroad, he returned to Ericsson in 1998, and since then he has been active in establishing and leading security research activities. He holds an M.Sc. in electrical engineering from the Royal Institute of Technology in Stockholm, Sweden, and a Ph.D. in Information Theory from Linköping University, Sweden.

Stefan Rommer, who joined Ericsson in 2001, is a Senior Specialist in IP mobile networks at Business Unit Networks, Product Development Unit Packet Core. He has also contributed to various wireless LAN and mobile packet core projects. Since 2006, he has been working with packet core standardization in 3GPP (SA2 working group). Stefan holds an M.Sc. in engineering physics and a Ph.D. in Theoretical Physics, both from Chalmers University of Technology, Gothenburg, Sweden.

Mats Näslund worked for Ericsson since 1999, when he joined the then newly formed Communication Security Lab, now Research Area Security. His research interests cover most aspects of fixed and mobile network security, and in 2009 he received Ericsson’s Inventor of the Year award. Mats holds a Ph.D. in computer science from the Royal Institute of Technology (KTH) in Stockholm, Sweden.

Bengt Sahlin joined Ericsson Finland in 2001. He has worked in the area of security ever since, and is currently Section Manager for the Network Security Section at Ericsson Research NomadicLab. Bengt holds an M.Sc. in computer science from Helsinki University of Technology (TKK) in Finland, and is a certified information systems security professional (CISSP). He is also the current chairperson of 3GPP SA3.

References
5. Algorithm descriptions can be found at: http://gsmworld.com/our-work/programmes-and-initiatives/fraud-and-security/gsm_security_algorithms.htm#nav-
Evolution of the voice interconnect

The migration to all-IP networks calls for an evolution of the interconnect between operator networks. This might lead to a new interconnect architecture featuring an intermediate carrier network based on either a pure IP transit architecture or a service hub architecture.

During the past decade, network operators have transformed their existing wireline and wireless core networks from Time Division Multiplex (TDM) or Asynchronous Transfer Mode (ATM) transport technology to IP transport. The evolution of the voice interconnect between operator networks to IP is another natural step towards an all-IP architecture. This transition will result in more efficient bandwidth usage, better speech quality, reduced costs, and a clear migration path to a full multimedia service interconnection.

**Background**

**Traffic forecast**

During the past five years there has been a continuous increase of 10-15 percent per year in international interconnect traffic, and mobile operators are playing a more important role in interconnect for two main reasons.

Firstly, worldwide the number of subscribers in mobile networks is still increasing, mainly driven by the demand in developing and emerging markets with low fixed-line penetration. Secondly, there is an increased trend towards fixed-mobile substitution in developed countries.

In 2002, the number of mobile subscribers worldwide surpassed the number of fixed subscribers worldwide. The increases in subscribers and minutes of use have resulted in the mobile terminating traffic surpassing the fixed terminating traffic in 2009.

Although traffic growth continues, PC-based voice clients such as Skype have started to use the internet to bypass traditional international transit networks. These new entrants have managed to capture approximately 8 percent of international voice traffic volume.

The increased competition has resulted in a reduction of the international interconnect price of around 7 percent per year.

**Existing interconnect architecture**

Today’s interconnect architecture is predominantly TDM-based and is a mixture of direct interconnects between operator networks, typically used within a country, and indirect interconnects that use an intermediate global carrier network to reach the rest of the world. With the migration towards an IP-based interconnect, operators and carriers are looking at new architectures rather than simply replacing TDM pipes with IP pipes (Figure 1).

The transformation from existing ISDN User Part (ISUP)/TDM interconnects to an IP-based interconnect has the following advantages:

**Evolution of the voice interconnect:**

Using bandwidth-optimized codecs will reduce bandwidth needs, and using IP instead of TDM will ensure better utilization of the point of interconnect. Both will significantly reduce capital costs.
Expenditure (capex) and operational expenditure (opex).

- Interconnect architecture: IP technology offers higher flexibility in choosing interconnect partners and architecture.

- Multimedia interconnect: Operators have started launching new multimedia based services such as presence and messaging. One of the key success factors will be interoperability between operators. This will require evolving the interconnect from a pure voice interconnect to a combined voice and multimedia interconnect.

The following sections examine the three areas in more detail.

**Evolution of the voice interconnect**

Today, voice interconnect is handled using legacy TDM connections with ISUP signaling and the G.711 codec. Over the years ISUP signaling has matured and now supports national and international variants, as well as redundancy mechanisms required for telecom grade service.

However, when it comes to the introduction of new services and interworking with IP-based networks, ISUP has severe limitations and adds to the complexity of the solution.

**Figure 2** shows a Session Initiation Protocol with encapsulated ISDN User Part (SIP-I) interconnect between two MSS domains. When SIP-I is used, in addition to the protocol mapping done for SIP, the complete ISDN User Part (ISUP) message is transferred within the body of the SIP message. This allows the receiving SIP User Agent to process the encapsulated ISUP messages, so that no service-related information is lost for legacy functionality. Similar to ISUP, market variants are supported using SIP-I screening, which allows modification of messages, parameters and parameter fields from the encapsulated ISUP content.

**Figure 3** shows an SIP interconnect between an IMS and MSS domain. When the Gateway carries out protocol interworking between SIP and ISUP or Bearer Independent Call Control (BICC) or Radio Access Network Application Part (RANAP), it performs a mapping between the message parameters of the SIP and ISUP protocol. This mapping can result in information loss.

![Evolution of the interconnect from TDM to IP is the natural next step after the transition to IP within operator domains.](image1)

![Reference architecture of an SIP-I and SIP Network to Network Interface (NNI).](image2)
related to the requested service and does not support all legacy services.

For voice interconnect between networks there is more to consider than just the choice of signaling protocol. The following aspects should also be considered for interconnects based on SIP-I and SIP:

- **Signaling policing**: policing is needed to protect the domain against unwanted messages. For ISUP interconnects, policing is performed on the MTP3 or SCCP protocol level. For an SIP or SIP-I it is done by the Interconnection Border Control Function (IBCF) on the SIP/SIP-I signaling level and by the Trunking Gateway (TrGw) on the media protocol level.

- **Bandwidth management**: in the case of an ISUP interconnect, there is always a predefined bandwidth usage of 64Kbps or 56Kbps for G.711 per channel. Due to the fixed timeslots on the payload path it is not possible to use more than the assigned bandwidth. For an IP-based interconnect the bandwidth needed depends on the codec used. Also the user plane connections share a common IP pipe. This means that proper bandwidth allocations and admission control mechanisms are needed to avoid exceeding the allocated bandwidth or causing deterioration of the voice services.

- **Voice transcoding**: as more codecs are now available, it is important to introduce negotiation mechanisms to select the best codec for each situation. This will reduce the number of transcoding stages, resulting in a Transcoder Free Operation (TrFO) whenever possible.

- **Fax, Video and CS Data Calls**: fax and data calls represent only a small fraction of the total number of calls but must still work properly. In order to support compressed speech without disturbing fax and modem operation, voiceband data traffic detection and fallback to G.711 is desired. Also the T.38 fax gateway is desired to support proper fax interworking over poor quality IP transport connections.

- **Charging and Accounting**: SIP/SIP-I based interconnect will reuse the existing charging mechanism based on call data records.

- **Emergency and Priority calls**: interconnects need to support emergency and priority calls. Emergency and priority calls are indicated with a certain ISUP
parameter (Continuous Packet Connectivity or CPC) and a certain SIP parameter.

Operators may implement IP-based voice interconnect using different codec types (Table 1) and enjoy the benefits of each. These can also be seen as evolution steps.

- **Usage of IP transmission (still G.711)**
  If transmission efficiency is not so important, an operator may introduce IP-based interconnect without support for any additional codec above the default G.711 codec.

- **Usage of bandwidth efficient codecs**
  As a next step, codecs with better bandwidth characteristics can be introduced on the interconnect to reduce OPEX. For example, G.729 is the most commonly used fixed network codec and can be used for compression.

- **Removal of the need for transcoding (end-to-end codec negotiation)**
  If the same codec is supported within the two connected networks then it is desirable to support this codec also on the interconnect. This can eliminate the need for transcoding, or at least minimize the number of transcoding stages, ensuring resource efficiency in MGWs, which reduces capex and improves speech quality. As an example, AMR coded speech, which is commonly used by mobile networks, can be used on interconnect to achieve TrFO end-to-end when connecting two mobile networks. TrFO reduces the need for transcoders and therefore reduces capex.

- **Improved voice quality with High Definition (HD) voice**
  Once interconnect is on IP transport and TrFO is achieved, new codecs with superior speech quality, HD-voice, can be introduced also for inter-Public Land Mobile Network (PLMN) calls. AMR-WB, which is optimized for mobile access, and G.722, which is widely used for fixed access, can be used for superior speech quality. These codecs can be transcoded to each other without losing wideband characteristics. HD voice on interconnect enables superior speech quality for calls between mobile and fixed access and can help to increase revenue and reduce churn.

Table 2 summarizes which codecs are used in which domain. It is clear that G.711 is the only codec that is common for all domains. A common codec via end-to-end codec negotiation is possible with AMR-NB, AMR-WB for a MSS-MSS and MSS-IMS interconnect. This is not possible for an MSS-wireline softswitch interconnect. G.729 is the optimal choice in this case, but requires transcoding in MSS due to the unavailability of the G.729 codec in mobile phones.

**Interconnect architecture**

With the migration to all-IP networks and IP-based interconnect, operators and carriers are looking at new interconnect architectures, rather than simply replacing TDM pipes with IP pipes.

In the new architectures, the concept of an intermediate carrier network exists in two forms:

- **Pure IP Transit architecture**: In this case, the carrier network essentially provides an IP interconnection “pipe” between operators, which will have a bilateral agreement for settlement. The carrier network is agnostic to the type of application or media flowing through the network and will typically charge the operators based upon the amount of data transferred.

- **Service Hub architecture**: In this case, the carrier network is service aware and can differentiate the type of application or media that is flowing through the network for QoS and charging purposes. The carrier network can provide a complete multilateral hubbing service where operators set up single contracts with carriers for their complete interconnection needs.

The future IP interconnect architecture is likely to involve a combination of:

- **Direct IP interconnect between two mobile operators (National traffic)**
- **Indirect interconnect via a third-party carrier IP backbone (International traffic – first using pure IP transit)**
- **Indirect interconnect via a third-party carrier network that also is service aware (Service Hub)**

In its pure IP transit form, as illustrated in Figure 4, the carrier network

![Figure 4](image-url)
is unlikely that models based upon a sustainable interconnect solution. It is necessary for mobile/fixed operators and also for the global carriers. This is necessary for interconnect fees.

In terms of reduced administration and perspective, this can provide benefits having fewer and larger points of interconnection. This can be used for real-time services such as voice.

In its service hub form, the carrier network can provide additional services such as service-based inter-operator settlement, service based routing and service interworking. This allows fixed/mobile operators to delegate a large part of interconnect administration to their chosen carrier.

The different architectures give network operators flexibility in choosing the appropriate interconnect architecture for particular destinations according to traffic and commercial requirements. The usage of IP technology gives higher flexibility, such as having fewer and larger points of interconnection. From an economy of scale perspective, this can provide benefits in terms of reduced administration and interconnect fees.

These architectures provide well-defined technical and business roles for mobile/fixed operators and also for the global carriers. This is necessary for a sustainable interconnect solution. It is unlikely that models based upon interconnections over the public internet can provide such a sustainable interconnect solution.

There is a need for industry-wide coordination as operator network IP-based interconnects are emerging. The GSM Association is active in coordinating the industry’s thinking on IP-based interconnection. The GSMA suggests that the existing carrier GRX IP networks are evolved to support voice and multimedia services. The IPX project builds upon the GRX to add QoS and so supports real-time CS and IMS services such as voice, video and presence. It is expected that within IPX there will be a service hub role for CS and IMS services. The service hub role is beneficial because an operator only has to establish one agreement with the IPX hub provider who then has agreements with all other operators. To provide a service hub function for these real-time services SIP Servers will be introduced into the IPX core.

**Multimedia interconnect**

Once IP-based interconnect for voice is established, there is a natural evolution from voice to multimedia across interconnects, shown in **Figure 5**. The Rich Communication Suite (RCS) initiative from GSMA provides enrichment to voice services. RCS also provides end-user services such as presence, instant messaging and video sharing.

The value of the RCS services will increase with the number of users and also with the number of interconnected networks (Metcalfe’s law). Operators benefited from this correlation in the past, when they started their SMS and MSS services, and at the same time secured interoperability. This requires RCS services to be supported across operator interconnects. These interconnects could be direct between two operators, or indirect using an intermediate carrier IP network, such as IPX.

RCS is a set of multimedia services based upon 3GPP IMS using SIP signaling and RTP and MSRP media. The interconnect needs to support these signaling and media protocols. RCS is likely to be the first service that makes IP interconnect mandatory, simply because TDM networks cannot transport the media types defined in RCS.

The different types of media in RCS have different bandwidth and delay requirements – video requires high bandwidth for media, presence has a high signaling rate. It is not sufficient to base inter-operator settlement upon packet count at IP level. These services require new settlement arrangements between operators and new KPIs to be defined for interconnects. There could be different settlement policies for different RCS services. The session border gateway has an important role to play in policing KPIs.

In effect, operator interconnects and intermediate carrier networks need to be service aware, both in terms of technical realization and commercial settlement mechanisms. This will allow RCS services to be deployed at national and international level, and will be the basis for the evolution of these networks and interconnects towards the full MMTel service.

With the introduction of new multimedia services there are benefits to be gained from using the service-aware hub carrier network. The service hub can provide interworking between services both at the media layer (transcoding) and also at the service layer – for example, interworking between legacy SMS/MMS and IMS messaging.

---

**Figure 5** Combined voice and multimedia interconnect.
**Conclusion**

The evolution of the voice interconnect creates opportunities for the operators to:

- reduce capex/opex by using IP technology with more efficient codecs
- optimize voice quality by reducing the transcoding steps and introducing new HD audio codecs
- reduce transmission costs by selecting a new interconnect architecture
- increase revenue by providing multi-media interworking between operator networks.

Ericsson believes that the evolution of the voice interconnect is an important part of the overall evolution of telephony. Ericsson is working with the major operators, carriers and standards bodies to ensure that this evolution is successful and that end users and the industry can benefit from the opportunities described above.

---

**John Baldwin**
- joined Ericsson in 2006 following the Marconi acquisition. He first worked for the Systems and Technology group within the Core & IMS unit of Business Unit Networks on strategies for connecting the mobile softswitch (MSS) to IMS networks. He was then involved in IP-based interconnects for MSS and IMS networks. At present, he is involved in the study of IP-based interconnects and new applications that use IMS networks and the internet. He holds a B.Sc. in mathematics from the University of Warwick, UK.

**Soner Yamen**
- joined Ericsson in 1995 as a support engineer. In 1999, after having served as a solution and project manager in the market unit, he transferred to system management at Ericsson Eurolab. In 2000, he joined product management, where he is currently responsible for the functional areas Transport and Network Quality & Efficient Infrastructure in product line switching at Business Unit Networks. He holds a M.Sc. in electrical engineering from the University of Wisconsin-Madison, US.

**Jörg Ewert**
- joined Ericsson in 1999 as a system manager for MSS operation and maintenance. He later became the leader of the O&M system design team. In 2005 he joined the MSS product management team, where he was responsible for network management and security. At present, he is responsible for MSS and IMS co-existence and SIP interconnects in product line switching at Business Unit Networks. Jörg has studied business administration at the University of Hagen, Germany, and he holds a M.Sc. and a Ph.D. in physics from the University of Göttingen, Germany.

---

**References**

2. ITU Q1912.5, 3GPP TS 29.164
3. 3GPP TS 29.163
Scalable parallelism using dataflow programming

Data flows and dataflow programming are attractive candidates for both modeling and designing parallel computing systems for base stations and mobile terminals.

Dataflow programming was invented to address the issue of parallel computing. Dataflow models are well suited to describe many forms of computation, particularly in the area of digital signal processing.

This article presents case studies about an MPEG-4 video decoder and a multi-standard OFDM receiver. Hardware and software have been synthesized from the dataflow models, and experimental results are presented.

Introduction

Support for multiple wireless standards, such as GSM, WCDMA and LTE, is becoming an essential feature for wireless telephony systems. Similarly, mobile terminals must support multiple media formats, such as different MPEG-4 profiles. Keeping power consumption and production costs at attractive points in the design space, while meeting the throughput requirements of such diverse functionality, is a considerable challenge. The trend is to move from fixed-function hardware to configurable and programmable solutions with a higher degree of shared resources and increased software content. Although parallel software is difficult to develop, let alone verify, the so-called “power wall” makes parallel software deployed on parallel architectures a necessity.

Models for functional programming and dataflow programming—the focus of this article—have the potential to facilitate parallel programming. To make this concept practical, however, development tools must evolve.

Signal processing is often modeled as flows of data. A circuit diagram, for instance, is a form of dataflow description, especially at the block diagram level. Many forms of computation are well suited to dataflow description and implementation. Some common examples include complex media coding, network processing, imaging and digital signal processing, and embedded control.

Digital radio is one application domain that can be successfully modeled using data flows. Dataflow modeling facilitates component isolation and “pluggable” processing chains. Performance-enhancing functionality, such as hybrid automatic repeat request (HARQ), can be implemented as an optional plug-in. Given that each new generation of mobile terminals is expected to support a larger number of radio standards, more functionality is being implemented in software or programmable hardware. Building a processing chain from pluggable modules—either standard-specific or universal with parameters—is therefore an attractive proposition.

Dataflow modeling is not only a means of creating transceiver implementations. The use of dataflow modeling can also reveal bottlenecks from a lack of parallelism inherent in a radio standard. This is especially useful when developing new standards. To support a wider range of maximum communication speeds, next generation radio standards will be scalable in terms of capacity and processing requirements. Ideally, it should be possible to build a high-capacity transceiver by duplicating the hardware of a baseline transceiver. To do this, the potential parallelism should scale in a linear way with processing requirements.

Parallel programming is not a new
proposition or challenge. At the end of the 1960s, researchers were struggling to make the transition to parallel computing. At that time, processing performance doubled about every 18 months, a trend that continued for the next three decades. For the most part, increasing clock frequencies drove this development. As a result of the rapid advance in performance, sequential programs proved sufficient for many computing disciplines, but this computer hardware development trend is over. Ever-increasing clock frequencies came to an abrupt end in 2005 due to, among other things, fundamental problems associated with power supply and heat.

Future advances in performance are expected to come from greater use of parallelism, which is facilitated by improved transistor density. Multicore computers are an example of this development, and forecasts for other sectors point in the same direction. For portable consumer electronics, the International Technology Roadmap for Semiconductors (ITRS) 2009 edition foresees exponential growth in parallelism, resulting in hundreds of parallel processing elements within five to ten years.

To leverage this development, the performance of an application or system needs to be scalable through parallelism. However, most applications have been built to accommodate the tradition of constantly increasing clock frequencies and are inherently sequential and thus not susceptible to parallelization, presenting a challenge to the software industry. In contrast, the telecom sector is in a good position, thanks to its practice of designing parallel systems. The task is to make efficient use of increasing parallelism, rather than enabling parallel execution.

Several research centers have been set up and many projects have been launched to address the challenges of multicore processors. Examples include the Parallel Computing Laboratory (ParLab) at the University of California at Berkeley, the Pervasive Parallelism Laboratory (PPL) at Stanford University and the HiPeac European Network of Excellence. Collaborations with external partners include ACTORS and DSL4DSP. Ericsson also actively researches in this area.

**Current practices**

C and related programming languages are dominant for software in consumer electronics, embedded multimedia and communications equipment. However, C’s control over low-level detail, usually considered a good thing, tends to overspecify programs. Besides specifying algorithms, C specifies how inherently parallel computations are sequenced, how inputs and outputs are passed between algorithms and, at a higher level, how computations are mapped to threads, processors and application-specific hardware.

Using analysis, it is not always possible to recover the original knowledge about the program and the chances for restructuring transformations are limited. Therefore, C is not a good starting point for parallelization.

Tools and frameworks, such as OpenMP and the message-passing interface (MPI), are used to facilitate the construction of parallel C programs for multicore and multiprocessor systems. These tools are, however, severely limited, as they can only expose parallelism when it has been explicitly identified by a programmer and they do not support programmers in assessing the correctness of parallelization.

In hardware design, the current practice is to create a low-level description, also referred to as Register Transfer Logic (RTL). The abstraction level of RTL tends to be low, which results in long development cycles impacting time to market.

Great gains in hardware efficiency can be made from architectural optimization. For this reason, it is desirable to raise the abstraction level of hardware descriptions, and as a consequence, the use of high-level tools.

Because RTL is inherently parallel, it is easy to translate a parallel description in a higher-level language into RTL. Unfortunately, descriptions in traditional programming languages, such as C, do not map equally well because of the difficulty in resolving the sequential nature of traditional programming languages into parallel hardware.

**Dataflow programs**

A dataflow program is defined as a directed graph, where nodes represent computational units or actors, and arcs represent the flow of data – communication channels that connect the actors. An actor is solely concerned with mapping its input to output. Dependencies are expressed by connecting actors; there is no other source of dependence. These properties make dataflow programs very flexible in terms of partitioning and sequencing of computations.

It has been shown that dataflow models offer a representation that can effectively support the tasks of parallelization and vectorization, thus providing a practical means of supporting multiprocessor systems and utilizing vector instructions. Interestingly, as dataflow programs grow in size they tend to expose more parallelism.

In parallel computing, a distinction is made between

- parallelism that scales with the size of the problem, known as data parallelism and
- parallelism that scales with the size of the program, known as task parallelism.

Scaling an algorithm over larger amounts of data is a relatively well-understood problem that applies to dataflow programs, as well as other programming models. While a dataflow program has a straightforward parallel composition mechanism, it is difficult, for example, to compose a C program, whose parts execute concurrently without interference.

There are several classes of dataflow programs. These classes differ in expressiveness and analyzability. At one end of the spectrum are Kahn process networks, which can express any kind of computation. Though not all the interesting properties of these networks can be established by program analysis (Kahn process networks are Turing complete). At the other end are synchronous dataflow networks, for which static schedules and bound memory requirements can be determined in order to make statements about the absence of deadlocks. The drawback is limited expressiveness, as synchronous dataflow cannot express control flow, such as input-dependent iteration.

Constrained dataflow models, such as synchronous dataflow, can be synthesized into particularly efficient...
Scaling performance to match resources

CAL actor language

The CAL actor language (CAL) is a domain-specific language that provides useful abstractions for dataflow programming with actors. It has been used in a wide variety of applications and compiled to hardware and software implementations. Work on mixed hardware and software implementations is ongoing.

The basic structure of a CAL actor is shown in the Add example below, which has two input ports, A and B, and one output port, Out. The actor contains one action that, when fired, consumes one token on each of the input ports and produces one token on the output port. An action may fire if the availability of tokens on the input ports matches the port patterns, which in this example, corresponds to one token on ports A and B.

```plaintext
actor Add() int A, int B => int Out
  action A:[a], B:[b] => Out:[sum]
  var int sum = a + b
end
end
```

An actor can take any number of actions. The untyped Select actor (see below) reads and forwards a token from port A or B, depending on the evaluation of guard conditions.

```plaintext
actor Select() bool selector, int A, int B => int Out
```

The above examples are quite trivial, as typical actors can be several hundred lines of code. Sorting, for example, can be implemented with a pair of actors (one that contains two actions and another that contains four actions) together with a network that links several instances in a chain according to the amount of data to be sorted.

CAL enables several ways of expressing and controlling the flow of data. Two actors may never share a state and their only means of communication is via the actor ports. The runtime system is responsible for how and when an actor is scheduled. An actor is simply a specification that describes the actions that will occur in response to the presence of data. The CAL actor model allows a complete separation between the scheduling of actors and the algorithm that an actor’s network specifies. The same network can be scheduled in a variety of ways, all resulting in the same functional result, but with different timing and computational properties.

Actors can be written in a variety of languages. CAL eliminates some repetitive coding, but is otherwise of comparable efficiency, in terms of lines of code. Another advantage of using CAL is that it specifies the network connections outside the actor code. This simplifies the need for any structural tinkering with algorithms and ultimately leads to more efficient implementations.

Reconfigurable video coding

Multimedia processing can easily be modeled using data flows. A recent development within MPEG makes CAL a suitable language for experiments in this area. The MPEG reconfigurable video coding (RVC) framework is a new ISO standard that aims to provide video codec specifications at the level of library components, instead of monolithic algorithms. The basic idea is to specify a decoder by selecting components (or actors) from a standard library of coding algorithms. The ability to dynamically configure and reconfigure codecs calls for new methodologies and new tools for describing the bitstream syntaxes and parsers of the new codecs. The RVC framework uses the CAL actor language to specify the standard library and instantiation of the RVC decoder model. RVC is composed of two ISO/IEC specifications. For more information about RVC, see Overview of the MPEG reconfigurable video coding framework.

Globally asynchronous locally synchronous (GALS)

A coarse-grain asynchronous architecture, such as GALS, makes it possible to exploit benefits of synchronous, as well as asynchronous design methodologies. Ordinarily, clock skew in synchronous digital designs is limited by the implementation of a balanced clock tree. The goal of a balanced clock tree is to make all clock delays equal. However, achieving equal delays in the clock tree becomes increasingly difficult as system complexity increases.

GALS designs reduce clock skew constraints and give smaller clock trees, because clocking is not a global issue in GALS. GALS might reduce induced noise from the clock to the analog domain, as noise from digital parts comes from several smaller clock domains rather than from one large clock domain. The switching noise from clock nets is therefore spread out over time. Additionally, the lack of global synchronization makes it possible to save energy in the clock net.

There is a strong coupling between GALS and dataflow descriptions. The use of first in, first out (FIFO) channels
to communicate between processes in dataflow descriptions can translate directly into a GALS specification. The dataflow processes are implemented as synchronous blocks and the FIFO channels are implemented as asynchronous communication channels.

CAL dataflow networks support multiple clock domains, with up to one clock domain per actor. Replacing the FIFO channels between clock domains with asynchronous FIFO channels transforms the dataflow into a GALS network, making the CAL dataflow an attractive candidate for modeling GALS architectures. Because GALS designs consist of modules with a well-defined interface, GALS is also conducive to modular design. Work to try to transform CAL dataflow to a GALS network is ongoing.

Case studies

The following case studies illustrate different applications of dataflow programming, where both the hardware and software have been synthesized from CAL dataflow models.

OFDM inner receiver

A digital radio receiver can be divided into an analog front-end, a digital front-end, the inner receiver and the outer receiver. The inner receiver, as shown in Figure 1, demodulates the baseband signal, converting analog waveforms to codeword probabilities. In this example, parts of an inner receiver were implemented for OFDM-based radio systems, to test the dataflow methodology for digital radio.

A basic OFDM inner receiver consists of a synchronizer, frequency error compensator, fast Fourier transformer (FFT), channel estimator, equalizer and demodulator. Of these parts, the implementation comprised the frequency compensation, synchronizer and FFT.

- the synchronizer, which estimates time position and frequency error with a coordinate rotation digital computer (CORDIC) rotator to compensate for digital frequency error; and
- FFT – in this case, a configurable FFT that supports a maximum symbol length of 8k samples.

The specification language used was CAL. The dataflow description was synthesized to an off-the-shelf FPGA-based development platform using Open Dataflow (OpenDF) and OpenForge tools. Test data was streamed over Ethernet to the development board and the result was displayed on an attached VGA display. An on-chip processor handled the Ethernet streaming and data display. The synthesized hardware is capable of processing 50 Megasamples/s, which is sufficient for real time. For more details on the implementation, see Reconfigurable OFDM Inner Receiver Implemented in the CAL Dataflow Language.

It was possible to reach the stated performance goals, but the CAL implementation used more resources than a comparable RTL implementation, to a large extent due to the relatively crude tools. Essentially, there is a one-to-one relationship between the CAL code and the RTL, placing the burden of optimization on the RTL toolchain. Several sources of overhead have been identified in the CAL to RTL translation, and work is ongoing to drastically reduce this.

In another case study, an MPEG-4 decoder was specified in CAL and implemented on an FPGA. The code generated from the CAL specification outperformed a reference in VHDL (a hardware description language), achieving better performance with less hardware resources. The CAL implementation required significantly less development effort.

CAL encourages reasoning relating to interfaces and structure. Contributing factors are: strict isolation of the actor’s internals, the asynchronous token interfaces and the hierarchical network modelling. These features also limit the impact of actor modifications.

As a result, it is comparatively easy to restructure a system until it meets the desired performance requirements. Ultimately, this leads to significantly reduced developer effort for a given functionality, with minimal or no penalty in area or performance; particularly for large and complex problems.

However, for small systems with regular structure and wide data paths, like the FFT, the overhead of the flow control and action scheduler appear to overshadow the structural gains.

Multicore MPEG-4 decoder

In a second case study, an MPEG-4 simple profile decoder was implemented in software on a 200MHz quad-core ARM11 MPCore, using the code-generation tools developed within the ACTORS project. The dataflow model was taken from the RVC toolbox.

![Dataflow model of the MPEG-4 simple profile decoder.](image)
The main functional units – themselves hierarchical compositions of actor networks (Figure 2) – consisted of:

- bitstream parser;
- reconstruction block;
- 2D inverse cosine transform;
- frame buffer; and
- motion compensator.

Efficient realization of a dataflow model on a multicore architecture, involves the challenge of how to partition the workload. While the semantics of the dataflow model allow any of the actors to be deployed on any of the cores without violating data dependencies, the way actors are partitioned impacts greatly on system performance. Two (sometimes conflicting) objectives that need to be considered are balanced load over the cores and minimal communication cost. Partitioning can be performed statically (offline) or dynamically (at runtime).

Naive execution of a dataflow model is associated with significant overhead. In the first, naive, attempt, each actor was executed in a separate thread, synchronized with FIFOs in shared memory. This way, the scheduler of the operating system (SMP Linux) managed load dynamically. In a second attempt, the actors were partitioned statically and each core executed a single worker thread. This approach improved performance by several orders of magnitude. Clearly the first approach suffered badly from the overhead of context switches. Efforts are being made to pursue the idea of partitioning actors dynamically, but as yet, a competitive solution has not been devised.

Figure 3 shows the results of partitioning the MPEG decoder onto one through four cores. Good speed-ups were achieved by increasing the number of cores. The partitions were, however, found manually, meaning that this solution remains unexplored for the most part. More advanced tool support would further speed up the search process and greatly improve the quality of partitions.

There is room for improvement when it comes to frame rate. The current solution is still immature, compiling actors separately and treating them as separate entities, even at runtime. Given the fine level of granularity to which the MPEG decoder is specified, the overhead of this execution model is high relative to the useful work it performs. The Model Compiler, a recent development of the ACTORS project, allows several actors to be synthesized jointly. For an earlier position on the model compiler, see Efficient realization of a CAL video decoder on a mobile terminal. For classic work in this area see references 7 and 20. It is anticipated that joint synthesis of actors will significantly improve performance by reducing the overhead associated with data transport and scheduling.

Conclusions

The historic trend of ever-increasing clock frequencies has come to an end. In the future, gains in performance will be derived from increased parallelism. Therefore, systems whose performance scales with increased parallelism make an attractive proposition.

A particular problem with current practices is over-specification. Software tends to be specified in such a way that computations are serialized, failing to expose the potential of parallelism. Low-level specification of hardware (such as RTL) is costly to develop and the specification of timing complicates the refactoring of designs.

Dataflow models are well suited for describing many forms of computation, particularly in the digital signal processing domain. As such, dataflow programming is an interesting approach that can be applied to highly parallel computing systems found in mobile terminals and base stations.

A dataflow program is composed of computational units called actors that may but not necessarily need to execute in parallel. This very flexible programming model allows for exploration of the design space. In particular, a dataflow program can be partitioned into:

- separately timed hardware components – which is useful in the context of partly asynchronous GALS architectures; or
- separate software threads to be executing on a multicore CPU.

Dataflow models have been used in two case studies to specify an MPEG-4 video decoder and a multi-standard OFDM receiver. The hardware and software were synthesized from the models. Much of this work was performed within the European FP-7 ACTORS and MultiBase projects, using tools and methodologies made available by the OpenDF initiative. MPEG RVC models were used in the video decoder case.
The ability to make rapid design iterations was of great value when developing hardware from the OFDM model. It was possible to explore an array of design ideas on the architectural level limited development time. The performance targets were met, but the resource requirements (corresponding to silicon area) exceeded those of a comparable RTL design. The software, which was synthesized from the MPEG RVC model, could be partitioned so that performance scaled with the number of available cores. However, performance on a single core is much lower than that of a comparable implementation in C.

The case studies illustrated some interesting properties of dataflow programming. In particular, the models could be partitioned and refactored in a flexible manner. Further improvement of the OpenDF tools is required to make the synthesized code more efficient. It appears that an increased multiplexing of hardware resources and a reduction of the execution overhead in software would greatly improve the usefulness of the tools.

References

3. www.actors-project.eu
4. dsl4dsp.inf.elte.hu
5. The OpenMP API Specification for Parallel Programming, http://openmp.org
21. www.multibase-project.eu
22. www.opendf.org
Next generation LTE, LTE-Advanced

Next generation LTE, LTE-Advanced or LTE Rel-10 is the next step in radio access technology.

> STEFAN PARKVALL, ANDERS FURUSKÄR, ERIK DAHLMAN

Whatever the name – next generation LTE, LTE-Advanced or LTE Rel-10 – the next step in LTE evolution allows operators to introduce new technologies without putting existing investments at risk.

LTE radio access technology is continuously evolving to meet the requirements of regulators, operators and users. The first fully commercial and operational 4G mobile broadband systems, currently being deployed, are based on the first release of LTE, 3GPP Rel-8, which was finalized in 2008.

Rel-9, finalized at the end of 2009, added support for broadcast/multicast services, positioning services, and enhanced emergency call functionality, as well as enhancements for downlink dual-layer beamforming.

Today, the main focus of 3GPP is the next generation of LTE evolution, Rel-10, often referred to as LTE-Advanced. Rel-10 further extends the performance and capabilities of the LTE radio access technology, and meets all of the requirements for IMT-Advanced as defined by ITU1,2.

In October 2010, ITU completed the assessment of submissions for global 4G mobile wireless broadband technology. LTE Rel-10 (submitted by 3GPP) was one of two technologies accorded the official designation of IMT-Advanced.

This article provides a brief introduction to IMT-Advanced, followed by a description of the extensions to LTE introduced as part of 3GPP Rel-10. It concludes with system-level results that illustrate the ability of LTE Rel-10 to fulfill and even surpass the IMT-Advanced requirements.

**ITU and IMT-Advanced**

IMT-Advanced is the term used by ITU for radio access technologies beyond IMT-2000. An invitation to submit candidate technologies for IMT-Advanced was issued by ITU in 20083. Together with this invitation, ITU defined a set of requirements to be fulfilled by any IMT-Advanced candidate technology4, some of which are shown in Table 1.

Anticipating the invitation from ITU, 3GPP initiated a study in March 2008 on LTE-Advanced, with the task of defining requirements and investigating potential technology components for the LTE evolution. Ericsson was very active in the 3GPP study, which was completed in 2009 and formed the basis for the 3GPP Rel-10 work on LTE5.

**LTE Rel-10**

LTE-Advanced is not a new radio access scheme distinct from LTE, but simply the evolution of LTE, providing improved performance and service capabilities. LTE Rel-10 includes all of the features of Rel-8/9 and several new ones, the most important of which are:

- carrier aggregation;
- enhanced multi-antenna support; and
- improved support for heterogeneous deployments, and relaying.

**Carrier aggregation**

The first releases of LTE provided extensive support for deployment in spectrum allocations of various characteristics, with transmission bandwidths ranging from 1.4MHz up to 20MHz in both paired and unpaired bands. In Rel-10, the transmission bandwidth can be further extended with so-called carrier aggregation (CA)3 where multiple compo-

---

**Box A Terms and abbreviations**

<table>
<thead>
<tr>
<th>3GPP</th>
<th>4G</th>
<th>4th Generation mobile wireless standards</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARQ</td>
<td>automatic repeat request</td>
<td></td>
</tr>
<tr>
<td>BS-to-RN</td>
<td>base station to relay node</td>
<td></td>
</tr>
<tr>
<td>CA</td>
<td>carrier aggregation</td>
<td></td>
</tr>
<tr>
<td>CSG</td>
<td>closed subscriber group</td>
<td></td>
</tr>
<tr>
<td>CSI</td>
<td>channel-state information</td>
<td></td>
</tr>
<tr>
<td>DFT</td>
<td>Discrete Fourier Transform</td>
<td></td>
</tr>
<tr>
<td>DL-related</td>
<td>downlink-related</td>
<td></td>
</tr>
<tr>
<td>E-UTRA</td>
<td>Evolved Universal Terrestrial Radio Access</td>
<td></td>
</tr>
<tr>
<td>FDD</td>
<td>frequency-division duplex</td>
<td></td>
</tr>
<tr>
<td>HARQ</td>
<td>hybrid ARQ</td>
<td></td>
</tr>
<tr>
<td>HSPA</td>
<td>High-Speed Packet Access</td>
<td></td>
</tr>
<tr>
<td>ICIC</td>
<td>inter-cell interference coordination</td>
<td></td>
</tr>
<tr>
<td>IMT</td>
<td>International Mobile Telecommunications</td>
<td></td>
</tr>
<tr>
<td>InH</td>
<td>indoor hotspot</td>
<td></td>
</tr>
<tr>
<td>ITU</td>
<td>International Telecommunication Union</td>
<td></td>
</tr>
<tr>
<td>ITU-R</td>
<td>ITU Radiocommunication Union</td>
<td></td>
</tr>
<tr>
<td>LTE</td>
<td>Long-Term Evolution</td>
<td></td>
</tr>
<tr>
<td>MAC</td>
<td>Medium Access Control</td>
<td></td>
</tr>
<tr>
<td>MBSFN</td>
<td>Multicast-Broadcast Single Frequency Network</td>
<td></td>
</tr>
<tr>
<td>OFDM</td>
<td>orthogonal frequency-division multiplexing</td>
<td></td>
</tr>
<tr>
<td>PHY</td>
<td>physical layer</td>
<td></td>
</tr>
<tr>
<td>RF</td>
<td>radio frequency</td>
<td></td>
</tr>
<tr>
<td>RLC</td>
<td>Radio Link Control</td>
<td></td>
</tr>
<tr>
<td>RMa</td>
<td>rural macro</td>
<td></td>
</tr>
<tr>
<td>TDD</td>
<td>time-division duplex</td>
<td></td>
</tr>
<tr>
<td>UE</td>
<td>user equipment</td>
<td></td>
</tr>
<tr>
<td>UL-related</td>
<td>uplink-related</td>
<td></td>
</tr>
<tr>
<td>UMa</td>
<td>urban macro</td>
<td></td>
</tr>
<tr>
<td>UMi</td>
<td>urban micro</td>
<td></td>
</tr>
<tr>
<td>WCDMA</td>
<td>Wideband Code Division Multiple Access</td>
<td></td>
</tr>
</tbody>
</table>
Component carriers are aggregated and jointly used for transmission to/from a single mobile terminal, as illustrated in Figure 1. Up to five component carriers can be aggregated, allowing for transmission bandwidths up to 100MHz. Backward compatibility is ensured as each component carrier conforms with the Rel-8 carrier structure. Consequently, to a Rel-8/9 terminal, each component carrier will appear as an LTE Rel-8 carrier, while a carrier aggregation-capable Rel-10 terminal can exploit the total aggregated bandwidth, thus achieving higher data rates.

In general, a different number of component carriers can be aggregated for the uplink and downlink. We can generally expect that a terminal will have different aggregation capabilities in the uplink and downlink directions.

There are three cases in terms of the frequency location of the different component carriers:

- intra-band aggregation with contiguous carriers (#2 and #3 in Figure 1);
- inter-band aggregation (#1 and #4 in Figure 1); and
- intra-band aggregation with non-contiguous carriers (#1 and #2 in Figure 1).

Aggregating non-adjacent component carriers means that the fragmented spectrum can be utilized, which in turn allows operators to provide high data rate services based on the availability of a wide overall bandwidth, even without a single wideband spectrum allocation.

From a baseband perspective, there is no difference among the three different aggregation alternatives, with LTE Rel-10 supporting them all. However, the complexity of the RF implementation varies greatly, with the first case being the least complex. As a result, while spectrum aggregation is supported by the basic Rel-10 specifications, implementation will be strongly constrained and will include specification of only a limited number of aggregation scenarios; only the most advanced terminals will support aggregation over a dispersed spectrum.

Figure 2 shows how scheduling...
and hybrid-ARQ retransmissions are handled independently for each component carrier. As a baseline, control signaling is transmitted on the same component carrier as the corresponding data. However, as an alternative, it is possible to use cross-carrier scheduling where the scheduling information is transmitted to the terminal on a different component carrier to the corresponding data transmission. This option could, for example, be used for heterogeneous deployments, as described later in this article.

To reduce terminal power consumption, a carrier-aggregation-capable terminal as baseline receives on one component carrier, the primary component carrier. Reception of additional secondary component carriers can be rapidly turned on or off in the terminal by the base station through MAC signaling. Similarly, in the uplink, feedback signaling is transmitted on the primary component carrier and secondary carriers are enabled when necessary for data transmission.

Enhanced multi-antenna support
Downlink spatial multiplexing in Rel-10 is enhanced to support up to eight transmission layers, together with an enhanced reference-signal structure. Relying on cell-specific reference signals for higher order spatial multiplexing is less attractive, as the reference-signal overhead is not proportional to the instantaneous transmission rank, but rather to the maximum supported transmission rank. Rel-10, therefore, introduces extensive support of UE-specific reference signals for demodulation of up to eight layers. In addition, feedback of channel-state information (CSI) is based on a separate set of reference signals – CSI reference signals. CSI reference signals are relatively sparse in frequency but regularly transmitted from all antennas at the base station, while the UE-specific reference signals are denser in frequency but only transmitted when data is transmitted on the corresponding layer.

Separating the reference-signal structures supporting demodulation and channel-state estimation helps to reduce reference-signal overhead, especially for high degrees of spatial multiplexing, and allows for implementation of various beam-forming schemes.

LTE Rel-10 also introduces the possibility for uplink spatial multiplexing with up to four layers, essentially extending the uplink peak data rates by a factor of four compared to earlier LTE releases.

Improved support for heterogeneous deployments
Increased use of mobile broadband has shifted the focus from theoretical peak rates to practical data rates experienced by users. The actual rate is dependent on several deployment factors, such as the terminal-to-base-station distance. Since the ability to improve link performance or increase transmission power is limited, a denser infrastructure is required, in many cases to support very high data rates. A denser network also directly increases the system capacity, or in other words, the total amount of traffic that can be handled by the network.

Straightforward densification of an existing macro network is one way of achieving the required density. However, for areas where users are highly clustered, a potentially more attractive approach is to complement a macro-cell layer providing basic coverage with additional low-output-power pico cells where needed, as shown in Figure 2. The result of such a strategy is a heterogeneous deployment with two or more overlaying cell layers.

The idea of multiple cell layers is in itself not new; it is a deployment strategy and not a technology component. As such, heterogeneous deployments are possible with LTE Rel-8 and Rel-9. However, Rel-10 provides additional features that improve the support for this type of deployment.

In a heterogeneous deployment with cells of very different output power, cell association (which cell a terminal should connect to) plays an important role. From an uplink data-rate perspective, connecting to the cell with the lowest path loss results in a higher data rate at a given transmit power. This is opposed to the traditional approach of connecting to the cell with the best downlink signal quality. Determining the best cell for downlink association depends on the load: at low load, connecting to the cell with the strongest received downlink offers the highest data rates, while at high loads, connecting to the low-power node may be preferable as it enables resource reuse among low-power nodes.

Cell association in a heterogeneous deployment is enhanced by a factor of four compared to earlier LTE releases.
deployment is therefore a non-trivial task when overall network performance is taken into account. Nevertheless, a cell association strategy not only focusing on maximizing the downlink signal quality can lead to new interference situations in the network. Essentially, the uplink coverage area can be larger than the downlink coverage area, implying that there is a region around the low-power node (illustrated by the dashed area in Figure 2) where downlink transmission from the low-power node to a terminal is subject to strong interference from the macro cell.

The signal-to-interference ratio experienced by the terminal at the outermost coverage area of the low-power node is a function of the difference in output power between the high-power macro and the low-power node, and can be significantly lower than what is experienced in a more homogeneous deployment.

For the data part of a subframe, this is not a serious problem as the inter-cell interference coordination (ICIC) mechanism presented as early as in Rel-8 can be used to more or less dynamically coordinate the resource usage between the cell layers and avoid overlapping resource usage. The cell layers can exchange information about which frequencies they intend to schedule transmissions on in the near future, thereby reducing or completely avoiding interference.

The control signaling in each subframe is more problematic as it spans the full cell bandwidth and is not subject to ICIC. LTE Rel-10 offers two ways to handle this:

- frequency-domain schemes; and
- time-domain schemes.

**Frequency-domain schemes**

In frequency-domain schemes, carrier aggregation (CA) is used to separate control signaling for the different cell layers. At least one component carrier in each cell layer is protected from interference from other cell layers by not transmitting control signaling on the component carrier in question. For example, referring to Figure 2, the macro base station transmits control signaling on component carrier f, but not on component carrier f2, and vice versa for the low-power nodes located within the macro cell.

Cross-carrier scheduling is used to schedule data on all the component carriers in each cell layer, subject to the normal ICIC mechanism. Essentially, this creates a frequency reuse for control signaling, while still allowing terminals to dynamically utilize the full bandwidth (and thereby supporting the highest data rates) for the data part.

For example, an operator with 20MHz of spectrum may choose to configure two component carriers of 10MHz each and use carrier aggregation as described above. In addition to the benefits of connecting to the low-power node (dashed area in Figure 2), carrier-aggregation capable terminals will have the same peak data rates as in the case of a single 20MHz carrier. Rel-8/9 can also benefit from seeing a larger pico cell but can only access one component carrier.

**Time-domain schemes**

In time-domain schemes (non-CA-based) there is one component carrier in each cell layer. Time-domain separation of control signaling in the different cell layers can be used to handle interference. Some subframes in the low-power cell layer are protected from interference where the macro layer has muted control signaling. However, for backward compatibility reasons, cell-specific reference signals still need to be transmitted from the macro cell. By employing time-domain separation, Rel-8/9 terminals in the dashed area in Figure 2 will connect to the macro and not the low-power node and can access the full bandwidth of the carrier.

The discussion above assumes that terminals in both frequency-domain and time-domain schemes are allowed to connect to the low-power node, known as open access, and typically the low-power nodes are operator-deployed.

The terms Home-eNB and femto base station usually describe low-power base stations deployed by users at more or less random locations (from the operator’s perspective). Home-eNBs rely on the users’ fixed broadband for backhaul and are often associated with a closed subscriber group (CSG), where access is limited to specific users or terminals that are part of the CSG.

The use of CSG results in additional interference scenarios. For example, a terminal located close to, but not admitted to connect to the Home-eNB (as it is not part of the CSG), will be subject to strong interference and may not be able to access the macro cell. The presence of a Home-eNB may cause coverage holes in the operator's macro network. Similarly, reception at the Home-eNB may be severely impacted by uplink transmissions from the terminal connected to the macro cell. Thus, to protect the macro layer from severe interference in the case of Home-eNB with CSG, it is preferable to use separate carriers for the Home-eNB layer, possibly in combination with frequency-domain operation as outlined above.

**Relaying**

LTE Rel-10 extends LTE radio access technology with support for relaying functionality (Figure 3). In case of relaying, the mobile terminal communicates with the network via a relay node that is wirelessly connected to a donor cell using the LTE radio interface technology. Note that the donor cell will typically not only serve the relay node, but also terminals directly connected to the donor cell. The donor-relay link may operate on the same frequency as the relay-terminal link (inband relaying) or on a different frequency (outband relaying).

With the 3GPP relaying solution, the relay node will, from a terminal point of view, appear as an ordinary cell. This has the important advantage of simplifying the terminal implementation and making the relay node backward compatible. Essentially, the relay is a low-power base station wirelessly connected to the remaining part of the network using the LTE radio access technology.

One of the attractive features provided by a relay is improving coverage in the LTE-based wireless backhaul by simply placing relays at problematic locations in, for example, indoor environments. If the traffic situation demands, the wireless donor-relay link could be replaced, for example, by an optical fiber to serve the relay so that precious radio resources in the donor cell could be used for terminal communication.

As the relay transmitter can cause interference to its own receiver, simultaneous donor-to-relay and relay-to-terminal transmission may not be feasible unless the outgoing and incoming signals are sufficiently isolated. Isolation can be achieved by well separated and well isolated antenna.
structures, or through the use of outband relaying.

Similarly, at the relay it may not be possible to receive transmissions from terminals and transmit them to the donor cell at the same time. To handle interference, Rel-10 creates a gap in the relay-to-terminal transmissions using MBSFN subframes, as shown in Figure 3.

In an MBSFN subframe, the first one or two OFDM symbols are transmitted as usual carrying cell-specific reference signals and downlink control signaling. The remainder of the MBSFN subframe is not used and therefore can be used for the donor-to-relay communication.

The benefit of using MBSFN subframes compared to blanking transmission in the whole subframe is backward compatibility with Rel-8/9 terminals. Blanking the whole subframe is not compatible with Rel-8/9, as such terminals assume that cell-specific reference signals are present in (part of) each subframe. In addition, MBSFN subframes are supported from Rel-8.

Since the relay needs to transmit cell-specific reference signals in the first part of an MBSFN subframe, it cannot receive the normal control signaling from the donor cell. Therefore, Rel-10 defines a new control channel, transmitted later in the subframe, as shown in Figure 3, to provide control signaling from the donor to the relay.

In the same way as normal control signaling, this control channel type, of which multiple instances can be configured, carries downlink (donor-to-relay) scheduling assignments and uplink (relay-to-donor) scheduling grants. As the assignments refer to data in the same subframe, early decoding of this control information is beneficial.

Performance Results

ITU has defined some requirements for IMT-Advanced technology. Some of the most basic of these requirements, together with the corresponding capabilities of LTE, are summarized in Table 1.

The first release of LTE meets all of the advanced requirements except those for bandwidth and uplink peak spectral efficiency. These requirements are addressed in Rel-10 through carrier aggregation and uplink spatial multiplexing, respectively.

3GPP has extensively evaluated the performance of LTE radio access technology in relation to the IMT-Advanced requirements. Examples of LTE system performance, for downlink and uplink, FDD and TDD, and for the different test environments specified by the ITU (indoor hotspot; urban micro; urban macro; and rural macro) are shown in Figure 4. LTE Rel-10 fulfills and even surpasses all of the IMT-Advanced requirements (indicated by the dashed lines in Figure 4). Detailed assumptions for the evaluations outlined in Figure 4 can be found in reference [6].

These performance results were achieved without any of the extended features in Rel-10. Thus, LTE generation Rel-8 fulfills the subset of IMT-Advanced requirements on average for cell-edge spectral efficiency.

This, however, does not imply that Rel-10 features, such as extended downlink multi-antenna transmission and relaying functionality, are redundant. Rather, these features take the capabilities of the LTE radio access technology beyond IMT-Advanced. By including more advanced features, such as extended multi-antenna transmission, LTE system performance is enhanced beyond what is illustrated. A wider range of deployment scenarios is also addressed, including those with relays and non-contiguous spectrum allocations.

* MBSFN (Multicast-Broadcast Single Frequency Network) subframes, present already in Rel-8, were originally intended for broadcast support, but have later been seen as a generic tool, for example, to blank parts of a subframe for relaying support.
Figure 4: Performance results for FDD, TDD, downlink and uplink.
Conclusion

This article has provided an overview of the evolution of LTE, also referred to as LTE-Advanced. By introducing several new features, including carrier aggregation, enhanced multi-antenna support, and relaying, LTE-Advanced significantly boosts the performance and service capabilities of LTE radio access technology. LTE-Advanced has also been approved by ITU as an IMT-Advanced technology, thus confirming the characteristics of LTE as a 4G technology.

Erik Dahlman

joined Ericsson Research in 1993 and is senior expert in the area of radio access technologies. He has been deeply involved in the development and standardization of 3G radio access technologies (WCDMA/HSPA) as well as LTE and its evolution. He is part of the Ericsson Research management team working with long-term radio access strategies. He is also co-author of the book 3G Evolution – HSPA and LTE for Mobile Broadband and, together with Parkvall, received “Stora Teknikpriset” in 2009 for contributions to the standardization of HSPA. He holds a Ph.D. from the Royal Institute of Technology (KTH) in Stockholm.

Stefan Parkvall

joined Ericsson Research in 1999 and is currently a principal researcher in the area of radio access, working with research and standardization of cellular technologies. He has been heavily involved in the development of HSPA and LTE, and is co-author of 3G Evolution – HSPA and LTE for Mobile Broadband. In 2009, he received “Stora Teknikpriset” (one of Sweden’s major technical awards) for his work on HSPA. He earned his Ph.D. in electrical engineering from KTH in Stockholm in 1996. He was previously an assistant professor in communication theory at KTH from 1996 until 2001, and a visiting researcher at the University of California at San Diego in the US from 1997 until 1998.

Anders Furuskär

joined Ericsson in 1990 and is currently a principal researcher within the field of wireless access networks at Ericsson Research. His current focus is on evolving HSPA and LTE to meet future demands on data rates and traffic volumes. He holds an M.Sc. and a Ph.D. from KTH in Stockholm.

References

1. ITU-R SG5, Invitation for submission of proposals for candidate radio interface technologies for the terrestrial components of the radio interface(s) for IMT-Advanced and invitation to participate in their subsequent evaluation, Circular Letter 5/LCCE/2, March 2008.
4. 3GPP TR 36.912, Feasibility study for further advancements of E-UTRA (LTE-Advanced).
5. 3GPP TR 36.913, Requirements for further advancements for E-UTRA (LTE-Advanced).
The telecommunications industry is characterized by its rapid rate of change, which creates new business opportunities. The Business Support System (BSS) is becoming the focal point for success, helping enterprises attract and retain customers and create attractive services in a constantly evolving market.

What is a BSS?
The microeconomic model of the late Herbert A. Simon (1978 Nobel prizewinner in Economic Sciences) states that the BSS is the “connection point” between external relations (customers, partners and suppliers) and an enterprise’s products and services. In turn, products and services must be associated with their constituent resources, such as network infrastructure, service applications, contents and factories. Simon’s reasoning states that it is crucial for organizations to control and maintain their business information.

**History and developments**
Twenty years ago, the telecommunications industry was much simpler: it was essentially one service, one business model, one type of customer – the subscriber – and no complex value network. This scenario has changed dramatically. Today, an enterprise must provide a wide range of services to many different customer types, using multiple business models and complex value networks. This environment of many customers and multiple business models creates business opportunities that can be turned into revenue, provided that the complexity can be managed.

In parallel with the growth in the industry, the BSS has also developed, albeit in a less controlled manner. Many present BSSs are multi-vendor and multi-generational, comprising several parallel silo-like systems. The time has come to reengineer the BSS to overcome these inherent problems, allowing the industry to develop innovative business models.

**Background**
Why does an enterprise need a BSS? Essentially, a BSS provides a set of tools to transform assets, such as content and communication networks, into commercial offerings and ultimately revenue.

BSS applications have traditionally been monoliths with their own view of information/data, processes, rules and document formats. In short, business assets and decisions have been taken over by system vendors and locked into the architecture, making them almost impossible to change and reuse in other business system contexts.

**Business first, technology second**
Every business is unique, with a unique environment and consequently a unique BSS. The BSS is defined from the enterprise’s full, cross-functional business perspective and provides strategic, tactical and operational support. A BSS enables the enterprise to gain control over its business, through measurements and analysis feedback loops that deliver a 360-degree view of all assets. This level of transparency allows the enterprise to control and maintain their business information.
The cross-functional simplified BSS of the future

Managing assets
The fundamental task of the BSS is accurate and synchronized control and sharing of business assets across the organization. Controlled management of offerings and external relations is also essential. Information is shared and controlled by the BSS through well-defined information ownership, one common information model and one set of shared data.

BSS Characteristics
The BSS handles the full life cycle for many enterprise assets, including creation, design, implementation, deployment, operation, analysis, update/improvement and termination. The duration of different business assets varies greatly, as does the volume of information handled. The life cycle of a business model, for example, is measured in years, whereas events from networks are measured in milliseconds. Similarly, data volumes can range from just a couple of bytes up to several petabytes.

The BSS delivers competitive time to market (TTM), time to consumer (TTC) and business agility to pursue new business models on demand, with complex B2B value networks, and low total cost of ownership (TCO). The BSS is built for continuous evolution of business models, putting the enterprise in the driver’s seat, deciding the way forward.

The BSS scales in multiple directions and independently. For example, the number of customers may grow while the number of products remains constant or vice versa.

Given the emerging multinational nature of the telecommunication industry, the BSS supports business in multiple regions, enabling local adaptation (see Box B) and enforcing corporate standards where necessary.

Designing and operating the business

FIGURE 1 The functional BSS Architecture.
of any enterprise requires the coordination of several processes, people and applications. Application support is becoming automated. The BSS adapts to ways of working and facilitates operational excellence, such as process support, user-BSS interaction and systems in combination.

The BSS supports value-chain management by connecting, controlling and analyzing business partners on a relationship level, a commercial level and a delivery level. Ericsson’s customers are searching for support to help them move from today’s BSS silos to BSS maturity. Ericsson’s BSS Architecture supports BSS transformations with an arbitrary functional starting point to fit a multi-vendor environment.

**BSS Architectural Framework**

Ericsson’s approach to this very complex area is the BSS Architectural Framework, which provides a common foundation for all stakeholder perspectives and can be described from the following three points of view:

- **BSS Logical Architecture**
- **BSS Realization Architecture**
- **BSS Libraries.**

The BSS Logical Architecture, shown in Figure 3, provides structure, terminology, principles and tools to bring business and technical stakeholders together by focusing on abstract and logical aspects of the BSS. It contains a range of architectural models, structures and views that form an architectural platform. It provides the ability to illustrate and discuss organizational issues in a coherent as-is and to-be manner, for example:

- **how to implement a major network transformation while maintaining 24/7 service performance;**
- **how to describe all steps of a BSS implementation from plan to operation;** and
- **asset and business model construction and life cycle management.**

BSS Logical Architecture may be used at different levels of abstraction, from overviews of process models to tangible and detailed workflows, covering a variety of activities, such as automated processing, manual tasks and user interactions. Essentially, the logical architecture describes the what and some of the why, not the how.

Ericsson’s preferred choice of architecture modeling language for the BSS Logical Architecture is ArchiMate®. This language is an open standard for enterprise architecture modeling produced by The Open Group. It has its origins in TOGAF but provides a more lightweight approach; we have found this to fit the need to be able to discuss relevant matters without going overboard in tool-related details.

One of the main functions of the BSS Logical Architecture is to provide structure for a BSS, and this achieved through separating external and internal aspects. Internal aspects are further divided into business, applications and infrastructure.

A key property of this model is the interconnection of these levels through service layers resulting in loose coupling and isolation between the layers.

At this point, it should be noted that most models tend to describe only the rather static view of a BSS in operation. We have found it necessary to add another dimension: the evolution phases to accommodate for the design, implementation, operation and termination of BSS functionality; in other words, BSS evolution to improve business agility.

**BSS Studio**

Architectures and models are only as useful as the tools available to manage them. Ericsson’s approach is to externalize processes, rules, events, business objects and the information from applications. These exposed facets are referred to as enterprise entities. This abstraction allows us to describe the BSS applications much more clearly in terms of pure functionality, and allows applications to expose a well defined set of services to the business layer.

To make full use of the enterprise entity concept, Ericsson has created the BSS Studio. Figure 4, a complete environment for enterprise-entity life cycle management, governance, analytics, publication, simulation, verification, performance and commissioning and decommissioning for BSS production. The studio is intended for an enterprise’s business users. Enterprise entities include:

- **actors – companies, functions and individuals;**
- **roles – customer, supplier and service provider;**
- **services – sales and contracting;**
- **processes – business processes, as defined by TMF in eTOM;**
- **objects – products, orders, contracts and accounts;** and
- **rules – decision-making processes.**
The cross-functional simplified BSS of the future are modeled in BPMn 4.0 and rules in SBVR.

**Figure 4** The externalization of Enterprise Entities and the BSS Studio.

Traditional BSS application with unique and secluded entities

Traditional BSS app
- Rules
- Processes
- Events
- Objects and data

BSS Studio
- Enterprise business information model
- Business object
- Business rule
- Business process
- Business event

Application data model
- Application service

**Figure 5** An order-handling process. The notation shown here is conceptual; processes are modeled in BPMN 2.0 and rules in SBVR.

Receive purchase order

Send purchase order reception ack.

Required entities
- Order handling
- Customer handling
- Product offering

Event and process
- PO reception
- Create customer order
- Check customer order

Created BO and LC
- Customer order

Application tasks
- Activate customer order
- Activate services
- Activate resources
- Activate billing
- Archive customer order

Send purchase order delivered

External
- Purchase order

As an example, Figure 5 shows how an order-handling process could be built.

The Enterprise Information/Data Model is the common and shared information framework for an organization. It is divided into a set of models, each belonging to a specific layer of the logical architecture described above, resulting in tailored information and data sub-models for business purposes, application use, and for optimal storage.

Much of the information in the enterprise business information model is also Master Data, in other words, the essential structural information and data of the Enterprise. Master Data management services are exposed to the enterprise at Business Level in the BSS Studio.

**BSS Realization Architecture**

The Realization Architecture, Figure 6, consists of basic infrastructure building blocks and components, whose purpose is to serve as a common foundation for the implementation and integration of BSS components and solutions.

The goal is to visualize the realization in terms of technology layers and their functionality. Both the Business and Application layers from the BSS Logical Architecture can be naturally mapped onto the BSS Realization Architecture, coupling most directly for the Infrastructure layer. Our preferred model for the BSS Realization Architecture is CORA 6, which is a vendor-agnostic N-tier architecture.

The CORA model describes elements with their interactions to fit the different architecture styles, such as N-tier, service-oriented and resource-oriented architectures. It is a general-purpose model, not only used in the BSS domain, and therefore some of the terms have counterparts in the Logical Architecture, however, the context normally resolves such overlaps.

The Channel access layer provides client-specific software to enable access to information systems. This layer is accessible for all actor groups, whether internal or external.

The Presentation layer provides presentation-specific software for displaying information to the user and handling user-initiated events.

The Composition layer provides...
composition-specific software, clustered in Orchestrated and Composed elements.

The Integration layer provides support for Synchronous Communication, Asynchronous Communication and Common elements.

The Application layer provides application-specific functionality; essentially, this is where the real work gets done. The layer is clustered into Entity, Task and Utility applications.

Finally, the Data layer provides support for Mediation (data access, aggregation, cleansing, transformation) and Data Storage elements.

Controlling the whole stack are two governing entities:

- Business Governance provides functionality to maintain full control of all business specifics in the BSS environment, covering all Enterprise Entities that are either in production or under construction;
- Security and compliance handles authority, authentication, behavioral compliance and audits;
- IT Governance manages IT assets down to the individual application component and/or service; and
- SOA governance is explicitly included.

The BSS Realization Architecture is a standard IT architecture taking advantage of standard and well-proven technologies. Essential BSS additions are made to support enterprise governance and business architecture. This gives robust technical qualities and predictable costs.

**BSS Libraries**

Ericsson BSS Libraries provide a platform for the enterprise-unique parts of a BSS. These libraries include, for example, a set of templates for the Business layer, as well as best practices. The libraries further cover base implementations of Enterprise Entities, Enterprise Business Information Model and Application Services. All libraries are built, maintained and evolved using the BSS Studio. Consequently, the BSS Libraries provide the flexibility to build the right solution for the enterprise, and the stable experience base that captures knowledge built in previous projects.

**Summary and conclusions**

The BSS Architecture puts business control into the hands of the operator, limiting dependence on system vendors and integrators. The Ericsson BSS Architecture gives the operator the power to design and manage its own BSS, from present to future business models, from people-oriented processes to details of assets. At the same time, it supports asset-centric and information-centric viewpoints, providing full traceability from concept to implementation. Finally, the Ericsson BSS Architecture directly supports an enterprise's way of working, enabling future growth in multiple dimensions. Ericsson firmly believes that this is essential for the telco operator of the future to achieve the operational excellence necessary for market success.
The cross-functional simplified BSS of the future

References

2. ArchiMate 1.0, http://www.archimate.org

Lars Angelin
> is an expert in the technology area of Multimedia Management at Business Unit Multimedia R&D Management, where he has been working since 2009. He holds an M.Sc. in engineering physics and a Tech. Licentiate in teletraffic theory from the Lund Institute of Technology. He has more than 24 years of work experience in the areas of innovation, concept development and strategies within the Telco and education industries. In 1996, he joined Ericsson Software Technology AB as a research engineer and in 2003 became a concept developer for telco-near applications, initiating and driving activities, most of them related to the M2M or BSS areas.

Ulf Olsson
> has a background in software architecture for distributed military Command and Control systems. He joined Ericsson in 1996, working mainly with architecture issues concerning packet-based systems, such as packet PDC, GPRS, UMTS and the CDMA2000 packet core network. He is now senior expert with Business Unit Multimedia, Portfolio Management and Technology group. He focuses on Data Management, Business Support Systems, IMS and developer-oriented issues. He holds an M.Sc. in engineering physics from the Royal Institute of Technology, Stockholm.

Per Tengroth
> is a principal solution architect in the BSS domain and has been working in Consulting and Systems Integration at Ericsson, responsible for end-to-end BSS solutions since 1996. He holds an M.Sc. in mathematics and a Master’s degree in accounting and finance administration, both from Gothenburg University. He has recently outlined the BSS offering within the BSS Transformation Program and participates in Ericsson Network Architecture Evolution Strategy. Per has been working in the BSS domain for more than 30 years.
Ericsson’s Business Communications Suite (BCS) is a set of applications that an operator can install to offer enhanced communication services to companies in the small and medium enterprise (SME) segment. With BCS, operators can extend their portfolios — and consequently revenue streams — with a branded end user experience spanning multiple devices into the rapidly growing Unified Communications (UC) market.

Here we present an overview of the Ericsson BCS package in the very important enterprise communications segment.

The business communications challenge
Attention to customer needs has always been a cornerstone of a successful business. In today’s business environment, characterized by accelerated pace and diversity, multiple devices, constant connectivity and rapidly expanding data volumes, this business principle is perhaps even more critical than ever for success. Organizations that can successfully maintain focus on their customers, whose needs are constantly changing, will distinguish themselves from the competition.

Unanswered phone calls and e-mails, information that cannot be accessed, or experts that cannot be reached result in questions that cannot be answered and ultimately lost business, lost revenue, and worse still, customers lost to the competition.

Ericsson’s BCS was designed to resolve such issues, which modern organizations confront daily, to enhance the efficiency and reach of a business with emphasis on usability and enriched user experience (UX).

BCS helps operators’ customers to be more competitive, in turn strengthening the operator’s enterprise customer base and supporting long-term business sustainability.

Functionality
BCS provides efficient business communication support tools, including:
- videoconferencing;
- multimedia;
- document sharing;
- interactive collaboration tools, such as instant messaging (IM) and presence indicators;
- mobile telephony; and
- IP telephony.

BCS supports user devices ranging from ordinary mobile phones to feature-rich, IP-based terminals providing a context-sensitive UX.

Using BCS, operators can offer attractive services to the enterprise customer market to support employee efficiency as well as offering cost control. In this way, BCS helps the operator attract and retain valuable business customers, as well as raising ARPU for business users.

Fully featured, easily integrated
A complete business communication solution encompasses a variety of features for end users and organizations. Such communication solutions need to be tightly integrated into an organization’s IT infrastructure.

To achieve this without excessive integration costs, Ericsson employs

<table>
<thead>
<tr>
<th>BOX A Terms and abbreviations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A-SBG</strong></td>
</tr>
<tr>
<td><strong>API</strong></td>
</tr>
<tr>
<td><strong>ARPU</strong></td>
</tr>
<tr>
<td><strong>BCS</strong></td>
</tr>
<tr>
<td><strong>CAMEL</strong></td>
</tr>
<tr>
<td><strong>CAPv2</strong></td>
</tr>
<tr>
<td><strong>CPU</strong></td>
</tr>
<tr>
<td><strong>DSL</strong></td>
</tr>
<tr>
<td><strong>EGEM</strong></td>
</tr>
<tr>
<td><strong>GUI</strong></td>
</tr>
<tr>
<td><strong>IFC</strong></td>
</tr>
<tr>
<td><strong>ILE</strong></td>
</tr>
<tr>
<td><strong>IM</strong></td>
</tr>
<tr>
<td><strong>IMS</strong></td>
</tr>
<tr>
<td><strong>IN</strong></td>
</tr>
<tr>
<td><strong>IRS</strong></td>
</tr>
<tr>
<td><strong>JDBC</strong></td>
</tr>
<tr>
<td><strong>JSON</strong></td>
</tr>
<tr>
<td><strong>JSR</strong></td>
</tr>
<tr>
<td><strong>LOTC</strong></td>
</tr>
<tr>
<td><strong>MMAS</strong></td>
</tr>
<tr>
<td><strong>NAT</strong></td>
</tr>
<tr>
<td><strong>OEM</strong></td>
</tr>
<tr>
<td><strong>OMP</strong></td>
</tr>
<tr>
<td><strong>P2P</strong></td>
</tr>
<tr>
<td><strong>PGM</strong></td>
</tr>
<tr>
<td><strong>PL</strong></td>
</tr>
<tr>
<td><strong>RCS</strong></td>
</tr>
<tr>
<td><strong>SAF</strong></td>
</tr>
<tr>
<td><strong>SC</strong></td>
</tr>
<tr>
<td><strong>SIP</strong></td>
</tr>
<tr>
<td><strong>SIP-UR</strong></td>
</tr>
<tr>
<td><strong>SME</strong></td>
</tr>
<tr>
<td><strong>Tel-UR</strong></td>
</tr>
<tr>
<td><strong>TLS</strong></td>
</tr>
<tr>
<td><strong>UC</strong></td>
</tr>
<tr>
<td><strong>UX</strong></td>
</tr>
<tr>
<td><strong>VoIP</strong></td>
</tr>
<tr>
<td><strong>WCE</strong></td>
</tr>
<tr>
<td><strong>XML</strong></td>
</tr>
</tbody>
</table>
an out-of-the-box approach. A key design goal has been to ensure that BCS-specific components and the supporting infrastructure are well integrated. Consequently, BCS applications are designed with inherent IT integration in mind. Configuration management and client software integration are part of the BCS package, as is the integrated company self-care provisioning web GUI. This interface gives a company administration control over user accounts and allows end users to manage their personal communications settings.

Even though BCS utilizes the core functions of IMS, it can run without the need for full-scale IMS implementation. Ericsson’s IMS Low Entry (ILE) package is a cost- and size-optimized IMS Core configuration that together with BCS offers an attractive alternative to operators. A default parameter setting supporting BCS within ILE minimizes integration overhead.

Time to deploy
This is a fundamental aspect of the decision-making process. In other words, the time it takes to set up and complete an initial delivery project to enable an early soft launch is of vital importance to the business customer.

Ericsson’s BCS supports early soft launch to help the operator’s customers transfer smoothly to the BCS environment.

Box contents
No two operators are alike. Consequently, Ericsson’s BCS has been designed as a set of applications from which operators can choose, to suit their customer base. These applications are:

- Business Voice;
- Collaboration;
- Conferencing; and
- Mobility.

Each application can be used in stand-alone mode, or together to provide an integrated, homogeneous and branded user experience across many devices and services.

**BCS Business Voice**

The voice application server offers a feature set designed to enable streamlined workflows and solve immediate “pain points” typically experienced by small companies. Some of the key features designed to enhance voice reachability include:

- call forwarding options and parallel alerting to separate numbers;
- short number dialing;
- support for On-Net/Off-Net indication via Diameter charging records;
- automated attendant function;
- call distribution group; and
- company self-care web GUI, with provisioning systems integration.

**BCS Collaboration**

Collaboration is a package of client software and supporting server components that together offer:

- directory search;
- presence;
- MS Exchange calendar integration;
- P2P IM and file transfer;
- software and configuration management of clients;

Client software can be divided into two main categories:

- IMS-native clients; and
- thin or Web Client Enabler (WCE) enabled clients.

WCE, shown in Figure 2, is an integral part of BCS Collaboration, allowing Ericsson and third-party developers to rapidly produce clients for new platforms that
do not yet have a full IMS stack. Operators can in turn provide IMS-based services to a much wider range of devices. The major platforms supported today include BlackBerry, Symbian S60, iPhone, MS Windows and the SNOM 870 VoIP desk phone, with more coming. The Windows PC client and Symbian S60 clients are IMS-native; iPhone and Blackberry are WCE-based clients. The SNOM 870 VoIP desk phone is a hybrid option, which relies on native IMS signaling for voice and WCE for presence and directory search functionality.

The BCS client and its functionality vary slightly for each platform. The following features are enabled:

**General**
- secure signaling connections (TLS);
- automatic software upgrade;
- automatic client configuration;
- MS Office applications integration.

**Contact management**
- favorite (“buddy list”) for contacts;
- access to local and directory contact details via unified search functionality with integrated presence display; and
- communication initiation from context (voice/IM/SMS/MMS).

**Presence management**
- display presence status of contacts;
- manage own presence, including activity and smiley-enabled text note;
- show “in a call” status.

**IM**
- smiley-enhanced IM;
- file transfer in IM session.

**Communication history**
unified history log.

**BCS Conferencing**
BCS Conferencing allows streamlined sharing of documents, video and audio using a PC. It provides an essential tool to help reduce the environmental impact of geographically distributed operations. The user experience includes:
- Web-based participation facilitated by a meeting-unique URL via, for example, an MS Outlook calendar booking or e-mail; and
- ad-hoc participation via the BCS PC client where a conference session can be set up by selecting participants from the presence-enabled directory search.

**BCS Mobility**
BCS Mobility enables the use of GSM phones as IMS endpoints for voice. It allows BCS Business Voice to provide telephony services to GSM phones, such as short number dialing and parallel ringing.

This means that all BCS devices are served by the same service implementation, ensuring a consistent user experience across all terminal types. These mechanisms can potentially be extended to provide IMS services to GSM phones.

The BCS Mobility application utilizes media gateways and CAPv2 IN integration to ensure that calls are brought to the application server residing in IMS, where originating and terminating services are executed.

**Deploying BCS**

**Ericsson BCS including IMS**
The recommended identity setup within IMS is to have a single Implicit Registration Set (IRS) consisting of one Tel-URI and one SIP-URI. In practice, this means that users can be identified with a telephone number (Tel-URI), or an e-mail address format (SIP-URI). The typical setup offers parallel alerting on all the user’s devices, such as mobile phone, PC client and desk phone.

An operator that has deployed BCS can offer GSMA Rich Communication Suite (RCS) services by reusing the supporting core infrastructure, such as IMS and Ericsson PGM (presence enabler). Furthermore, an RCS-enabled business customer or family member communicating with a BCS company user will be able to enrich the communication with, for example, IM and presence.

**The BCS Application Server**
All BCS applications, except conferencing, run in a standard Java environment, supporting key communication standards, such as JSR-154 (the web servlet API) and JSR-289 (the SIP servlet API) along with JDBC.

**OMP**
BCS applications are deployed and verified on the Open Multimedia Platform (OMP) and Multimedia Application Server (MMAS).

OMP provides BCS with an efficient Linux-based cluster optimized for telecom-grade environments. OMP and MMAS are used by several of the
Applications within the cluster are implemented according to Service Availability Forum (SAF) standards. SAF controls which applications run on the cluster, on which blade and on how many blades. Common fault management and performance management interfaces are aligned between applications running on OMP.

The currently recommended hardware is Sun Blade 6000, supporting up to 10 blades per chassis, each blade equipped with two Intel Xeon series multi-core x86 CPUs.

A simplified view of how the BCS applications (except BCS Conferencing) can be deployed on a 10-blade cluster, is shown in Figure 5. The System Controller (SC) blades boot from local redundant disks and manage the cluster's payload (PL) blades. Except for the database blades the PL blades are diskless. The database blades also boot from the SC blades and use local (redundant) disks for persisting objects into the database.

**Scalability**

Depending on the desired characteristics, BCS applications can be deployed in other constellations with upward scalability to 18 PL blades with two SC blades spread over two Sun blade chassis.

For deployments where more than 18 PL blades are needed, the system can be deployed across multiple OMP clusters. For extremely large deployments, a single BCS application is installed within an OMP cluster; the user base can be partitioned, for example, using IMS Initial Filter Criteria (IFC). This way, different users can be allocated to different OMP clusters using standard IMS mechanisms.

An 18-blade BCS Voice 2.1 installation is expected to handle roughly one million business users depending on the traffic model.

**Deployment in reality**

Offering IP-based services, including presence, requires a constantly active IP connection, which can negatively impact battery life, especially for mobile devices. BCS clients employ an optimization scheme that significantly reduces the load on the presence enabler (for example, the Ericsson PGM product), while providing a substantial improvement in battery life.

Initial tests on an optimized Symbian S60 show an increase in standby time by a factor of about 25, transforming the presence capability from a seldom used function into a practical feature.

BCS PC client users typically connect over a shared DSL line with a basic gateway located at their company premises. These gateways normally provide firewalls and Network Address Translation (NAT) services, which complicate or restrict IP connectivity.

![Figures 1, 2, and 3: Open Media Platform, Sun Blade 6000, and enhanced communication services](images/ericsson_review_2010_38_enhanced_communication_services.png)
Christian Olrog

> is a solution system manager and chief architect for business applications in Solution Area Consumer and Business Applications (SA-CBA). In 1999, he earned his M.Sc. in general physics from the Royal Institute of Technology (KTH) in Stockholm. He joined the department of New and Special Business Operations at Ericsson in 1999 and has been involved in research and development in different parts of Ericsson in areas ranging from Wireless LAN standardization and IP security to embedded devices and Enterprise PBXs.

Ulfr Olsson

> has a background in software architecture for distributed military Command and Control systems. He joined Ericsson in 1996, working mainly with architecture issues concerning packet-based systems, such as packet PDC, GPRS, UMTS and the CDMA2000 packet core network. He is now senior expert with Business Unit Multimedia, Portfolio Management and Technology group. He focuses on Data Management, Business Support Systems, IMS and developer-oriented issues. He holds an M.Sc. in engineering physics from KTH in Stockholm.

Ericsson’s BCS clients are enhanced, utilizing advanced functionality in the A-SBG to provide seamless access even when such gateways are present.

Conclusion

Ericsson BCS offers operators the opportunity to enhance and leverage existing investments to attract businesses and their corresponding high-ARPU, low-churn users.

Ericsson BCS' brings together several advanced technologies in a combined package that offers the possibility to present an operator’s brand across multiple devices with a focus on user experience.

References

1. Converged service for fixed and mobile telephony, Ericsson Review 2, 2009
2. Open multimedia platform framework, Ericsson Review 1, 2009

25 years ago

> In 1985, Ericsson Review published an article on the field trial of an optical fiber cable TV system carried out jointly by the Swedish Telecommunications Administration and Ericsson. In this project, Ericsson developed and supplied the digital transmission system, the equipment for the analog subscriber lines and the optical fiber cable. For part of the field trial, Ericsson developed demo equipment using wavelength division multiplexing (WDM) for fiber optic subscriber lines. The equipment demonstrated one of several possible solutions for subscriber connections in future interactive networks. The subscriber lines were arranged in a star-shaped structure and employed WDM to provide a complete multi-service subscriber connection using only one optical fiber.

50 years ago

> In 1960, an article was published about a new method of constructing transmission equipment. The technology had been developed to meet equipment demands from telephone administrations. It provided optimum performance for a reasonable technical outlay and fulfilled the following requirements:
- occupying a small space;
- consuming low power; and
- needing little maintenance.

The new method was a response to the rapid expansion of the telephone networks at the time. Figures for the Swedish national network showed that the number of circuit kilometers increased three times from 1948 to 1958, with 50 percent of this increase occurring after 1955.

75 years ago

> In 1935, the Great Depression had a dramatic effect on the number of telephones sold. North America, which at the time accounted for 57.2 percent of the world’s telephones, and Oceania were most affected. The decrease was significant enough to lower global telephone density from 1.8 to 1.54 telephones per 100 inhabitants, even though in all other parts of the world density increased slightly. The study on telephone sales, which appeared in Ericsson Review, showed a constant rise in telephone density up until 1930, with the decline starting in 1931 and ending in 1933, in line with the bottom of the depression.

www.ericssonhistory.com