

The long-term evolution of 3G

Erik Dahlman, Hannes Ekström, Anders Furuskär, Jonas Karlsson, Michael Meyer, Stefan Parkvall, Johan Torsner and Mattias Wahlqvist

The Third Generation Partnership Project (3GPP) has begun charting the long-term evolution of 3G to ensure the competitiveness of 3G technology during the next 10 years and beyond. The fundamental aims of this evolution – to further improve service provisioning and reduce user and operator costs – will be met through improved coverage and system capacity and by improving data rates and reducing latency.

The authors describe technologies that promise to provide these improvements, including orthogonal frequency-division multiplexing (OFDM), multi-antenna solutions, evolved quality-of-service (QoS) and link-layer concepts, and an evolved system architecture. The authors also present the results of a performance evaluation which confirms that the long-term requirements can indeed be fulfilled.

Background and targets

Third-generation mobile systems (3G) based on WCDMA radio-access technology are being deployed on a broad scale all around the world. A first step in enhancing or evolving this technology entails introducing high-speed downlink packet access (HSDPA) and an enhanced uplink (E-UL), giving a radio-access technology that is highly competitive.

However, knowing that user and operator requirements and expectations will continue to evolve, the 3GPP has begun considering the next major step or evolution of the 3G standard (sometimes called *Super 3G*) to ensure the long-term competitiveness of 3G. The 3GPP recently launched a Study Item entitled *Evolved*

UTRA and UTRAN. The study will investigate means of achieving major leaps in performance in order to improve service provisioning and reduce user and operator costs.

It is generally assumed that there will be a convergence toward the use of internet protocols, and all future services will be carried on top of IP. Therefore, the focus of the evolution is on enhancements to the packet-switched (PS) domain. In particular, the 3GPP aims to deliver a set of specifications for evolved 3G radio access in 2007. Initial product availability is envisioned around 2009-2010.

The main objectives of the evolution are to further improve service provisioning and reduce user and operator costs. More specifically, some key performance and capability targets for the long-term evolution are

- the potential to provide significantly higher data rates compared to HSDPA and E-UL, with target peak data rates of more than 100Mbps over the downlink and 50Mbps over the uplink;
 - improved coverage – that is, high data rates with wide-area coverage;
 - the potential to significantly reduce latency in the user plane in the interest of improving the performance of higher layer protocols (for example, TCP) as well as reducing the delay associated with control plane procedures (for instance, session setup); and
 - greater system capacity – threefold capacity compared to current standards.
- One other key requirement of the long-term evolution is to allow for a smooth migration to these technologies. This can be ensured by giving
- operators the ability to deploy the new system in existing (already paid for) spectrum. This puts technical requirements on spectrum flexibility, allowing deployment in many different allocations of spectrum (2G and 3G);
 - operators the ability to reuse sites and investments in site and transmission equipment;
 - operators the ability to maintain their base of end users by smooth service phase-ins and phase-outs. This puts requirements on service continuity and mobility between systems;
 - operators the ability to deploy the new technology in areas where it is profitable. Elsewhere, they can rely on existing systems for coverage and, to some extent, capacity. This puts requirements on mobility between systems; and

BOX A, TERMS AND ABBREVIATIONS

2G	Second-generation mobile system	PS	Packet switched
3G	Third-generation mobile system	QAM	Quadrature amplitude modulation
3GPP	Third Generation Partnership Project	QoS	Quality of service
AML	Adaptive multilayer	QPSK	Quadrature phase shift keying
ARQ	Automatic repeat request	RAB	Radio access bearer
E-UL	Enhanced uplink	RAN	Radio access network
E-UTRA	Evolved UTRA	RLC	Radio link protocol
FDMA	Frequency domain multiple access	RNC	Radio network controller
FFT	Fast Fourier transform	SGSN	Serving GSN
GBR	Guaranteed bit rate	SIC	Successive interference cancellation
GGSN	Gateway GSN	SISO	Single input, single output
GPRS	General packet radio service	S-PARC	Selective per-antenna rate control
GSN	GPRS support node	TCP	Transmission control protocol
HSDPA	High-speed downlink packet access	UE	User equipment
IFFT	Inverse FFT	UMTS	Universal mobile telecommunications system
MIMO	Multiple input, multiple output	UTRA	UMTS terrestrial radio access
OFDM	Orthogonal frequency division multiplexing	UTRAN	UMTS terrestrial radio access network
PDP	Packet data protocol	WCDMA	Wideband code-division multiple access
PDU	Protocol data unit		

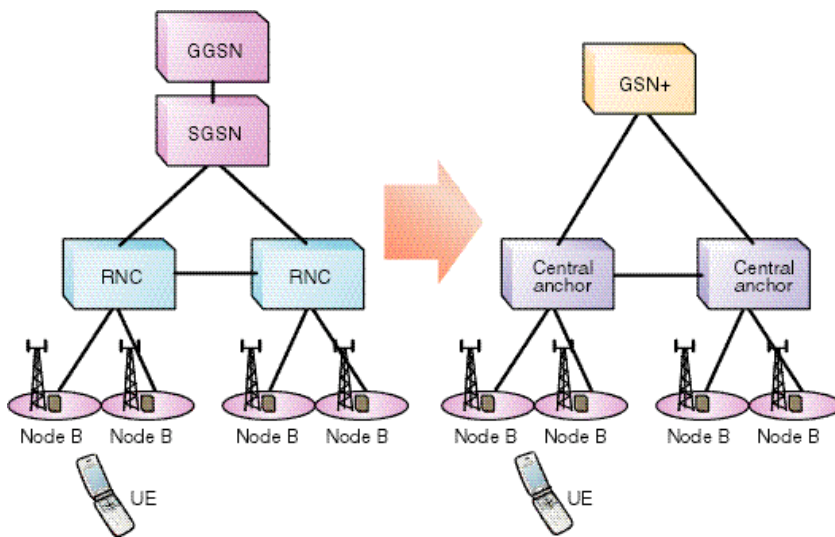


Figure 1
 Left: Architecture of 3GPP Rel-6.
 Right: Possible evolution of the architecture; the number of nodes along the user plane data path has been reduced from four to three. RLC protocol on top of the hybrid ARQ layer operates between the UE and a central anchor node.

- manufacturers the ability to reuse investments in development. This results in stable and competitively priced equipment and shorter time to market.

To reach the performance and capability targets, 3GPP must consider some new radio-transmission technologies as well as updates and modifications to the architecture of the 3G radio network. Ericsson believes that the following building blocks can help fulfill the stated targets:

- simplified system architecture;
- evolved QoS and link-layer concepts;
- the use of adaptive multilayered OFDM (AML-OFDM) as a new radio-access technology; and
- advanced multi-antenna solutions.

Technical overview

Architecture evolution

Given the requirements to reduce latency and cost, it makes sense to consider a system architecture that contains fewer network nodes, because this reduces the overall amount of protocol-related processing, number of interfaces, and cost of interoperability testing. Fewer nodes also means that one can more easily optimize radio interface protocols, for example, by merging some control plane protocols. Finally, shorter signaling sequences results in more rapid

session setup. Figure 1 shows the architecture of 3GPP Release 6 (left) and a possible evolution of this architecture (right).

In Release 6 (Rel 6), the gateway GPRS support node (GGSN) serves as an anchor node in the home network. All traffic is typically routed to the home network to maintain a concise service environment. This allows operators to filter traffic and provide security. The radio network controller (RNC) manages radio resources and local mobility, controls the bearers, and optimizes the transport network. It also serves as a termination point for some radio protocols. The serving GSN (SGSN) functions as an anchor node in the visited network and manages mobility and sessions.

In the evolved architecture, the GGSN functionality continues to reside in the home network to ensure roaming and consistency in the service environment. Similarly, the Node B continues to handle the lower layers of the radio interface. Therefore, a logical evolution might be to merge the SGSN and RNC into a central anchor node. An alternative solution might be to distribute the functionality of the SGSN and RNC, thereby completely eliminating these nodes.

The evolved architecture might also require a central anchor node in the visited network, in order to ensure mobility, security and transport network efficiency. We discuss this in greater detail below.

Mobility

Assuming the same level of mobility as in present-day 2G and 3G networks, this central node would ensure good handover performance with little service interruption. Also, it would hide the movements of user equipment (UE) from the home network.

Security

It is easier to cipher user data and protect the integrity of signaling and sensitive subscriber information and ciphering keys in a central site than in the Node B. The assumption is that the evolved network must provide at least the same level of security as present-day networks.

Efficiency in the transport network

A central node would greatly benefit the transport network:

- IP-related headers could be compressed at a central point, yielding gains over the radio interface and over the last mile link to the Node B.
- During handover between two Node Bs, user data could be forwarded using fewer resources. Without the central node, the same data packet might have to traverse the last mile link to the base station up to three times.

The central anchor node could contain functionality from the RNC and SGSN. However, it might be advantageous to move some of the SGSN functionality to the GGSN node. We denote the enhanced GGSN *GSN+* (Figure 1, right). Although it might be possible to split functions between the *GSN+* and central anchor, it is generally accepted that SGSN functionality will be required in a visited network (for roaming) and must be located in the central anchor. All other functionality could be located in the *GSN+*. To minimize delays, radio interface protocols involving the UE should be located in the visited network.

If one adheres to these principles, we see that the Rel-6 mobility management model can be simplified with little impact on the *GSN+*. As a consequence, the evolved architecture has the potential

- to reduce latency in the user plane because there are fewer nodes and less protocol packing and unpacking;
- to reduce complexity because there are fewer interfaces to implement and test. This also benefits interoperability testing;
- to reduce call or bearer setup time thanks to the merging of control plane protocols; and

- to simplify mobility management with little impact on the *GSN+*.

Quality of service

Given the requirement to reduce setup delays, we will now describe one possible evolution of the current 3GPP QoS concept. Apart from allowing for substantially reduced setup delays, the concept aims to give operators an easy and effective means of controlling QoS.

A logical connection through the evolved 3G network is associated with a given QoS level. For the purposes of this description, we call this connection a “tunnel”. A user may have multiple tunnels, possibly associated with different QoS levels. In today’s 3G networks, tunnels comprise a PDP context and an associated radio access bearer (RAB).

One can reduce session setup delays by establishing tunnels in advance; that is, by establishing the tunnels before initiating the user-sessions associated with them. Pre-establishment requires less network signaling during session setup, resulting in shorter session setup delay. The tunnels could be established when a user switches on his or her terminal. This approach may even be used for tunnels associated with guaranteed bitrate (GBR) media, such as VoIP. For pre-established bearers to be feasible and effective, one should

- split the admission control and tunnel establishment procedure into two separate procedures. Otherwise, resources for the pre-established tunnels are reserved unnecessarily;
- give operators better control over which tunnels may be pre-established as well as when they should be established. This may be done by defining network-controlled procedures for setting up the tunnels;
- give operators better control of the QoS levels associated with the (pre-)established tunnels. This may be done by defining network-controlled procedures for setting up the tunnels; and
- give operators better control of the multiplexing application flows into the (pre-)established tunnels. This may be done by giving the network control over the packet filters used to map traffic into the different tunnels (in both the UE and the *GSN+*).

A QoS architecture of this kind would give operators full control of the QoS levels associated with each tunnel, of when the tunnels are to be established, and of the appli-

cation flows mapped into the tunnels. In other words, it gives operators the tools they need to effectively support QoS control.

Link layer solutions

The Rel-6 link layer protocols provide effective support for the peak data rates of HSDPA and E-UL. But looking ahead, the evolved radio access network (RAN) must support envisioned peak data rates of more than 100Mbps. Furthermore, apart from higher peak data rates, the design of the link layer protocols must consider the RAN architecture, reliability, and radio-resource and transport-network efficiency.

The main building block of the link layer protocols is the retransmission function, which ensures reliable and radio-resource-efficient transmission. Hybrid automatic repeat request (ARQ) protocols are known to work well in this context. Therefore, as in Rel-6, a hybrid ARQ protocol will be used between the UE and Node B. If transmission errors occur, the protocols can retransmit efficiently using incremental redundancy and soft-combining techniques.

One drawback of the hybrid ARQ protocol in Rel-6 is the high cost of achieving the low residual block error rates (for example, 10^{-5}) required by higher-layer protocols (such as TCP). This is due to the low reliability of the associated feedback signals. Because these signals are sent frequently, increasing their reliability by sending them with more transmission power is too costly. Three options for circumventing this problem have been considered:

1. Putting a second ARQ protocol layer (the radio link control protocol) on top of the hybrid ARQ layer to operate between the UE and a central anchor node (Figure 1, right). Rel-6 already employs this concept. Apart from its support of increased data reliability, the RLC protocol can provide a high level of security by ciphering up to the central anchor node. It can also support lossless handovers and retransmit data lost in the transport network.
2. Putting a second ARQ protocol layer (the RLC protocol) on top of the hybrid ARQ layer to operate between the UE and Node B. This option increases reliability but without the additional benefits of Option 1. Where RLC retransmissions are concerned, this solution results in lower delays than Option 1. However, given residual error rates (above hybrid ARQ) in the order of 10^{-3} , the impact of this reduced

delay is negligible. A disadvantage of this approach is the need for dedicated mechanisms for transferring context during handovers, which increases complexity.

3. A single-layer ARQ protocol stack with enhanced mechanisms for lowering the residual block error rates, for instance, repeated or coded feedback signals. A drawback of this solution is that mechanisms of this kind introduce additional delay and are more costly in terms of transmission power. In addition, the mobility-related drawback of Option 2 applies.

Seen in terms of the evolved architecture, Option 1 is the preferred design for the link layer protocol: a two-layered ARQ protocol stack where the RLC is terminated in the central anchor node. This design is similar to that used for Rel-6. However, the significantly increased peak data rate will require changes to the RLC protocol. The Rel-6 RLC uses relatively small, fixed-size protocol data units (PDU) and is too inflexible to operate over a wide range of data rates. Small PDUs, in turn, lead to excessive header overhead; large PDUs introduce excessive padding overhead for small packets such as voice over IP (VoIP) frames or TCP acknowledgements. Therefore, it is proposed that each payload or IP packet should be encapsulated in exactly one RLC PDU, the size of which is variable. This change will further increase spectral efficiency by reducing overhead and padding.

Physical layer and radio resource management

Important targets for the long-term evolution of 3G are support of high data rates and the potential to operate in different spectrum allocations. AML-OFDM is an interesting choice of technology (Box B) for providing the necessary high data rates (with correspondingly large transmission bandwidths) and flexible spectrum allocation. The technology also enables the system to adapt the transmission parameters in the frequency domain, thereby satisfying requirements for spectral efficiency. It is also suitable for broadcast transmission.

By varying the number of AML-OFDM subcarriers, one can support different allocations of spectrum ranging from 1.25MHz to 20MHz. The fine frequency granularity offered by AML-OFDM thus facilitates smooth migration, for example, of 2G spectrum. In principle, a GSM operator can migrate on a carrier-by-carrier (200kHz GSM) basis using only a fraction of the available



Figure 2
A musical analogy to OFDM transmission (top) and single-carrier transmission (bottom). See also Box B.

OFDM subcarriers. Furthermore, operation is possible in paired and unpaired spectrum because AML-OFDM supports both time-division and frequency-division duplex operation.

Downlink: OFDM with frequency-domain adaptation

Figure 3 shows the basic structure of the downlink AML-OFDM time/frequency structure. The selected subcarrier spacing of 15kHz allows for operation in a wide range of environments with different propagation characteristics. It is also well-matched to the WCDMA clock rate, thus allowing for simplified implementation of dual-mode UTRA/E-UTRA terminals. To minimize delays, a short 0.5ms frame duration has been selected. A cyclic prefix of 4.7µs is sufficient for handling the delay spread for most unicast scenarios and adds only modest overhead. Reducing the number of OFDM symbols in a subframe makes it possible to extend the cyclic prefix to 16.7µs for very large cells with a radius of 120km or more and with large amounts of time dispersion. Broadcast services transmit the same information from multiple (synchronized) base stations and will also benefit from the long cyclic prefix to handle asynchronism. To a terminal, the received signals from multiple base stations appear as multipath propagation that can be and is exploited by the OFDM receiver.

Channel variations in the time domain are exploited through link adaptation and channel-dependent scheduling as is done in

WCDMA/HSDPA, giving a substantial increase in spectral efficiency. With the evolved radio access, we can go one step further by adapting the transmission parameters not only in the time domain but also in the frequency domain. OFDM techniques can yield large gains in performance where the channel varies significantly over the system bandwidth. Therefore, frequency-domain adaptation becomes increasingly important as system bandwidth increases. Information about downlink channel quality is obtained through feedback from terminals. The Node B allocates the downlink time-frequency resource (to a given user) and dynamically selects an appropriate data rate by varying the output power level, the channel-coding rate, and the modulation scheme. The downlink supports quadrature phase shift keying (QPSK), 16 quadrature amplitude modulation (16QAM) and 64QAM.

Uplink: single-carrier FDMA with dynamic bandwidth

For uplink transmission, power-efficient user-terminal transmission is necessary to maximize coverage. For this reason, single-carrier frequency-division multiple access (SC-FDMA) with dynamic bandwidth is preferred. During each time interval, the base station assigns terminals a unique frequency for transmitting user data and ensuring intra-cell orthogonality, thus avoiding intra-cell interference. For the most part, time-domain scheduling is used to separate users. However, frequency-domain

BOX B, OFDM

- Orthogonal frequency-division multiplexing is a modulation technique that is highly resistant to frequency-selective fading. It is based on multicarrier transmission, where the total data stream is split into several low-bit-rate streams transmitted on separate subcarriers. OFDM is relatively easy to implement.
- In contrast, as the name implies, single-carrier systems, transmit the overall, high-bit-rate data stream over a single carrier. In musical terms (Figure 2), multicarrier transmission could be expressed as the transmission of a sequence of chords as

opposed to the transmission of a sequence of tones.

- The OFDM transmitter is typically implemented using low-complexity inverse fast Fourier transform (IFFT). A cyclic prefix is then added to each symbol prior to transmission. Using an FFT, the receiver can perfectly detect the transmitted signal provided the maximum delay spread in the channel is shorter than the length of the cyclic prefix. Adaptive multilayer OFDM is based on OFDM with added support for multistream transmission and for adapting the transmission parameters to channel conditions.

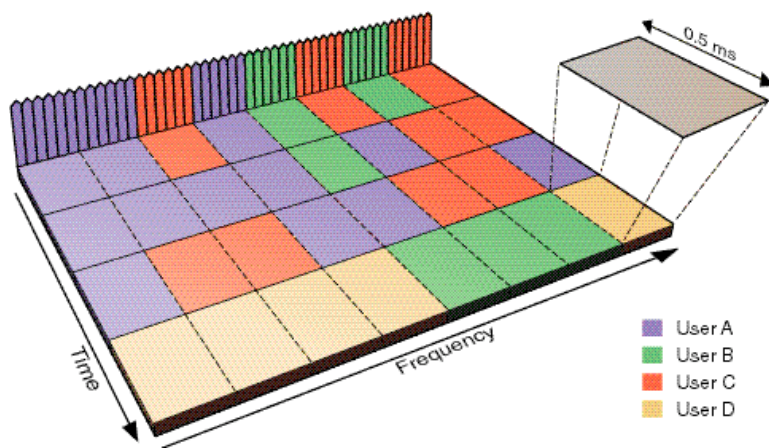


Figure 3
Structure of time/frequency in the downlink.

scheduling may also be used for terminals with either limited power or little data to transmit. Frequency-domain adaptation is not usually used in the uplink due to a lack of information about the channel – this is because terminals cannot continuously transmit a pilot signal that covers the entire frequency domain. Slow power control is used to compensate for path loss and shadow fading. Thanks to the orthogonal uplink transmissions, there is no need for fast power control to handle the near-far problem.

Interference due to multipath propagation is handled at the base station, aided by the insertion of a cyclic prefix in the transmitted signal. The transmission parameters, coding and modulation are similar to those of the downlink transmission.

Multi-antenna solutions

Multi-antenna schemes, including beamforming and multilayer transmission, play a significant role in increasing data rates, coverage and capacity. The potential to use the spatial domain is great and has not been fully exploited to date.

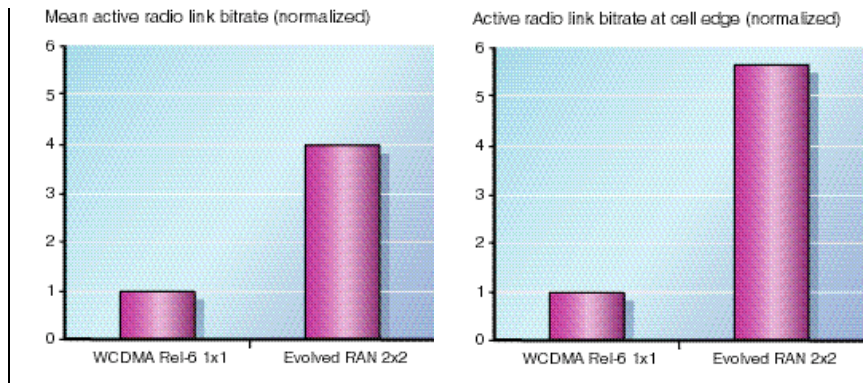
Multilayer transmission, sometimes also referred to as multiple input, multiple output (MIMO) transmission, can be used to increase data rates by transmitting parallel streams to a single user. Such techniques, which are mainly applicable in scenarios where the received signal-to-noise ratio is high and the channel contains rich scattering (small cells or indoor deployments) will primarily be used in the downlink. The ter-

minal separates the data streams by exploiting the channel properties and its knowledge of the coding scheme used in the base station. The multilayer transmission scheme for the long-term evolution must thus be standardized. One promising technique, *selective per-antenna rate control* (S-PARC), adapts the number of layers and data rate per layer to instantaneous channel conditions.

Beamforming, or the use of multiple antennas to form beams, increases the signal-to-noise ratio at the receiver. This technique can be used to increase the coverage of a particular data rate or to increase the spectral efficiency of the system. The increased signal-to-noise ratio is due in part to a larger gain in the direction of the user, and in part to better control of the distribution of spatial interference in the cell. Beamforming can be applied to the downlink and uplink. A standard should be developed that makes beamforming transparent to the terminal (so there will be no need to standardize a particular solution). Algorithms, on the other hand, can evolve over time and be tailored to fit particular needs.

One may even combine these two techniques. Two data streams, for example, can be transmitted by two groups of antennas, where beamforming is employed in each group. Beamforming can thus be used to increase the received signal-to-noise ratio, and multilayer transmission can be used to exploit the increased signal-to-noise ratio as a higher data rate.

Relative gains in mean and cell edge (5th percentile) active radio link bit rate for the evolved RAN with 2x2 MIMO over WCDMA Rel-6 with single input, single output (SISO).



Performance evaluation

Performance is evaluated by measuring the bit rate of the active radio link under different traffic loads. The bit rate of the active radio link is the bit rate a scheduled user re-

ceives. When multiple users share the channel, the bit rate they experience above the MAC layer will be lower.

This measurement is used to assess quality, capacity and coverage. The standard for all comparisons is WCDMA Rel-6, which

is assumed to use single-transmit and single-receive antennas. The evolved RAN concept uses a two-transmit, two-receive antenna (Eigen value-based MIMO) solution with successive interference cancellation (SIC) receivers. No frequency-domain adaptation techniques are employed. Each system is assumed to occupy 5MHz of spectrum. In each case many protocol aspects above the physical layer have been omitted yielding optimistic absolute values. Notwithstanding, it is possible to obtain relative assessments of the gains associated with OFDM and MIMO.

Figure 4 shows the relative gains achieved in the bit rate of the active radio link averaged over the cell (left) and at the cell edge (right). The traffic load in each case is the same. The targets set by 3GPP call for a threefold increase in bit rates averaged over the cell and at the cell edge. The performance evaluation shows that the evolved RAN concept exceeds these targets. Comparisons of capacity and coverage show gains of similar magnitude.

Although the results of the performance evaluation are preliminary and exclude enhancements to higher layer protocols, they

indicate the potential to improve user quality, capacity and coverage, or put another way, to reduce the overall cost of infrastructure for the given coverage and capacity requirements.

Conclusion

To ensure the competitiveness of 3G technology in a long-term perspective, the current standards must evolve. The 3GPP calls this the long-term evolution of 3G.

The fundamental target of the evolution is to reduce user and operator costs and to improve service provisioning. Promising technologies to reach these targets include an evolved architecture and QoS concept, resulting in lower latency, shorter setup times and reduced cost. An evolved link layer and an OFDM-based physical layer that supports multi-antenna solutions will enable higher bit rates and improved coverage and system capacity.

Performance evaluations indicate that the proposed technologies will satisfy the requirements of wireless communication throughout the next decade and beyond.