Providing reliable and efficient VoIP over WCDMA

Mårten Ericson, Lotta Voigt and Stefan Wänstedt

The architecture of the IP Multimedia Subsystem (IMS) defined by the Third Generation Partnership Project (3GPP) provides a solid base for solutions for interconnecting fixed and wireless networks and supplying new services over them, such as voice over IP (VoIP).

Compared with circuit-switched voice, the main end-user benefit of VoIP is the ease with which end users can enrich basic voice service with multimedia services during an ongoing session. Improved end-user flexibility will also probably lead to an increase in traffic and operator revenues. VoIP also has the potential to help operators lower their operating costs by converging their networks into a single all-IP network.

The challenges associated with deploying VoIP over mobile systems include matching the efficiency and quality of circuit-switched voice service while maintaining the inherent flexibility of IP-based services. In the process, one must consider every aspect of the VoIP realization, from compression of IP header overhead to radio access bearer, including quality enhancements and the handling of jitter.

The authors provide an overview of VoIP and its transmission over mobile radio links with available packet-based access (GPRS/EDGE, WCDMA) and dedicated bearers (HSDPA/E-UL). They also describe the capabilities a radio network must have to provide a flexible packet-switched voice service with adequate quality and capacity.

Introduction

The notion of offering VoIP service over mobile networks is not new. The 3GPP has been discussing this topic for years, but has yet to define an actual implementation, in part due to a lack of suitable radio access technology.

Interest in this technology generally falls into one of two main groups or drivers:

- rapid growth in the number of VoIP users on the internet; and
- desire on the part of mobile operators to make their businesses more efficient and profitable.

Compared with circuit-switched voice service, the main end-user advantage of VoIP is the ease with which one can enrich the basic voice service with multimedia services, such as messaging and video during a VoIP session. A continuously growing number of services will influence and possibly change the way people communicate. The addition of presence information, for example, will help determine the form of communication that users employ. Moreover, in the middle of an ongoing session, end users might change from one form of communication to another by adding video to a voice session. Improved end-user flexibility will also probably lead to an increase in traffic and operator revenues.

The advent of VoIP enables established operators to converge their networks into a single all-IP network and to lower their overall operating costs. Similarly, newcomer operators solely need one network for every type of communication (voice and data), which means smaller investments. Being able to deliver every media component over a unified IP infrastructure changes the way one creates and deploys services. Indeed, it changes the very nature of the IP networks themselves.[1]

One price to pay for having this vastly more flexible service is the requirement to transfer IP headers all the way to the recipient (receiving client), which results in a larger total bit rate. Making VoIP over mobile systems into an efficient service with the quality and coverage end users have

BOX A, TERMS AND ABBREVIATIONS

3G	Third-generation mobile system
3GPP	Third Generation Partnership Project
A-DPCH	Associated DPCH
AMR	Adaptive multirate
ARQ	Automatic repeat request
CDMA	Code-division multiple access
CN	Core network
CQI	Channel quality indicator
CSCF	Call/session control function
DCH	Dedicated channel
DPCH	Dedicated physical channel
DTX	Discontinuous transmission
EDGE	Enhanced data rates for GSM evolution
E-UL	Enhanced uplink
F-DPCH	Fractional DPCH
GGSN	Gateway GSN
GPRS	General packet radio service
GSM	Global system for mobile communication
GSN	GPRS support node
HSDPA	High-speed downlink packet access
HSS	Home subscriber server
I-CSCF	Interrogating CSCF
IMS	IP Multimedia Subsystem
IP	Internet protocol
P-CSCF	Proxy CSCF
MAC-hs	Media access control for HSDPA
QoS	Quality of service
PDU	Packet data unit
PF	Proportional fair
RLC	Radio link controller
RNC	Radio network controller
ROHC	Robust header compression
RRC	Radio resource control
RTCP	RTP control protocol
RTP	Real-time transport protocol
S-CSCF	Serving CSCF
SGSN	Serving GSN
SID	Silence descriptor
SigComp	Signaling compression
SIP	Session initiation protocol
TB	Transport block
TPC	Transmit power control
TTI	Transmission time interval
UDP	User datagram protocol
UE	User equipment
UMTS	Universal mobile telecommunications system
UTRAN	UMTS terrestrial radio access network
VoIP	Voice over IP
WCDMA	Wideband CDMA

come to expect from circuit-switched voice service is thus a challenging task. Doing so entails more than merely optimizing radio network performance. In fact, reaching competitive efficiency affects every node in the associated chain.

One other challenge is the prioritization of media flows to guarantee a quality presentation of real-time media. The ability to differentiate between media flows allows operators to provide, guarantee, and charge for different kinds of services.

The introduction of high-speed downlink packet access (HSDPA) and the enhanced uplink (E-UL) will greatly increase transport capabilities in terms of capacity, quality and flexibility. These enhancements of the radio interface are a prerequisite for a resource-efficient deployment of flexible, real-time IP services.

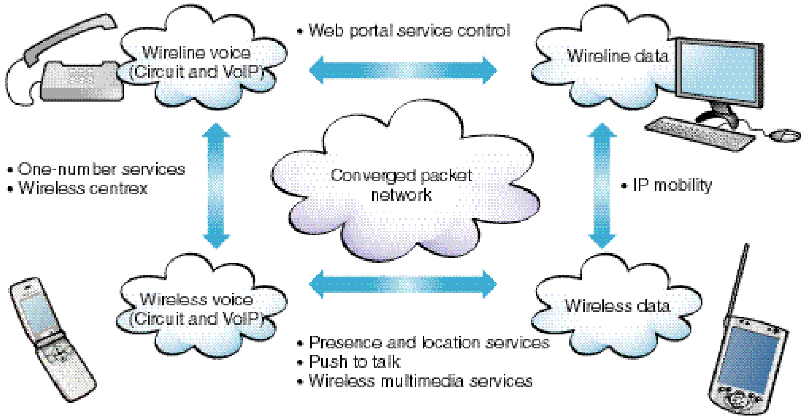


Figure 1
Example of all-IP.

IMS

The 3GPP has adopted the session initiation protocol (SIP) as a basis for a new session-control layer in the core network of third-generation mobile systems (3G). Anxious to take advantage of SIP's real-time session-initiation features, the mobile industry has developed specifications for an IP Multimedia Subsystem (IMS). The vision is an independent (decoupled from access layers) platform for IP multimedia services that facilitates rapid deployment of new IP-based services. IMS introduces a service layer that allows operators to efficiently deploy and manage IP multimedia services as well as maintain control over subscriber applications in order to introduce and charge for differentiated services (Figure 1).

The wireless and wireline industries anticipate that the 3GPP IMS specification will enable the introduction of SIP services, such as push to talk[2], presence, and VoIP combined with video or image sharing. Having services that cover wireline and wireless access gives operators a needed advantage for growing their customer base in a short time.

Architecture overview

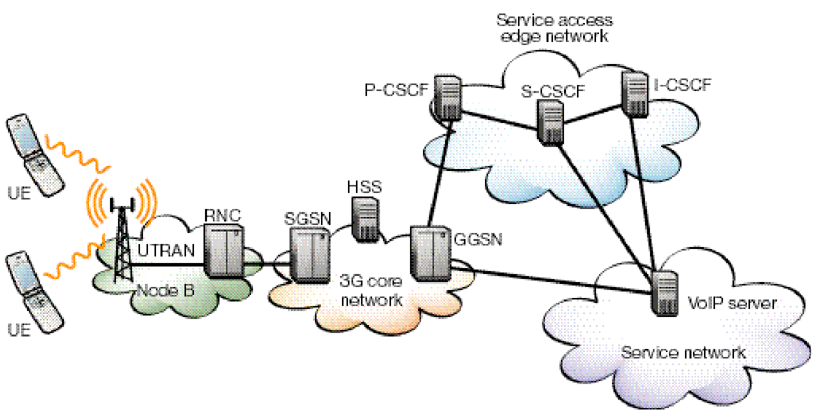
Figure 2 shows a simplified overview of the system nodes that deal with traffic in the user plane. VoIP is an IMS service that employs SIP signaling nodes to establish voice sessions. There are three SIP signaling nodes, as follows:

- Proxy call/session control function (P-CSCF);
- Serving CSCF (S-CSCF); and
- Interrogating CSCF (I-CSCF).

Traffic in the user plane is routed directly from the gateway GPRS support node (GGSN) to the VoIP server. The proxies and SIP nodes are located logically at the edge of the service access network.

A simplified view breaks the UMTS/WCDMA network down into two parts: the core network (CN) and the UMTS terrestrial radio access network (UTRAN). The core network, which consists of the GGSN, serving GPRS support node (SGSN), and home subscriber server (HSS), provides an IP

Figure 2
UMTS/WCDMA system overview.



transmission plane from the service network or internet to the radio network controller (RNC) that serves the mobile user. The UTRAN consists of the RNC and Node Bs. A third component, user equipment (UE, for example, a mobile handset), gives end users access to the network and services.

During a VoIP session the media is sent end-to-end from one client to another. Most session control signaling occurs before and after the voice conversation (session initiation and session termination). IMS session setup generates considerable traffic over the air interfaces between the inviting IMS client and the P-CSCF and over the air interface between the invited IMS client and the P-CSCF. To keep setup reasonably short, the inviting IMS client, the P-CSCF, and the invited IMS client each support and make use of SIP compression (SigComp). During the conversation, the system might need to send or receive additional messages, for example, to re-register, to update application parameters, or to provide notification of incoming calls.

The RTP control protocol (RTCP) is used to report on-hold senders and recipients and to detect the activity of links at the remote end. Presence applications might also generate SIP signaling.

VoIP characteristics

Real-time, interactive audio over IP networks often suffers from packet loss and jitter (variations in network delay). Present-day quality-of-service (QoS) requirements for mobile VoIP are set by circuit-switched services. High-quality circuit-switched voice usually has end-to-end delay of at most 250ms, less than 2% voice frame loss, and virtually no jitter.

Delay and jitter

Delay and its consequences are some of the most important aspects affecting VoIP service. The most important nodes in this context (that is, those that contribute the most delay) manage voice encoding and packets, the radio network, and possibly IP transport if the geographical distance between users is great. Distance also affects delay in circuit-switched voice.

The mobile handset encodes voice frames, assigns them to IP packets and transmits them over the air interface. The encoding delay is determined by the time (ms) it takes to encode the voice frames and by the duration of the look-ahead interval.

Voice packets sent mouth-to-ear over one or more wireless channels are affected by delay and loss. To compensate for variations in delay, the receiving terminal employs a play-out buffer. In essence, packet flows are held back to ensure continuous play-out. Packets that arrive too late are discarded. There are two kinds of play-out algorithms: fixed and adaptive. Adaptive play-out schemes are more common in VoIP systems because they can adjust play-out delay for each burst of conversation. The scheduled play-out delay is a tradeoff between buffer losses and end-to-end delay. The value is selected so as to maximize the quality of voice communication. Large play-out delay decreases packet loss from late arrivals but hampers interactivity between the communicating parties; small play-out delay improves interactivity but causes greater buffer losses and degrades voice quality. After leaving the dejitter buffer, the voice frames are depacketized and fed to the decoder.

Bit rates

The adaptive multirate (AMR) voice codec is mandatory for voice services in GSM and WCDMA systems. It is also a reasonable assumption for VoIP service. During bursts of conversation, with AMR mode 12.2kbps, the VoIP application generates 32-byte voice payload at 20ms intervals. During silent periods, a 7-byte payload carries a silence descriptor (SID) frame at 160ms intervals. A typical VoIP protocol stack, which employs the real-time transport protocol (RTP), is encapsulated in the user datagram protocol (UDP). This, in turn, is carried by IP. The combined effect of these protocols is a 40-byte IPv4 header or a 60-byte IPv6 header.

Radio networks have limited transport capacity. Obviously, a header overhead of 60 bytes seriously degrades the spectral efficiency of VoIP service. Each IP packet contains RTP/UDP/IP headers. Therefore, if one 20ms voice frame is put into each IP packet, the ratio of overhead to payload is between 60:40 and 70:30.

Robust header compression (ROHC) reduces the size of the IP/UDP/RTP headers to as little as 1 byte (Figure3). The effective compression is based on the assumption that most fields in the combined IP/UDP/RDP header are constant or introduce constant change throughout the session. Maximum compression (1 byte) can only be reached by imposing limitations; a more typical compressed header size is 3 or 4 bytes. The

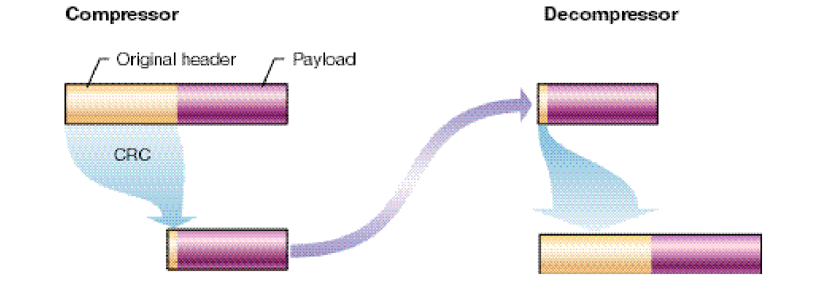


Figure 3
Robust header compression (ROHC).

ROHC operation is based on synchronized compression and decompression contexts. Irregularities in the transmitted stream, for instance by DTX operation or lost packets, introduce slightly larger compressed headers.

ROHC is very robust, handling practically every common packet loss event in mobile environments. In extreme conditions a feedback mechanism can be used to facilitate recovery when synchronization is lost between compressor and decompressor. Recovery entails sending a larger compressed header with context and update information. Under regular operating conditions, ROHC yields a very low bit-rate requirement. The bit rate for VoIP, for example, needs be only 10 to 20% greater than for circuit-switched service with the same coder. In all likelihood, however, the bit rate will vary during a conversation due to header compression and DTX. Moreover, the service should be able to accommodate the addition of new services to the session, which puts greater bit-rate requirements on the bearer. The selected VoIP (or IMS) bearer must thus be able to transport a data stream with varying bit rates.

High-speed packet access

HSPDA

High-speed downlink packet access (HSDPA) is shared channel transmission that includes features such as fast link adaptation, fast hybrid automatic repeat request (ARQ), and channel-dependent scheduling. HSDPA was designed for services with high throughput requirements but it can also be used for VoIP. Generally speaking, VoIP and high-throughput services have different characteristics that require different treatment. In the first place, VoIP media flows must be given priority over interactive traffic. Given that HSDPA maintains individual scheduling buffers to exploit changes in radio conditions per link, it can also be used to take service characteristics into account – for example, in scheduling schemes, by mapping different services or media flows on separate priority queues. The use of separate priority queues makes it possible to optimize HSDPA scheduling for VoIP. Below are some other features of the HSDPA radio access solution which are necessary for efficient VoIP transport:

- radio link controller unacknowledged mode (RLC UM) – to block retransmission at the RLC level;
- optimized packet discard timer;
- delay-based priority – to include the delay of packets in the scheduling process;
- code multiplexing – to transmit to more than one user per transmission time interval (TTI); and
- reduction of power overhead, for instance, by means of a fractional dedicated physical channel (DPCH).

Power allocation

WCDMA Release 5 can handle dedicated and HSDPA channels simultaneously. VoIP is supported over HSDPA; support for circuit-switched voice is retained over dedicated channels (DCH). HSDPA makes dynamic use of available base station transmit power.

Channels

HSDPA can employ up to 15 codes with a spreading factor of 16. The number of codes is a configured value according to the resources needed for packet-data services. Users generally share the HSDPA resource in time. All codes and power are thus allocated to one user during the TTI (2ms). Notwithstanding, the codes can also be multiplexed, which means it is possible to transmit to several handsets during a given TTI. Each handset is required to listen to up to four different code "sets". An HSDPA control channel is used for each code set. This benefits services such as VoIP that do not have large packets to transmit and do not require all the code resources. Notwithstanding, because each HSDPA control channel also needs power, the number of simultaneous channels should be limited (even for VoIP).

A dedicated channel, called associated DPCH (A-DPCH), is associated with each active HSDPA user. As with all other dedicated channels, the A-DPCH is power-controlled. Ordinarily, the spreading factor is set to 256. The A-DPCH is used for signaling and transmit power control (TPC) commands for controlling power in the uplink. But because it accounts for a relatively large amount of the total base station transmit power and codes, 3GPP has developed the F-DPCH, which only transports the uplink TPC bits; signaling is sent inband.

Scheduling

Because HSDPA is a shared resource, user access to the channel must be scheduled. The media access control for HSDPA (MAC-hs) sublayer, located in the Node B, handles scheduling using a 2ms TTI. The data rate for each terminal can be estimated from the reported channel quality indicator (CQI). Two sets of algorithms are used: QoS-aware algorithms and basic algorithms, which cannot differentiate between services or the QoS demands of different users. The proportional fair (PF) algorithm calculates priority p for each user i:

$$p_i = \frac{R_i(n)}{r_i(n)}$$

The user with the greatest priority p is then scheduled for transmission in the next TTI. $R_i(n)$ is the estimated supported bit rate in the next TTI; n is the current scheduling interval (TTI); and r_i is the filtered user throughput for user i.

For conversational services (VoIP) over HSDPA it is beneficial to include a time-delay factor. To measure delay, the scheduler time-stamps each packet as it arrives at the priority queue. The greater the delay, the greater the priority. Hence,

$$p_i = \frac{R_i(n)}{r_i(n)} f(D_i(n), D_{th}(n)),$$

where $D_i(n)$ is the delay for user i at TTI n; D_{th} is the delay threshold; and $f()$ is a delay function. The delay function $f()$ must be designed so that the benefits of transmitting during fading tops (high $R_i(n)$) are not cancelled.

Enhanced uplink

The enhanced uplink (E-UL) is described in an extension to WCDMA Release 6. Compared to a dedicated channel, the enhanced uplink reduces delay in the uplink, increases uplink data rates, and increases uplink capacity.

Short TTIs reduce overall delay and speed up retransmissions. The shortest TTI is 2ms. The E-UL also supports the 10ms TTI used by dedicated channels.

Fast hybrid ARQ with soft-combining reduces the number of retransmissions and the time between them.

Fast scheduling allows for rapid reallocation of resources between handsets by exploiting the burstiness of packet data transmissions. It also allows the system to rapidly adapt to variations in interference.

The shared resource on the uplink is interference headroom, not power and codes. Scheduling over the E-UL is different from that over HSDPA, in part because the shared resource differs. For symmetric, real-time services, such as VoIP, it may be advantageous to omit scheduling altogether and instead allow the UE to send data at any time by respecting the configured bit rate and ignoring scheduling commands from Node B (similar to DCH). This approach minimizes signaling overhead and scheduling delay. However, in scenarios with mixed services another strategy might be necessary to maximize overall capacity to meet QoS requirements.

Delay budget

Low delay is one of the most important criteria for maintaining high-quality VoIP service. But to attain high VoIP capacity for HSDPA the scheduler must have sufficient time to manage scheduling. UTRAN must thus cater for the functionality to minimize latency in every involved node. Figure 4 shows an example of the delay budget for the major nodes. Of the assumed 250ms end-to-end delay, about 100ms is available for scheduling in the downlink. E-UL is important because it decreases delay

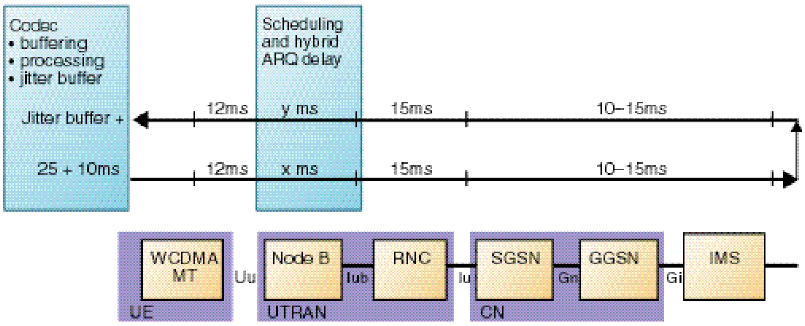


Figure 4
End-to-end delay assuming E-UL and HSDPA.

in the uplink compared to the direct channel. As mentioned above, real-time, interactive audio over IP networks often suffers from packet loss and variations in network delay (jitter). The scheduler adds jitter because it distributes resources among users according to radio conditions, bit rate, or some other criterion. Some packets must thus wait to be scheduled, which introduces delay jitter. What is more, packets that have not been correctly received must be retransmitted with hybrid ARQ until they have been correctly received or are obsolete. Hybrid ARQ retransmissions add even more delay jitter to the packet flow.

VoIP capacity example

In the following example, the resultant maximum end-to-end delay, including jitter buffer, is slightly larger than that of circuit-switched voice, or around 300ms. Note: This capacity comparison does not include SIP or radio resource control (RRC) signaling.

In a shared media transport channel, such as HSDPA, VoIP media must be given priority over other media. Scheduling must be designed carefully and parameter choices should optimize VoIP and interactive media. One way of attaining high capacity is to increase available time in the scheduler. However, to retain low end-to-end delay, all other nodes must reduce their processing time. The jitter buffer at the terminal side must also cope with a variety of delay intervals.

Simulations are used to estimate capacity for VoIP over HSDPA. Figure 5 shows the capacity in terms of satisfied users and system delay for a specific scenario. The chan-

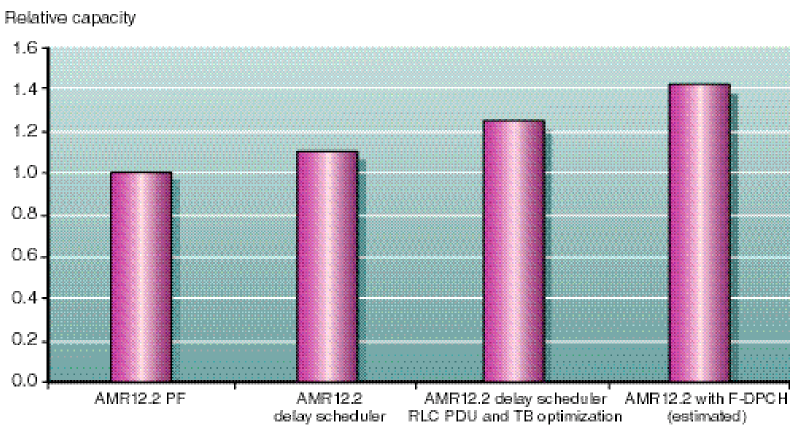


Figure 5
Simulated capacity of VoIP over HSDPA using different schedulers.

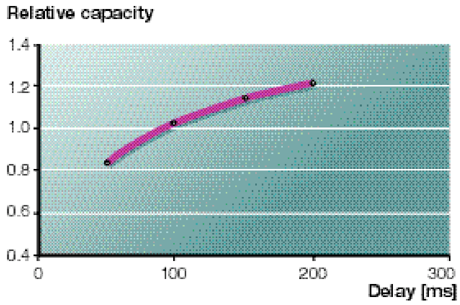


Figure 6
Capacity of VoIP over HSDPA as a function of the delay.

nel model is Typical Urban for 5MHz; UE velocity is 3km/h. Voice activity is 50%, and IP header compression has been employed. Likewise, the AMR 12.2 codec has been employed. In this example, the size of the transport blocks (TB) has not been optimized for the AMR 12.2 codec. Consequently, some padding is necessary.

The proportional fair scheduler is not an optimal scheduler for delay-sensitive traffic, such as voice. It does not maximize the VoIP capacity and may give excessive delay variation. The introduction of a delay scheduler, however, increases capacity considerably and keeps the delay within reasonable limits. The optimization of RLC packet data unit (PDU) and MAC-hs transport block sizes further increases capacity.

The VoIP simulations use A-DPCH with spreading factor 256 per user. A-DPCH accounts for a large amount of the total base station transmit power overhead. The introduction of F-DPCH, by contrast, decreases the transmit power overhead, which means more power can be allocated to VoIP.

With these enhancements we can show that VoIP has the potential to match or even exceed the capacity of circuit-switched voice, assuming full optimization of the radio-access network and UEs.

Figure 6 depicts the relationship between maximum allowed delays in the scheduler and capacity (same scenario as above). Capacity increases as scheduling time increases, thanks to an improved ability to transmit during fading tops. Capacity can thus be traded for increased dealy.

Coverage

HSDPA and E-UL enable hybrid ARQ retransmissions, which further increases coverage. Coverage can thus be traded for delay. The downlink coverage of VoIP over HSDPA is expected to match the cell planning for a typical WCDMA network.

With E-UL, a 2ms TTI gives the shortest delay but also provides less coverage than a 10ms TTI. The drop in coverage may be overcome by multiple hybrid ARQ retransmissions in the uplink, but due to increased delay the initial benefit of a 2ms TTI is lost. On the other hand, with a 10ms TTI and one retransmission, the uplink coverage will match the cell planning of a typical WCDMA network. To ensure wide area coverage, in the event that not all cells have HSDPA or E-UL, VoIP can be transported over WCDMA DCH radio bearers or EDGE.

DCH

The choice of radio bearer governs the flexibility of combinational IP services. A bearer that is optimized for high capacity (for a predefined codec) generally limits flexibility. The DCH solution must be able to adapt to changing bit-rate requirements, for example, when a new service is added. A general advantage of a DCH bearer is that it has almost fixed delay (over radio, assuming RLC UM). A general disadvantage is that it is somewhat rigid or unforgiving of variations in bit rate. For this reason and to attain greater capacity, HSDPA is the preferred choice for VoIP and other IMS services.

EDGE

In general, VoIP over EDGE faces the same kinds of challenges as VoIP over HSDPA, for example, latency, efficiency and coverage. To avoid excessive latency, one can run RLC UM as is done in HSDPA. Some benefits of EDGE are the highly developed support for QoS and the inherent flexibility of this support.

GSM/EDGE is widely deployed, and many of today's terminals can, in principle, support VoIP. This makes GSM/EDGE an attractive alternative for enhancing coverage for VoIP over DCH/HSDPA. Where DCH/HSDPA coverage is scarce or unavailable, GSM/EDGE can be used to extend the coverage area for VoIP service.

Test beds

Ericsson is running trial systems with multimedia based on VoIP. These trial activities, which are based on WCDMA R99 and HSDPA together with E-UL, give operators confidence regarding the overall end-to-end service.

The multimedia service over HSDPA and E-UL is complemented with mobile broadband, which includes fast web browsing service and the uploading and downloading of e-mail at high data rates.

Conclusion

VoIP over wireless is a key component behind the all-IP network paradigm and new service flexibility. The idea of all-IP networks supports the convergence of fixed and mobile networks, thereby reducing operators' capital expenditures (CAPEX) and operating expenses (OPEX).

A transition to packet-switched transport gives mobile networks unprecedented flexibility – flexibility that will enable a plethora of combinational services that greatly enhance the user experience. Presence information, for example, will tell users whether or not a recipient is currently available. A benefit of potential new services is greater traffic volumes, which translates into greater operator revenues.

Although VoIP reduces operating costs, capacity over radio is also an important aspect. When used in combination, HSDPA and E-UL provide capacity for VoIP that is on par with that of present-day circuit-switched voice services. More importantly, HSDPA and E-UL facilitate all possible combinations of IMS services. For cells without HSDPA coverage, VoIP traffic can be transported over packet-switched WCDMA radio bearers or EDGE bearers.

Ericsson will time the introduction of E-UL to coincide with the availability of terminals. Early terminals and PC cards will be available during the second half of 2006. Handsets will enter the market in early 2007. By evolving WCDMA, Ericsson stays true to its established tradition of supplying future-proof products.

REFERENCES

1. Olsson, U.: Toward the all-IP vision. Ericsson Review, Vol. 82(2005):1, pp. 44-53
2. Medman, N., Svanbro, K. and Synnergren, P.: Ericsson Instant Talk. Ericsson Review, Vol. 81(2004):1, pp. 16-19