Ericsson Instant Talk

Niclas Medman, Krister Svanbro and Per Synnergren

Push to talk is a quick and informal way of communicating person-to-person and with groups. With a simple push of a button, users can activate voice communication with friends and family—much the same way as using walkie-talkies or private mobile radios. But because it is a mobile communication service, push to talk also enjoys the range and wide area coverage of traditional mobile services.

Ericsson Instant Talk is a voice-over-IP (VoIP) service set up using the session initiation protocol (SIP). The technical realization is based on a total business approach that gives operators a complete solution consisting of the Instant Talk Application Server, the IP Multimedia (IPMM) system, terminal clients, and professional services.

Ericsson's practice of basing products on open standards—to ensure interoperability and rapid uptake of service—also applies to Instant Talk. The solution is fully compliant with the push to talk over cellular (PoC) specification. It has also been built with service evolution in mind, giving operators the ability to add new features and services to enrich the service and the IPMM.

Introduction

Push to talk is the common name for half-duplex voice services activated by pressing a button. PoC is the name of the open specifications for this service. In the consumer segment, push to talk allows users to stay in touch with friends and coordinate leisure activities, such as visits to the cinema or simultaneous communication with a group of family members. In the enterprise segment, it can be used to share information in a group—for instance, a field technician can use it to ask colleagues for help or advice.

Ericsson's Instant Talk solution complies with the PoC specifications. Its users can make person-to-person calls or create an *ad hoc* group call. At present, it is available for implementation in GPRS/EDGE and CDMA2000 networks. WCDMA solutions will follow. PoC and Instant Talk are based on the IP Multimedia System (IMS)—evidence that Ericsson is taking concrete steps toward realizing IP services on IMS in mobile networks.[1,2]

SIP-based service

Ericsson Instant Talk is an important step toward making IP-based services an integral part of the mobile service offering and of making the Internet protocol a dominant transport technology for mobile applications. The solution is controlled by SIP, whose philosophy of client-to-client control over multimedia session establishment while drawing on support from servers in the network distinguishes it from other common application layer protocols (FTP, HTTP).[3]

Because SIP can establish, modify and terminate multimedia sessions between two or more clients, it facilitates one-to-many communication, which is a key feature of Instant Talk. Clients use the session description protocol (SDP), contained in the body of SIP messages, to describe what kinds of media they can use and how the media should be transported. End-to-end negotiation of media types and transport permits tandem-free operation—that is, the coded bit stream that contains the talk burst is sent directly to the encoder in the receiving terminal without being decoded in the base station subsystem (BSS). Thanks to SIP end-to-end negotiation, new media coders/decoders (codec) may be introduced in terminals as they become available.

Signaling compression

When designing SIP, the Internet Engineering Task Force (IETF) had the large bandwidth and low latencies of the public Internet in mind, which is to say the size of SIP messages was not a priority. But in the context of mobile environments the size of SIP messages can be a drawback. In mobile systems, radio spectrum is an expensive resource that must be used wisely. Operators must always try to maximize data transport

BOX A, TERMS AND ABBREVIATIONS

3GPP	Third-generation Partnership Project
AMR	Adaptive multirate
ASCII	American Standard Code for Information Interchange
AUC	Authentication center
BSS	Base station subsystem
CDMA	Code-division multiple access
CS	Circuit-switched
CSCF	Call/session control function
DNS	Domain name service
EDGE	Enhanced data rates for GSM evolution
EVRC	Enhanced variable rate codec
FTP	File transfer protocol
GLMS	Group list management server
GPRS	General packet radio service
HLR	Home location register
HSS	Home subscriber server
HTTP	Hypertext transfer protocol
IETF	Internet Engineering Task Force
IMS	IP multimedia subsystem
IP	Internet protocol
IPMM	IP Multimedia
IS/IT	Information system/information technology
MRF	Media resource function
OMA	Open Mobile Alliance
PoC	Push to talk over cellular
QoS	Quality of service
RTCP	RTP control protocol
RTP	Real-time transport protocol
SDP	Session description protocol
SIP	Session initiation protocol
UDP	User datagram protocol
UDVM	Universal decompression virtual machine
URI	Uniform resource indicator
WCDMA	Wideband CDMA

efficiency over the air interface. This problem has been addressed through signaling compression (SigComp), another IETF standard initiated by Ericsson. SigComp is a versatile compression framework that can use any compression algorithm to compress ASCII-based protocols, such as SIP, to a fraction of their original size, reducing both bit rate requirements and transport latency.

SigComp sessions are initialized during the registration phase of the Instant Talk service. The SigComp compressor generates a decompression byte code that contains a preferred compression algorithm and transmits it to the decompressor. The decompressor, called the universal decompression virtual machine (UDVM), decodes the byte code and adopts the decompression algorithm, making it ready to receive compressed SIP messages. Extended operations (an important feature) enable SigComp to learn from SIP messages.[4] Therefore, SigComp becomes increasingly efficient as compression and SIP signaling proceed.

SigComp is a critical component of Ericsson Instant Talk because it can significantly reduce the number of bits sent over the wireless link (compression ratios of 8:1 are not unusual).[5] It greatly reduces the transmission delay of SIP messages, and consequently, the time it takes to establish an Instant Talk session.

Technical realization

Ericsson has taken a total business approach in its technical realization of Intant Talk, providing the infrastructure, terminal clients and professional services, and ensuring availability of terminals. Guided by this approach, Ericsson will continue to optimize performance by implementing enhancements in the radio access network, mobile core network, application servers and clients. Ericsson's ability and willingness to influence the nodes in the end-to-end path of telecommunications systems set it apart from other vendors of push to talk solutions. The commercial launch of Ericsson Instant Talk will include vertical service assurance.

One more important factor is that Ericsson bases its products on open standards. Therefore, handset interoperability will not be an issue. Operators and end-users will be able to choose terminals and terminal vendors freely.

Instant Talk consists of three main parts: the IPMM, the Instant Talk application server, and the handset client.

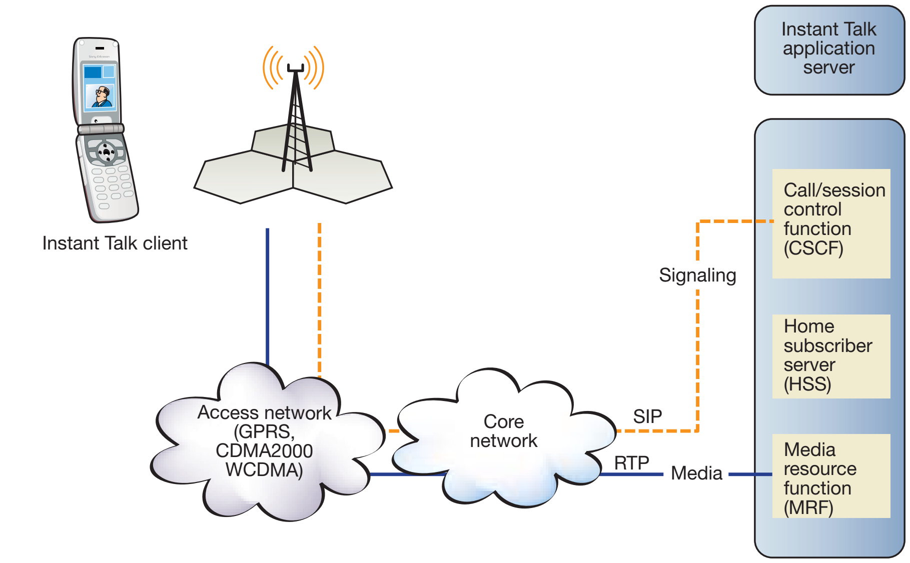

Figure 1
Ericsson Instant Talk high-level architecture.

IP multimedia system

Figure 1 shows a schematic drawing of the nodes involved in the Instant Talk service. At the heart of the solution is the Ericsson IPMM system (Figure 2), which complies with the principles of the IMS standard (3GPP) that was drafted to bring SIP-based communications to the wireless market.[2] The IMS can be deemed a "new" domain that has been added to the mobile core network to support a wide range of SIP-based applications, such as Instant Talk, instant messaging, and presence services.

The IPMM architecture includes the call/session control function (CSCF), the media resource function (MRF), and the home subscriber server (HSS). The CSCF is the teminal's first point of contact in the IPMM domain. All SIP signaling is routed through the CSCF, which also performs SigComp. The role of the CSCF is to handle subscriber registration, and to support the establishment, modification and release of Instant Talk sessions. The CSCF ensures interoperability with telephone systems and network addressing mechanisms by querying domain name service (DNS) servers to map SIP uniform resource identifiers (URI) or E.164 numbers to network addresses.

In an IPMM-based service, media is directed to, replicated in, and distributed from

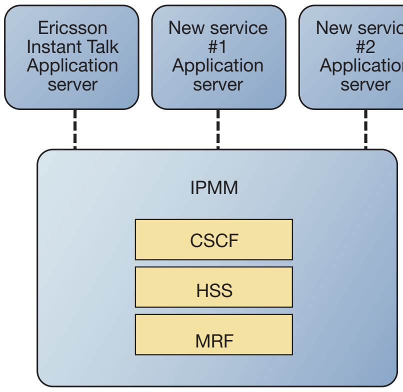

Figure 2
The IPMM system is built with service evolution in mind.

the media resource function. Because Instant Talk is a half-duplex service, the media resource function must prevent two or more users from sending media at the same time. This is called talk burst control. The MRF employs a request/response mechanism to control transmission rights. Users who want to transmit must wait until their requests have been granted. The MRF can also revoke transmission rights when a user abuses the service.

The HSS maintains the Instant Talk subscriber profile by keeping track of the core network node tasked with handling the subscriber. An evolution of the home location register (HLR) and the authentication center (AUC) used by all IMS services, the HSS also handles Instant Talk subscriber authentication and authorization functions.

Instant Talk application server

The Instant Talk application server is basically a database tool that handles subscriber data during call set-up

- to ensure that the called party is a subscriber;
- to determine who is to be included in the group during the call;
- to determine whether or not users have activated Do-not-disturb mode; and
- to check whether the user has activated manual or auto-answer mode.

The application server also stores and passes down rules and regulations to the MRF. Local policies stored in the application server typically include timer values, such as remaining talk time in a user's account.

The Instant Talk application server also supports group list management server (GLMS) functionality, which enables users with list management operations to create, modify, retrieve and delete the groups and contact lists needed for the Instant Talk service. The GMLS also provides storage space for groups and lists.

Client

Ericsson works with third parties to integrate clients into terminals from different vendors. Pre-integrated clients can be optimized for the hardware platform on which they run. For instance, to set its clients apart from downloadable clients, Ericsson can make use of the native voice codec implemented for circuit-switched calls, to increase performance and battery life.

To ensure maximum voice quality, even in environments with limited bandwidth, Ericsson employs the adaptive multirate (AMR) codec and the enhanced variable rate codec (EVRC). The AMR codec is used in EDGE, GPRS and WCDMA terminals; EVRC is used for CDMA2000 clients.

Besides SIP, Ericsson Instant Talk clients use the real-time transport protocol (RTP) to carry real-time data generated by the voice codec. Frame bundling is used to reduce the effect of relatively large headers in the IP/UDP/RTP layers. This means that several voice codec frames are sent in one RTP packet. As dictated by IETF, the frame-bundling feature is supported in the AMR and EVRC payload format. Built-in functionality in the MRF makes it possible to order clients to change the number of frames in an RTP packet to adapt to current network conditions. The MRF employs the RTP control protocol (RTCP) to monitor the network.

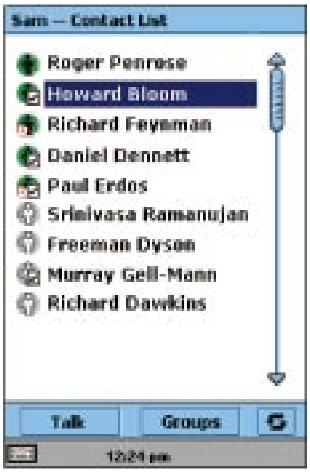

Figure 3
Mobile presence makes everyday communication easier.

Standardization

The Ericsson Instant Talk solution is based on the PoC industry standard, initially developed by an industry consortium made up of Ericsson, Motorola, Nokia, and Siemens mobile, and supported by AT&T Wireless Services, Cingular, Sonim Technologies, and Sony Ericsson. It is an open, published and demonstrated interoperable multi-vendor specification. PoC Phase I includes detailed specifications that mandate the requirements and architecture of the solution. In additon to the Stage 1 and 2 specifications, the consortium also developed a complete Stage 3 suite of PoC specifications that stipulate signaling flows, user-plane flows (including talk burst control mechanisms), and group and list management methods. One of the specifications applies specifically to the radio access network, describing the recommended QoS attributes of the radio link and the operation of the voice codec. To avoid market segmentation, Ericsson has striven for an open standard that avoids lock-in and creates interoperability between operators and between terminals from different vendors.

Ericsson and the other members of the consortium have submitted the Phase 1 specification to OMA. The OMA process is moving forward, and the PoC Phase 1 specification has formed a basis for the OMA PoC standard. Further development will include new features in the PoC specifications, such as a standardized presence solution (Figure 3), an enhanced authentication method, and a network-to-network interface to enable charging between operators and to make the PoC solution compatible with additional radio access networks (WCDMA and CDMA2000). As the OMA PoC standard evolves, the Ericsson Instant talk product will be updated.

Conclusion

Ericsson Instant Talk is the first of many IP-based applications that will be available over IPMM, which is based on the 3GPP IMS standard. The applications running on IPMM are controlled using SIP and SigComp (to make SIP efficient in mobile communications).

Instant Talk consists of three parts: the IPMM system, which can be reused in future applications, the Instant Talk Application Server, and the handset client. Instant Talk is fully compliant with PoC specifications, which ensures interoperability and facilitates rapid uptake of service.

REFERENCES

1 Heino Hameleers and Christer Johansson, “IP Technology in WCDMA/GSM core networks," Ericsson Review, Vol. 79(2002):1, 14-27, 2002.
2 Peter Granström, Sean Olson and Marc Peck, "The future of communication using SIP," Ericsson Review, Vol. 79(2002):1, 28-35, 2002.
3 Gonzalo Camarillo, “SIP demystified,” McGraw-Hill, 2002
4 Mats Nordberg, Hans Hannu, Jan Christoffersson and Lajos Zaccomer, "Improving SigComp performance through extended operations," Proceedings of VTC Fall 2003, Vehicular Technology Conference, Orlando, October 2003.
5 Jan Christoffersson, Hans Hannu, Stefan Forsgren, Stefan Wänstedt and Krister Svanbro, "Reducing call setup delays using SIP/SDP compression," Proceedings of RVK 02 - Radiovetenskap och Kommunikation 02, Stockholm, June 2002.