Proceedings of the 4th WSEAS Int. Conf. on Information Security, Communications and Computers, Tenerife, Spain, December 16-18, 2005 (pp548-553)

A lightweight and scalable VoIP platform based on MGCP/H.323 interworking and QoS management capabilities

T. DAGIUKLAS[1], K. IOANNOU[2], A.GARMPIS[3]

[1] Department of Telecommunications and Network Systems
TEI of Messologi
Nafpaktos
GREECE

[2] Department of Electrical and Computer Engineering
University of Patras
Rio, Patra
GREECE

[3] Department of Applied Informatics in Management & Economy,
Technological Educational Institution of Messolonghi,
Messolonghi, 30100
GREECE

Abstract: - This paper presents a lightweight VoIP platform based on H.323/MGCP interworking and QoS capabilities. The lightweight feature is accomplished by introducing two new schemes in the platform; the use of the SCTP as a transport protocol and the use of the Compressed-RTP as a compressed implementation of the RTP. The MGCP/H.323 interworking provides the platform with the essential scalability so that many different VoIP scenarios can be implemented. The whole architecture consists of a Call Agent that provide central management to the whole system and of a Media Gateway that is responsible for carrying out all the essential media transformation. Finally the VoIP platform uses a QoS management module in order to guarantee certain level of voice quality to the users. This module is able to deal with situations where the network load is significantly affected by the voice packets produced by the proposed platform. The whole architecture bears many innovative characteristics and presents the way for a normal interworking of all these modules. Furthermore, it takes into account that many products in the market are already based in either H.323 or MGCP and tries to propose a way for a smooth co-operation between these two protocols.

Key-Words: - Call Agent, Multimedia over Packet-based Networks, Quality of Service

1 Introduction

We are experiencing a technology revolution in the field of telecommunications and information technology. Existing Voice Networks using circuit switching have dominated the telecommunication area for the many decades. However the last years, the packet based traffic have steadily increased and will soon exceed the voice traffic [1],[2]. As a result, the convergence of the circuit switch networks with the packet-based networks became of great importance. The telecommunications operators and the Internet Service providers are both affected by this fact. Several protocols have been deployed towards the way of convergence. The Voice over IP platforms started growing rapidly. This explosion leads to many different solutions expressed by a great number of protocols. Of course this fact created several problems as the lack of interoperability and scalability. Nowadays there are some dominant protocols like H.323, SIP and MGCP and many products are based on them.

This paper proposes an architecture that will lead to the interworking of different protocols based on MGCP and H.323. The use of these two protocols are imported in order to obtain a scalable platform that consists of a Media Gateway (MG), where all the media transformations take place and of a Call the platform resides. This intelligence includes the

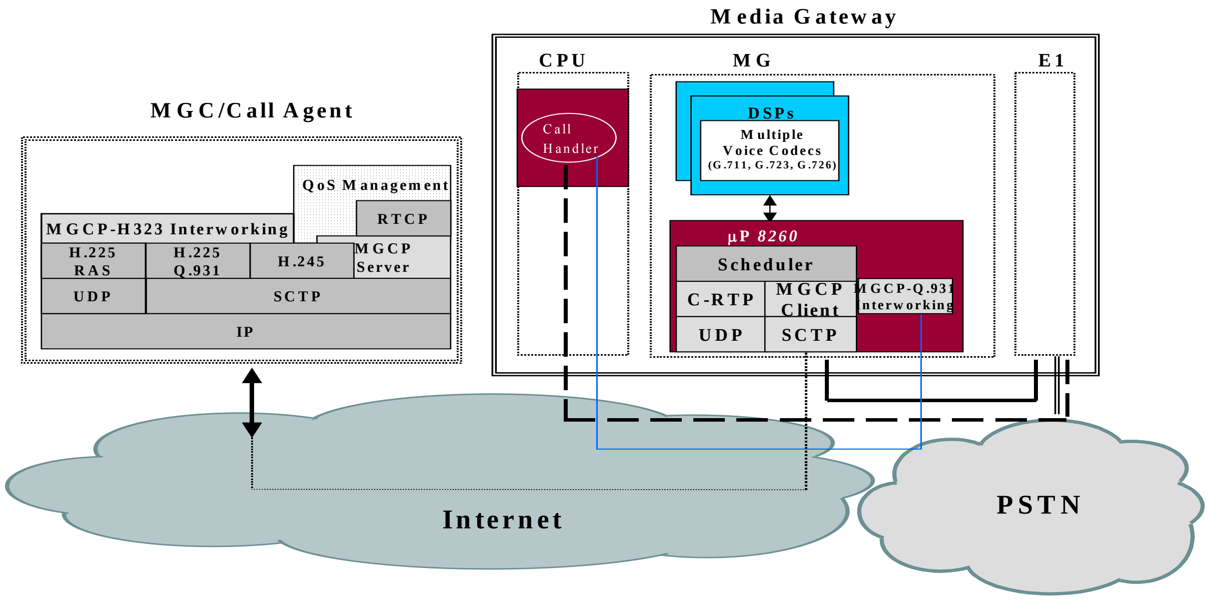

Fig.1: VoIP platform architecture overview

Agent where all the intelligence and the control of signalling protocol, the interworking module and the QoS management. The MG from its side is able to produce voice packets of different coding and transmit them using the CRTP instead of the simple RTP. It also provides minimum signalling functionality as it terminates the PSTN network using an E1 line.

The paper is organized as follows: Section 2 presents an overview of the architecture and of all the different components that compose the System. Section 3 presents a more detailed view of the MGCP and H.323 interworking, section 4 the use of SCTP and section 5 the use of C-RTP. Section 6 presents the QoS management, section 7 describes the Message Sequence Charts (MSCs) during end-point registration, call initiation, modification of the parameters during the call and call termination and finally section 8 presents the conclusions.

2 Platform overview

The proposed enhanced VoIP platform can participate in the following call scenarios: PSTN to PSTN, IP Phone to PSTN, PSTN to IP Phone and IP Phone to IP Phone. The protocol used to initiate VoIP calls at the client is the H.323 in parallel with the MGCP, which is responsible for the communication between the Call Agent and the Media Gateway. The whole system architecture is presented in fig.1. The Call Agent is responsible to handle the VoIP Calls and provides the necessary termination points. It includes the MGCP/H.323 Interworking Module which is responsible for the conversion of the H.225/Q.931 and H.245 messages to MGCP commands and vice versa. The QoS Management is the software module that deals with all the quality issues in the VoIP platform, such as packet loss or jittering that may lead to a change in the used voice codec.

The Media Gateway consists of the MGCP protocol in order to send and receive MGCP commands, the C-RTP module that provides the compression to the RTP packets by extracting the information from the headers as long as it is able to rebuild the packet upon the reception [11]. The SCTP protocol should also be supported and the Call Handler is able to gather all the ISDN/Q.931 messages and passes them to the Enhanced Call Agent.

The Media Gateway exhibits the following characteristics:

Compact PCI Backplane: The VoIP architecture is based on the Compact PCI (cPCI) backplane. cPCI is a superset of the desktop PCI, with enhanced electrical specs targeting the most demanding industrial and embedded applications [3]. Each cPCI bus is limited to eight slots for electrical loading reasons. This can be easily expanded with PCI-PCI bridge chips. One advantage of bridge chips is that each side of the bridge can be performing data transfers to boards on its side of the bridge simultaneously.

Thus, a dual PCI system separated by a bridge chip can be transferring data at a total of twice the usual 132 Mbytes/second rate. It is protocol compatible with the desktop PCI, but it has some additional signals (TDM Switching bus-H.110-integrated

Proceedings of the 4th WSEAS Int. Conf. on Information Security, Communications and Computers, Tenerife, Spain, December 16-18, 2005 (pp548-553)

within the available backplane connectors-J4) and different connector type, as illustrated in Fig 2.
The cPCI is intended as an industry-oriented bus for application in telecommunications, computer telephony, etc. Additionally because of its extremely high bandwidth, the cPCI bus is particularly well suited for many high-speed data communication applications such as routers, gateways, converters and switches [4]

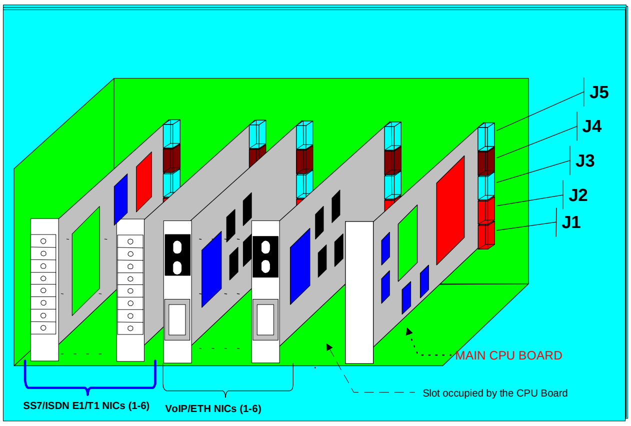

Fig.2: VoIP Gateway Switching Buses

Modular: The VoIP Gateway architecture uses a passive backplane with parallel bus and switching bus. The parallel bus is used for control, management and signalling data, while the switching bus for real time user data, as illustrated in Fig 3.
This way the parallel bus will be relief from data transfers not relative with control, management or signaling while the delay path will be reduced.

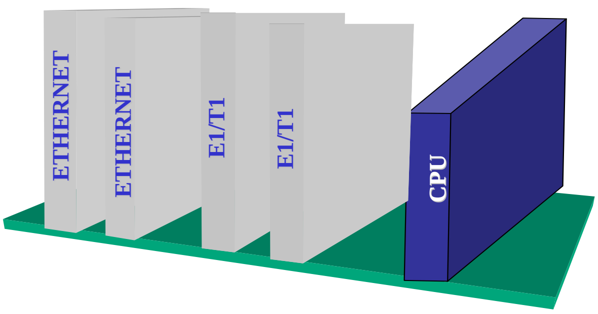

Fig. 3: cPCI VoIP Architecture

Distributed Architecture: The proposed architecture minimizes the processing expected for peripheral tasks by the main CPU. The main CPU performs the signaling translation and it is connected to the other peripheral boards using only the parallel bus. Multiple network interface cards (i.e. E1 Card, Ethernet Card) can be plugged in the VoIP gateway to support the needs and requirement which may differ from application to application, as illustrated in Fig 3.
In case where, more capacity is needed, extra network interface cards may be plugged in the cPCI backplane. When the maximum capacity of the backplane is reached, extra demand can be accommodated through the use of additional cPCI backplanes. It is the responsibility of the peripheral boards to deliver signaling data to the main CPU.
A dedicated switching backplane, the H.110 bus allows users' data exchange between the peripheral boards without any load for the CPU and the parallel bus.
Within this approach the following issues are considered with respect to the architecture of each individual peripheral board:

- The choice for use of an on-board processor is related with the efficient handling of the on board resources and data paths. For example each peripheral board should occupy the parallel bus for the minimum time with the optimum way i.e. using bulk transfers.
- Each peripheral board must be able to exchange user data using the dedicated switching bus. The type of the switching bus is of major importance for high performance and cost reduction.
- No matter the choice for the switching bus, there must be provision for future integration of additional switching resources.

Hot-Swapability: The power and signal pins on the CompactPCI connector are staged so as to allow the specification in the future to support hot swapping, a feature that is very important for fault tolerant systems and which is not possible on standard PCI. The cPCI backplane, allows on site replacement/additions. The cPCI is used to route control and signaling data between the network interface boards and the CPU.

Reliability: This corresponds to five nines availability (99,999). This can be accomplished by adopting passive backplane, using a very reliable Real-Time Operating System. Administration scheme based on PICMG specifications (System Management Specs) will monitor diverse parameters like temperature levels, voltage levels at the power supplies, fan rotation rates.

3. MGCP and H.323 Interworking

3.1 Use of MGCP

The most important part of the proposed platform is the H.323 to MGCP interworking module. This

module resides in the same physical entity with the Call Agent.
In fact it can be seen as an add-on to the Call Agent that also provides all the essential communication with the H.323 Gatekeeper that may be used. The architecture is presented in Fig 4.

MGCP Notification Request: This command is sent by the Call Agent to request the MG to send notifications upon the occurrence of specified events in an endpoint.
MGCP Create Connection: This command is sent by the Call Agent to inform the MG about the creation of a new voice channel.

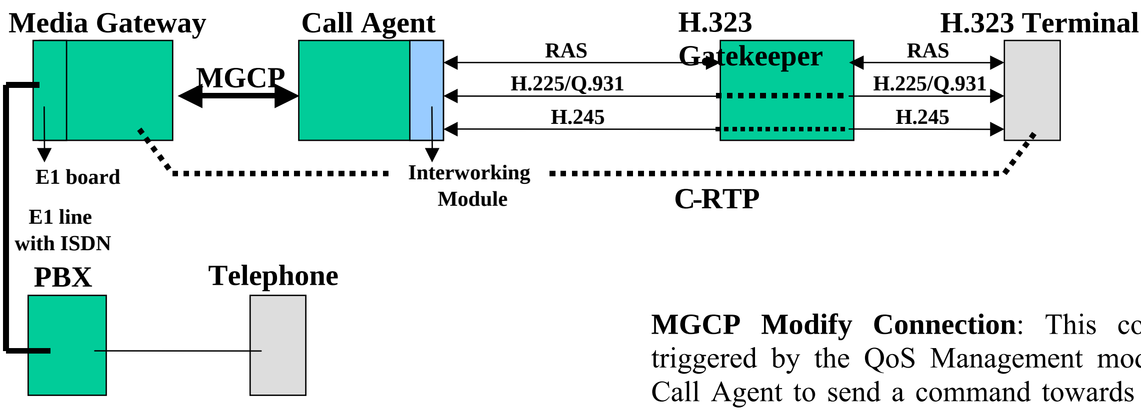

Figure 4: End to end VoIP platform architecture

The Media Gateway Control Protocol (MGCP) has been proposed by IETF to implement a decomposed and scalable VoIP architecture, where media transformation is separated from call control [5]. This separation is clearly presented in our platform where the MGCP is used for the communication between the Call Agent and the Media Gateway(s). Note that the MGCP is a client-server protocol, which means that each time one entity is sending, and the other is receiving and responds appropriately. Another important characteristic of our architecture is the Call Agent's capability to control more than one Media Gateways concurrently. The Interworking module registers the Call Agent to the H.323 Gatekeeper as if it is a normal H.323 entity, using the RAS protocol. In this way, the Call Agent can interact with the Gatekeeper and exchange messages. This is presented in the following section with the appropriate Messages Sequence Charts (MSCs). The next paragraph describes the use of the MGCP commands in the proposed platform and the essential transformation to H.323 messages (Q.931 and H.245).

3.1.1 MGCP Server

The MGCP Server resides at the Call Agent and is triggered by the Interworking Module in order to send the appropriate MGCP command. It handles the following commands:

MGCP Modify Connection: This command is triggered by the QoS Management module of the Call Agent to send a command towards the MG in order to instruct the DSPs to generate less bit by employing a higher compressed voice encoding scheme (e.g. switch from G.711 to G.727).
MGCP Endpoint Configuration: This command is used to inform the MG about the bearer/coding characteristics on the line side. In the proposed architecture, the EndpointConfiguration is also used to inform the MG about a H.225/Q.931 Alerting message from the IP side.
MGCP Delete Connection: As its name implies, this command is used by the Call Agent to terminate a connection and inform the MG to release the appropriate resources at the DSPs.

3.1.2 MGCP Client

The MGCP client resides at the MG. Upon the reception of a command that is sent by the Call Agent the MG sends back a response that carries several response parameters. The commands sent by the MG are the following:

MGCP Notify: This command is sent by the gateway to the Call Agent when the observed events (e.g. off-hook detection, collected digits and continuity tones) occur.

MGCP Delete Connection: The MG as the Call Agent has the capability to tear down a connection.

All the pre-referred commands are followed by the returned codes of each command.

4. Use of the SCTP as a transport protocol

In H.323, TCP is employed as to carry out the signaling messages from H.225/Q.931 and H.245 protocol [6], [7], [8]. Although TCP is a reliable protocol, it introduces large overheads, which deteriorates signaling performance. Besides, it is strictly oriented to sequencing delivery of the packets and continuous retransmissions, which are not always required in signaling transportation; at least, not in the way that TCP implements them. To give just an example, we know that different connections depend on different signaling streams and as a result the sequencing is essential only between the packets-messages of the same connections although all packets are carried through the same channel (TCP). Our scope is to use the meaning of the SCTP association and thus avoiding e.g. the head-of-blocking phenomenon, where one packet can block the transmission of all the forthcoming packets no matter whether they belong to the same signaling connection or not. Moreover to that, the employment of TCP can cause unnecessary delays with a great amount of acknowledgements, retransmission and duplicated packets. Furthermore, the stream-oriented nature of TCP can cause some kind of inconvenience when either trying to mark the messages' boundaries or when dealing with the transfer time. Moreover to that, TCP is quite vulnerable to denial of service attacks, and its sockets are not capable of providing highly available data transfer capability using multi-homing hosts. In a few words, TCP is not the most appropriate protocol for signaling transportation. For this purpose, IETF SigTran WG has standardized a new protocol (SCTP-Scream Control Transport Protocol) for the transmission of signaling information in IP-based networks [9]. In the proposed VoIP gateway, SCTP is employed to transmit H.323 signaling (H.225-Q.931 and H.245 messages).

The main advantages, as they are shown in real-time operation, are in the Network load and in the speed of Call Control (Call Setup, maintenance and Call Completion). The protocol stack of the proposed platform is shown in fig 5:

<table>
<tr><td>H225/RAS</td><td>H225/Q931</td><td>H245</td></tr>
<tr><td>UDP</td><td colspan="2">SCTP</td></tr>
<tr><td colspan="3">Network Layer</td></tr>
<tr><td colspan="3">Link Layer</td></tr>
<tr><td colspan="3">Physical</td></tr>
</table>

Fig. 5: VoIP Decomposed Architecture

5. Use of the Compressed RTP

RTP is the protocol introduced by IETF to support real-time multimedia transportation. In spite of its great advantages, RTP sometimes leads to great bandwidth consumption. To be more exact, RTP is carried over UDP that is carried over IP. As a result, the whole IP packet contains an IP header of 12 bytes, a UDP header of 8 bytes and the RTP header of 12 bytes. Thus, 32 bytes are spent for header information and the voice data can sometimes be less than 20 bytes (depends on the codec and on the frequency sampling used for the carried voice packet). To avoid this useless load, we introduce an algorithm for a Compressed RTP scheme based on [10]. The main idea is extracted by the fact that most of the RTP fields remain the same or has a constant value changing during the connection. So we can just carry the constant difference and reconstruct the initial fields at the receiver. In this way, we only send the first RTP packet in its original form and then only C-RTP packets are transmitted until a change in one of the header fields is detected. In that case the whole RTP packet is transmitted and another compression session begins. The gain from this algorithm is that the header is reduced in 2 bytes. This lightweight characteristic is very important in VoIP environments with heavy voice-packet load, where the Media Gateway sometimes manages up to 120 calls concurrently.

6. QoS Management

The other main characteristic of the enhanced Call Agent is the QoS Management. This is a software module that runs in the Call Agent and its main purpose is to monitor the overall network load and the quality of each call connection separately using the E-Model [11]. It is able to take decisions on the voice codec used in order to guarantee the call quality. Its implementation is based on a UDP server that continually listens to a pre-defined port where the RTCP packets of each connection arrive. When the RTCP packets arrive it grabs them and analyses their fields. Thus it extracts the values of the NTP Timestamp, RTP Timestamp, Interarrival Jitter and Packet loss fields. These values are used to estimate the packet loss, the jitter of each connection and the network load which may lead to a possible network congestion. Once the network load reaches at a certain level, the voice connections are instructed to employ a higher voice compression scheme, leading to the reduction of the generated bits. The level of degradation during congestion periods depends of the VoIP service category that each voice connection belongs and the mechanism used to handle the congestion.

Service Category	Encoding Scheme	Network Load
Best	G711 (64 Kbps)	
High	G727 (40 Kbps)	70%
Medium	G727 (32 Kbps)	80%
Low	G727 (16 Kbps)	90%
Poor	G728 (8 Kbps)	100%

Table 1: VoIP Service Categories

Without loss of generality, four different service categories are considered, as illustrated in Table 1.

The provided quality on each connection depends on the Voice Codec used.

A lower quality voice-encoding (e.g. switch from G.711 to G.727 at 40Kbps) scheme is employed when the estimated network load has reached at certain level. The last column presents the percentage network load as it is estimated from the entirety of the connections and the provided bandwidth. In the same time, the QoS Management module determines the expected VoIP QoS from each connection. This is accomplished by using the R-factor, as proposed by the ITU-T's E-Model [11]. The R factor ranges from 0 to 100, depends on the echo, the background noise, the signal loss, the codec impairments and is calculated from the following formula:

$$R=100-Is-Id-Ief+A$$

where Is is the signal-to-noise impairments associated with typical SCN paths, Id is the impairment associated with the delay of the path and Ief is an equipment impairment factor associated with losses within the gateway and A is the expectation factor covering those intangible quantities that are difficult to quantify.

For the sake of simplicity, according to the ITU-T G.107 recommendation, this equation can be simplified as follows [12]:

$$R = 94.2 - 0.024d + 0.11(d-177.3)*H(d-177.3)-Ie$$

Note that d is the one way delay time and H is either 0, when the parentheses is evaluated negative, or 1 for a non-negative value [9]. Once the network load drops at low level and the R-factor has fallen below the ordinary levels, the voice connections employ a higher quality voice encoding scheme to compensate for quality distortion during the congested periods.

7. Signaling Events in the VoIP Platform

This section focuses on presenting the operation of our VoIP platform for a real call from the PSTN to the IP network, based on the architecture of Figure 1. Such a Call demonstrates the full functionality of the platform and the interoperability of all the different entities. The physical entities that take part in this call are a PSTN telephone, a VoIP terminal, an external PC hosting the Call Agent and a compact PCI rack that has an E1, a VoIP and a CPU board.

The MSCs presented below, shows the messages exchange during the registration, call initiation, call modification and call termination. All the software entities presented in this paper, participate in this Call scenario.

References

[1] IEEE Communications Magazine, *Special Issue on Internet Telephony,* April 2000.

[2] Varshney, U. Snow, A. McGivern, M. and Christi Howard, "Voice Over IP', Communications of the ACM, June 2002

[3] *PICMG*, "Compact PCI Core Specification", PICMG 2.0 R3.0, March 2000

[4] PICMG, "Compact PCI Hot Swap", PICMG 2.0 R2.0

[5] *IETF RFC 2705,* 'Media Gateway Control Protocol', March 2003

[6] H. Liu and P. Mouchtaris, "VoIP Signaling: H.323 and Beyond", IEEE Communications Magazine, October 2000.

[7] *ITU-T Recommendation H.323,* "Packet-based multimedia communications systems", September 1999.

[8] *ITU-T Recommendation H.225.0,* "Call Signalling protocols and media stream packetization for packet-based multimedia communications systems", September 1999.

[9] *IETF RFC 2960,* "SCTP, Stream Control transmission Protocol" October 2004.

[10] *IETF RFC 3096,* "Requirements for robust IP/UDP/RTP header compression", July 2001.

[11] *ITU-T Recommendation G.107,* "The E-Model, a computation model for use in transmission planning", December 1998.

[12] R. Cole and J. Rosenbluth, "Voice over IP performance monitoring", ACM Computer Communications Review, Vol. 31, April 2001.