Evaluation of Media Server in SIP-Based Voice Conferencing

Fatemeh Samsami and Ahmad Akbari

Abstract—**VoIP and internet users are more and more interested in the multimedia communication services and this has made the academic researches in this domain to be risen dramatically. Media Servers play a key role in multiple communications and voice and video conferencing. Thus, implementation and evaluation of media servers has a great importance in the world of VoIP and Internet Telephony.**

In this paper we configured an open source media server called SEMS to run a voice conference. Then we evaluated our conference server using CPU load, packet loss and jitter. We carried out our tests using different speech Codecs like G.711, G.726, GSM, and iLBC and different number of participants. We also use an open source traffic generator called SIPp in order to generate SIP and RTP traffic simulating the participants of the conference.

We show when the number of participants in conference exceeds a certain value that is called capacity of server then the quality of communication decreases dramatically where an overload control is needed. The capacity of the conference media server is evaluated and obtained in this report for several Codecs and deployment situations.

Index Terms—**SEMS, SIP, CPU load, codecs.**

I. INTRODUCTION

According to the increasing interest to the world of Internet Telephony or VoIP, academic research in this field is raising dramatically. One of the main components in this world is media server which can provide different functionalities. The most important role of a media server can be considered as conferencing.

In conferencing framework so many participants can be attended and their media streams would be mixed together by media server and redistribute again to all participants. Media server can be evaluated from different point of views. In our research we used a free and open source media server which has been released by IPTEL organization in Germany. The name is SIP Express Media Server known as SEMS [1]. In the meantime we have a free and open source traffic generator called SIPp [2]. We evaluated SEMS on be half of Jitter, packet loss and CPU load versus increasing number of participants in conference.

There are a few researches in the field of media servers'

evaluations[2] [7]. M. Ahmed *et al.*, [7] in their research, which is very close to our paper evaluated Asterisk as an open source PBX and showed the capacity of this server for different Code.

This paper structure would be like this: the next part is about introducing two most important protocols in our platform which are SIP [8] and RTP [9]. Codecs are introduced in this part too. In Part III we explain conferencing and introduce SEMS as our free media server. Our evaluation and test results are described in part IV. Ultimately in part V we mention the conclusion and future works which can be considered as the point of research in future.

II. OVERVIEW

In this part we introduce the most important protocols in our experiments which are SIP and RTP. Additionally we have a brief overview of media server, conferencing. Then we explain Codecs and show a comparison of different Codec.

A. Session Initiation Protocol

Session Initiation Protocol is a signaling protocol used to initiate, modify and terminate the multimedia sessions between two or more parties. The SIP invite messages can be used to establish a call session over the IP networks. These messages carry descriptions which different parties can adopt on various media types as shown in Fig. 1.
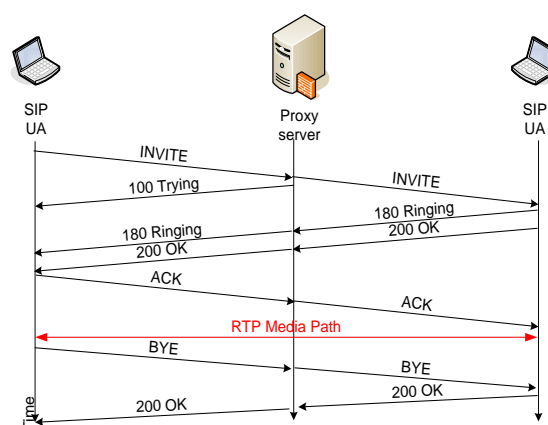

Fig. 1. How SIP communicate between parties [10].

SIP has different messages to establish and terminate a call like INVITE, Trying, Ringing, 200 OK, ACK and BYE.

SIP was accepted as a 3GPP signaling protocol and permanent element of the IP Multimedia Subsystem (IMS) architecture for IP-based streaming multimedia services in mobile systems [11], [12].

B. Real-Time Transport Protocol

RTP is a Real Time Streaming Protocol and provides end-to-end network transport functions suitable for applications transmitting real-time data, such as audio, video, simulation data, over multicast or unicast network services. RTP does not address resource reservation and does not guarantee quality-of-service for real-time services. The data transport is augmented by a control protocol (RTCP) to allow monitoring of the data delivery in a manner scalable to large multicast networks, and to provide minimal control and identification functionality. RTP and RTCP are designed to be independent of the underlying transport and network layers. The protocol supports the use of RTP-level translators and mixers. The RTP header fields are shown in Fig. 2.

Fig. 2. RTP header fields [13].

C. Codec

A Codec [14]-[18] is an analog-to-digital (A/D) and digital-to-analog (D/A) converter. Poor voice quality can actually be a problem. G.711, G.726-32, GSM, iLBC are evaluated in our experiments as they are supported by our Media Server. A brief comparison of these Codecs is listed in Table I.

TABLE I: COMPARISON OF DIFFERENT CODECS

Codecs	G.711	G.726	GSM	iLBC
Bit Rate (kbps)	64	16, 24, 32, 40	13	13.3, 15.2
Complexity (MIPS)	0.2	8.7	1.4	15
Frame Size (ms)	20	20	20	30, 20
Sample rate (KHz)	8	8	8	8
Standard	ITU-T	ITU-T	ETSI	ITU-T

III. MEDIA SERVERS AND VOICE CONFERENCING

The necessity of using the multimedia communication services for internet users is increasing day by day. Since the media servers play the key role in the world of multimedia and VoIP services, thus the evaluation of media servers may be extremely important.

A. SIP Express Media Servers (SEMS)

This media processing functions are the core responsibility of a media server. SEMS is an Open Source, free and extensible media server that functions as a Voice mail, ISDN gateway, Conference, IVR and announcement server.

SEMS is also developed by [IPTEL], easy to implement and works fine with SER. SEMS is a free media server which helps you adding voice services to your VoIP system. The current version provides voice mail, conference and announcement. SEMS can be installed on a separate machine or on the same machine where the proxy server is running. It requires SER instance for its functioning. Together with SER (Sip Express Router), it turns every Linux box into a full-featured IP telephony system. It uses a centralized conferencing model.

B. Conferencing

Over the past few years, IETF SIPPING working group has been working on standards for the conference management and control using IETF session initiation protocol (SIP). In Fig. 3, we show the general conferencing architecture standardized by IETF SIPPING working group. As you see the conferencing architecture consists of a conference server and participants.
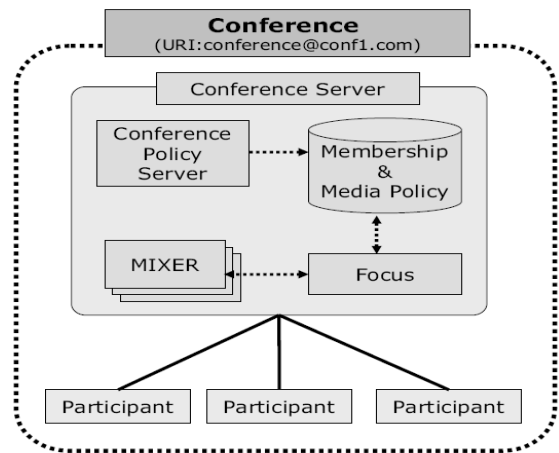

Fig. 3. The general conferencing architecture [19].

As said, to insert images in *Word,* position the cursor at the insertion point and either use Insert | Picture | From File or copy the image to the Windows clipboard and then Edit | Paste Special | Picture (with "Float over text" unchecked).

The authors of the accepted manuscripts will be given a copyright form and the form should accompany your final submission.

The focus is a SIP user agent which is addressed by a conference URI and identifies a conference. It maintains a SIP signaling relationship with each participant in the conference and is responsible for ensuring, in some way, that each participant receives the media that make up the conference. It also implements conference policies.

A focus is responsible for the management of the conference using SIP signalling protocols. It handles the requests from participants by referring to the conferencing polices which are stored in the membership and media policy databases. The media policy databases and the membership are managed by the policy server. The conference policy contains the rules that guide the decision-making process of the focus for the management of various conference requests from the participants.

A mixer receives a set of media streams of the same type, and depending upon the policies defined by focus, combines their media in a type-specific manner and redistributing the result to each participant. Mixer is responsible for managing

the multimedia streams, and generation of output streams which can be distributed to participants. A mixer can be located either in the focus or in the participant's user agent. In both cases, a mixer is controlled by the focus.

Conferencing with centralized servers having direct connection to the participants are easier to handle for the end systems and simplify keeping track of the participants and state machine for the SIP Conference

The Media server plays the key role, it transcodes, mixes and redistributes appropriate media stream to the participants based upon the configurations made in the manage server [19]-[22].

IV. PROPOSED METHOD TO EVALUATE MEDIA SERVER

In one system we install SEMS, while using its conferencing application. On the other side we configure SIPp according to these four selected Codecs; it works same as a soft phone but has the ability of sending so many participants toward conference server simultaneously.

In order to trace SIP and RTP packets in network card we installed Wireshark [23] as a powerful network protocol analyzer developed by an international team of networking experts. Meanwhile Wireshark is used to calculate packet loss and jitter. Our Test Bed including the name and version of operating system, media server and traffic generator is mentioned in Table II.

TABLE II: TEST BED

Name	Description
System	Pentium IV- 1.86GHZ / 3096MB RAM
Operating System	Fedora 9.0-Kernel 2.6.25
Traffic Generator	SIPp 3.1 2008
Media Server	SEMS V 1.1.1

Topology of this test bed is shown in "Fig. 4". The position of media server and SIPp is clear then. All our tests are fulfilled in the same condition and after warm-up process. It should be noticed that call rate in all experiment is 50 calls per second and the duration of RTP transmission to the server is 30 second. Note that just one conference takes place.

On the other hand we considers to states: with VAD[1] and without VAD. VAD also known as speech activity detection or speech detection is used to detect the presence of speech in an audio signal.

VAD has an important role as a pre-processing stage in numerous audio processing applications. For example, in VoIP[2] and mobile telephony applications, VAD can reduce bandwidth usage and network traffic by transmitting audio packets only if speech is detected.

The first test is CPU load or Stress test for different Codecs with VAD, in which we evaluate the CPU load while increasing the number of participants. The diagram in "Fig. 5" shows that the G.711 Codec has the best performance in our test bed. We evaluated this experiment in our laboratory with

[1] Voice Activity Detection
[2] Voice over IP

LAN Network and as the diagram shows the role of Codec's complexity are clearer than other parameters same as bit rates in these conditions.
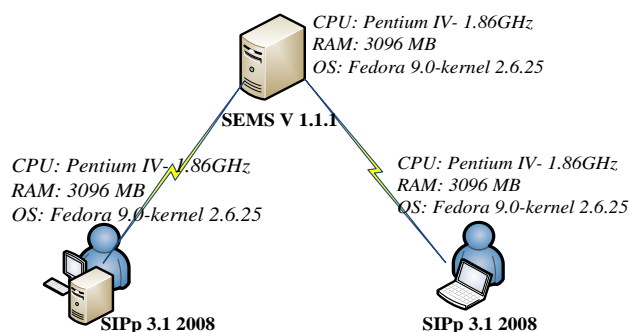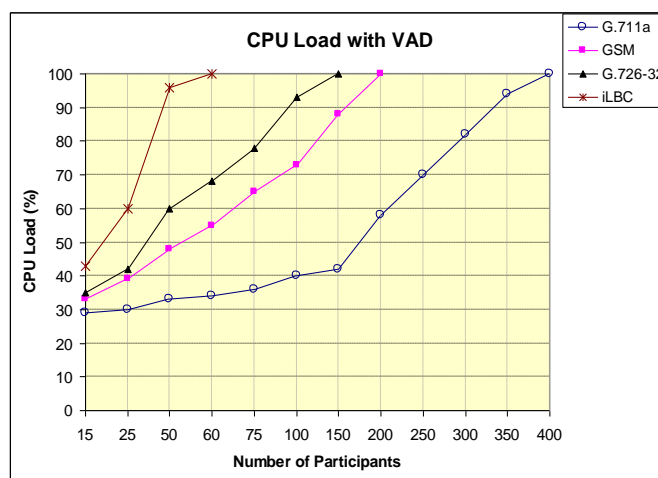

Fig. 4. Topology of test bed.

Fig. 5. CPU Load with VAD versus increasing Number of Participants

If we consider CPU Load of 80% acceptable for server, then we can have this number of participants in conference for different Codecs:

- iLBC: 40 users
- G.726-32: 75 users
- GSM: 120 users
- G.711-a: 300 users

Each result is deduced from statistic mean of 10 tests and after omitting the computational errors and out of range cases.

It should be noticed that CPU load can be changed due to the operating system type, the hardware like CPU, cache and even the type of time allocation algorithm by Kernel.
In "Fig. 6" you see the same experiment but for the state of No_VAD. If we consider CPU Load of 80% acceptable for server, then we may have following number of participants in conference for different Codecs:

- iLBC: 32 users
- G.726-32: 50 users
- GSM: 68 users
- G.711-a: 170 users

The infraction of VAD in media servers is obvious then. For G.711a with VAD we may have 300 participants in conference but without VAD we have just 170 participants.
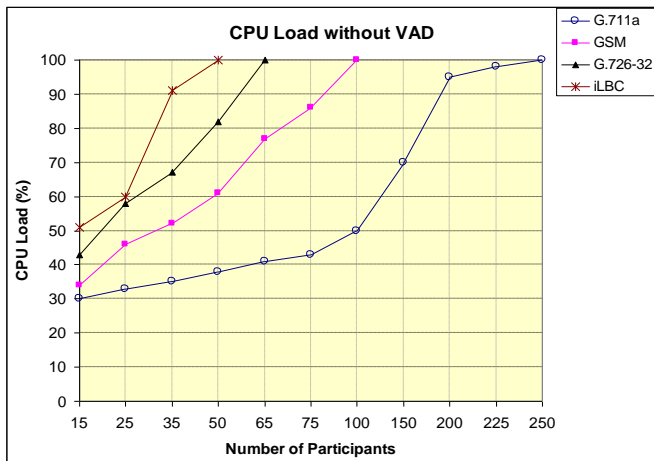
Fig. 6. CPU Load without VAD versus increasing Number of Participants.

The behavior of packet loss and jitter in a conference is evaluated too. Generally, packet loss could be the consequence of signal degradation in the network platform, saturation of communication links, incorrect hardware and even the routing policies in the router. But as we have LAN network, thus the packet loss would be the consequence of media server overload and missing of packets in SEMS buffer.

In "Fig. 7" we consider packet loss of 3% acceptable for our test, and then we may have following number of participants in conference for different Codecs:

- iLBC: 55 users
- G.726-32: 115 users
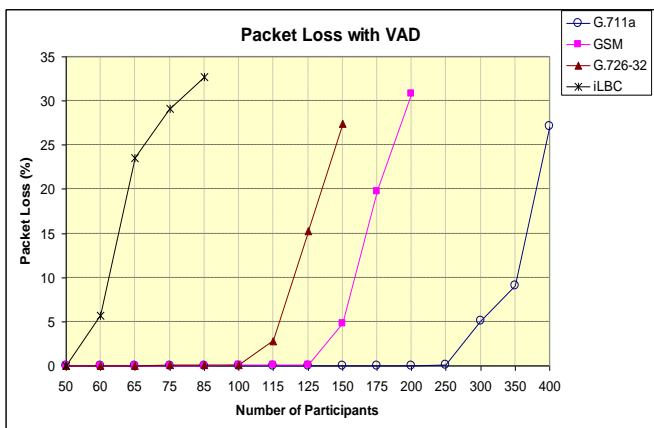- GSM: 140 users
- G.711-a: 280 users

Fig. 7. Packet Loss with VAD versus increasing Number of Participants.

In "Fig. 8" you see the simultaneous diagram of CPU Load, packet loss and jitter for G.711a with VAD state.

As you see all three diagrams start to increase dramatically nearly form the point of 300 participants in the conference.

In general, jitter is usually used as a measure of the variability over time of the packet latency across a network and can be happened due to the network traffic, routing change even devices wrong set-up to face incoming queue traffic. The main reason of increasing jitter in our test bed is increment of number of participants more than server capacity during composition of media streams and redistribution them again to the participants.
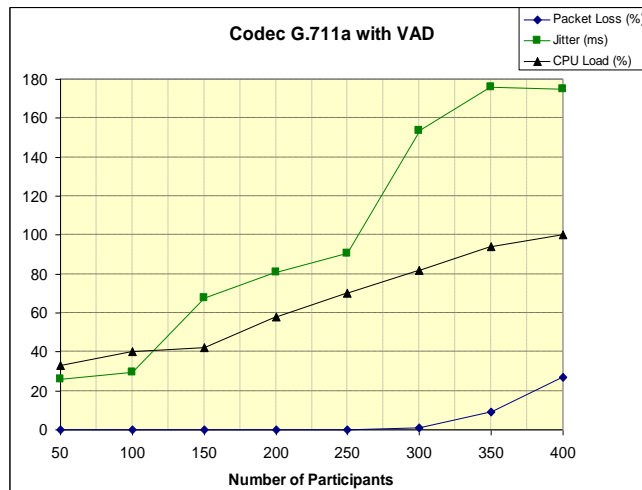

Fig. 8. CPU Load, Packet Loss and Jitter with VAD versus increasing Number of Participants.

V. CONCLUSIONS

In this paper we evaluated a media server in different conditions and deduced while increasing the number of participants in a conference from a specific point server will be saturated and not be able to accept any new user. The evaluation of RTP load over a media server can present a conception based on QoS and higher lever than signaling load. For signaling tests all evaluations are exactly logical according to successful or unsuccessful calls and the number of calls, but RTP load which is in fact the main core of calls in conference can show the quality rather than quantity. The main goal in real-time communications is higher quality. We deduced in our research after saturation point of media server although call establishment is possible yet but quality decrement will prevent us to have a proper communication and actually there is not a good quality communication. By defining new scenarios and making load balancing policies and before overload of server we can distribute loads over other servers steadily. Defining such loading points can be extremely important for enterprises and these service providers.

On the other hand as our test experiments are fulfilled over a LAN network in laboratory, therefore we don't have a real network packet loss and jitter, and the impact of Codec's complexity is more than other parameters like bit rate. Since iLBC has the most complexity of 15 MIPS, in LAN conditions the CPU Load would be saturated much sooner. Also the packet loss and Jitter would have the worst condition in comparative with other Codecs.

Eventually the role of VAD is clear then and the most importance part of the result is finding the capacity of SEMS for different Codecs.

For future work we are planning to have VAD in server side too. As in this research just client side had the ability of VAD. Also we will evaluate the other conferencing models in future, since in this paper we had just the centralized conferencing model for SEMS.

REFERENCES

[1] SIP Express Media Server. [Online]. Available: http://www.iptel.org/sems

[2] SIP. [Online]. Available: http://sipp.sourceforge.net/doc3.0/reference.html

[3] H. Wook, S. G. Kang, and D. Kim, "Performance Enhancement of SIP proxy server by using lhash for matching transaction," in *Proc. The 9th International Conference on Advanced Communication Technology*, vol. 2, pp. 1290-1293, 2007.

[4] M. Femminella, R. Francescangeli, F. Giacinti, and E. Maccherani, "Design, Implementation, and performance evaluation of an advanced SIP-based call control for VoIP services," in *Proc. IEEE International Conference on Communications*, pp. 1-5, 2009.

[5] F. Samsami, "Evaluation of SIP-based voice conferencing," M. Eng. thesis, Iran University of Science and Technology, Tehran, Iran, Feb. 2010.

[6] A. H. Ashouri, F. Samsami, A. Akbari, "E-Learning Media Server Evaluation and its architecture modeling with signaling load tests," ICeLT, IUST, Tehran, Iran, Dec 2009.

[7] M. Ahmed and A. M. Mansor, "CPU dimensioning on performance of Asterisk VoIP PBX," 11th communications and networking simulation symposium, Ottawa, Canada, pp. 139-146, Apr 2008.

[8] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler, "SIP: Session Initiation Protocol," RFC 3261, June 2002.

[9] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications", RFC 3550, July 2003.

[10] M. Jahanbakhsh, V. Azhari, A. Akbari, and M. Homayouni, "Evaluation of different SIP Signaling network topologies using measurement of call setup quality parameters," *ICT Scientific and Research Lett.,* vol. 1, Amirkabir University of Technology, Iran, Dec. 2009.

[11] F. Yao and L. Zhang, "OpenIMS and Interoperability with Asterisk/Sip Express VoIP Enterprise Solutions," MSc Thesis, Agder University college, Norway, May 2007.

[12] E. J. Ramos, "Analysing the Media Control interfaces and Mobile Media Gateway for the IP Multimedia Subsystem (IMS)," Helsinki University of Technology February 2008.

[13] R. M. Perea, *Internet Multimedia Communications Using SIP,* M. Kaufmann, Ed. Burlington, MA 01803, USA: Elsevier, 2008.

[14] R. Salami, "Low Bit Rate Speech Coding for Multimedia and Wireless Communications," International Workshop on Circuits, Systems and Signal Processing for Communications, Tampere, Finland, April 2006.

[15] H. Schulzrinne *et al*., "RTP Profile for Audio and Video Conferences with Minimal Control," RFC 3551, July 2003.

[16] G. Vaudreuil *et al*., "Toll Quality Voice - 32 Kbit/s ADPCM," RFC 2422, September 1998.

[17] H. Haverinen *et al*., "Extensible Authentication Protocol Method for GSM," RFC 4186, January 2006.

[18] S. Andersen *et al*., "Internet Low Bit Rate Codec," RFC 3951, December 2004.

[19] Y. H. Cho *et al*., "Distributed Management Architecture for Multimedia Conferencing Using SIP," in *Proc. IEEE First International Conference on Distributed Frameworks for Multimedia Applications*, pp. 98-105, 2005.

[20] J. Rosenberg, "A Framework for Conferencing with the Session Initiation Protocol (SIP)," RFC 4353, February 2006.

[21] O. Levin, "High-Level Requirements for Tightly Coupled SIP Conferencing," RFC 4245, November 2005.

[22] R. Even and N. Ismail, "Conferencing Scenarios," RFC 4597, August 2006.

[23] Wireshark. [Online]. Available: http://www.wireshark.org/

Fatemeh Samsami was born in Yazd province of Iran which is the second historical city in the world in 1979. The author has been under graduated in Electronic Engineering from Yazd university of Iran in 2004 and graduated in Information and Communication Technology (ICT) from Iran University of Science and Technology located in Tehran capital city of Iran in 2010.

She has worked in Jahan Electronic Company (JEC) located in Yazd as a Foreign Commercial Manager for nearly 2 years. It was an excellent experience in life career since she had connection with so many important and esteemed companies from Iran and all over the world. At the moment she teaches Mathematics, Telecommunications and English at Azad University of Mehriz.