# A Practical Analysis of Asterisk SIP Server Performance

Razvan Rughinis Politehnica University of Bucharest Romania Email: razvan.rughinis@cs.pub.ro

### Abstract

This paper presents an analysis of the Asterisk server performance as a SIP server, as well as of bandwidth consumption in multiple scenarios. Server load has been tested using a certain type of codec, which requires a voice quality close to classic telephony (PSTN). Therefore, one can infer minimum VoIP requirements for an Asterisk SIP server by analyzing the maximum number of possible calls and the maximum bandwidth.

Keywords: IP telephony, Asterisk, SIP, SIPp, IAX.

## **1. Introduction**

With the development of VoIP telephony, the Asterisk servers have become common in personal and organizational environments as well. Besides relying on the benefits of open source technology, Asterisk servers have also proven to be resilient and to function well over heterogeneous architectures [1]. Still, any fast expanding Internet technology becomes vulnerable to security threats [2, 3] and to increasing demands for quality of service. One of the main limits for addressing these concerns derives from the relatively high bandwidth consumption of VoIP, since bandwidth requirements as well as processor load increase significantly when measures providing Confidentiality, for Integrity and Availability are put into place [4].

Given the importance of low resource consumption for future improvements in the use of Asterisk servers, this paper attempts a performance evaluation of an Asterisk server, focusing on bandwidth consumption and processor load under several utilization scenarios.

### 2. Performance evaluation

We have analyzed the performance of an Asterisk server in multiple scenarios of use as an SIP server. The laboratory setup consisted of the following equipments: two PCs Pentium(R) D CPU 2.8GHz 1GB RAM, two NICs 1Gbps, and one switch 3560g. The following pieces of software were deployed: Asterisk PBX v1.4.18, Debian 4.0 / kernel 2.6.18-6-

Cristian Iconaru RoEduNet Romania Email: cristian.iconaru@roedu.net

686, Microsoft Windows XP [Version 5.1.2600], SIPp v3.1, X-Lite v3.0.

#### 2.1. General bandwidth constraints

ITU-T G.711, referred to as PCM, is the most frequently used codec in present-day telephony system. There are two variants of this codification method:  $\mu$ -law, used in the US and in Japan, and a-law in the other countries. Each of these methods transmits 8000 samples per second – which means that a total of 64000 bits may be transmitted per second.

Therefore, in order to digitally transmit voice, the G.711 codec uses 64 Kbps of the available bandwidth. In real usage there is an overhead load besides voice traffic, which leads to a real bandwidth usage of 87.2 Kbps. This bandwidth is used in both directions. This means that a minimum 128 Kbps canal is required for this codec to function normally.

# **2.2. External connection bandwidth constraints for the maximum call rate**

An Asterisk server similar to the one used in the laboratory configuration may support up to 250 simultaneous calls, equivalent to 500 simultaneous channels.

Call-rate(length) Port Tot		reen otal-calls 4500	Remote-h	Change Screen ost :5060(UDP)
0 new calls during 1.000 s per 750 calls (linit 750) 0 Running, 3033 Paused, 250 Wo 0 dead call msg (discarded) 3 open sockets	ken up	. ms schedu Peak was 75 I out-of-ca	0 calls, a	fter 3 s
INUITE	3750 3750 3750 3750 3750	Retrans 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Timeout 0 0 0 0 0 0 0 0 0	Unexpected-Msg 0 0 0 0 0 0 0 0 0 0 0 0
[+¦- *¦/]: Adjust rate	[q]: Sa	ft exit	[p]: Pa	use traffic

Start Time Last Reset Time	Statistics Screen   2008-06-18 15:51:26:812   2008-06-18 15:51:49:853	[1-9]: Change Screen 1213793486.812500 1213793509.853500	Call initiation IVR activation
Current Time	2008-06-18 15:51:50:849	1213793510.849500	CPU History
Counter Name	Periodic value	Cumulative value	
Elapsed Time Call Rate	00:00:00:996 69.277 cps	00:00:24:037 93.606 cps	
Incoming call created OutGoing call created Total Call created Current Call	0 69 750	2250 2250 2250	
Successful call Failed call	33 Ø	1500 0	
	00:00:10:002   00:00:10:003   rate [q]: Soft exit	:   00:00:10:005   00:00:10:013 - [p]: Pause traffic	

Figure 1. Details regarding the Asterisk server

A single call creates two channels, but one may use more, according to the Dial Plan logic. The above discussed scenario requires a 43,600 Kbps bandwidth, since each channel requires a 87.2 Kbps bandwidth.

If one uses a T3 connection (44,736 Kbps) this means that around 256 calls can be initiated, given the availability of 513 channels. Therefore, one Asterisk server is enough to saturate such a connection.

A significant processor load is noticeable in the IVR activation periods, for 250 calls per second, for the given Dial Plan.

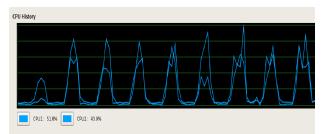


Figure 2. Processor load for 250 calls/second

The load is lower when calls are effectively initiated by the server.

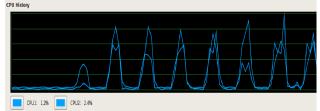


Figure 3. Processor load for 250 calls/second if calls are initiated by the server

The spikes in the figures below illustrate IVR activation (voice transmission), while lower segments represent the de facto call initiation.

Figure 4. Spikes in processor load for 250 calls/second

Tests indicate a large number of large size SIP packets and an overall traffic of 16 Mbps.

C Ethereal: Summ	iary			
File Name: Length: Format: Packet size limit:	58254850	bytes pdump, Etherea	AL5~1\Temp\etherXXXX50NVCU I, etc.)	
Time First packet: Last packet: Elapsed:		8 15:51:24 8 15:53:52		
Capture Interface: Dropped packets: Capture filter:	i packets: 0			
Display Display filter: Marked packets:	sip 0		,,	
Traffic		Captured	Displayed	
Between first and l	ast packet	147,578 sec	145,172 sec	
Packets		162813	72942	
Avg. packets/sec		1103,237	502,452	
Avg. packet size			500,190 bytes	
Bytes		55649818	36484873	
Avg. bytes/sec Avg. MBit/sec		377088,599 3,017	251321,744 2,011	
		Clo	)se	

Figure 5. Test information for 250 calls / second

Figure 6 illustrates results for real voice traffic (RTP). One can see a large number of voice packets, with an overall traffic of around 8 Mbps.

🚽 Echereal. Summ				
File Name: Length: Format: Packet size limit:	58254850	bytes pdump, Etherea	u.S~1\Temp\etherXXXXS0NVCU I, etc.)	
Time First packet: Last packet: Elapsed:		18 15:51:24 18 15:53:52		
Capture Interface: Intel(R) PRO/1000 PL Network Connection (Microsoft's Packet Scheduler) Dropped packets: 0 Capture filter: not tcp port 3389				
Display Display filter: Marked packets:	rtp 0			
Traffic		Captured	Displayed	
Between first and l	ast packet	147,578 sec	135,169 sec	
Packets		162813	89035	
Avg. packets/sec		1103,237	658,696	
Avg. packet size		341,000 bytes 55649818	214,000 bytes 19053490	
Bytes		377088.599	140960.926	
Aug. butec/cec				
Avg. bytes/sec Avg. MBit/sec		3,017	1,128	

Figure 6. Test results for real voice traffic (RTP) for 250 packets/second

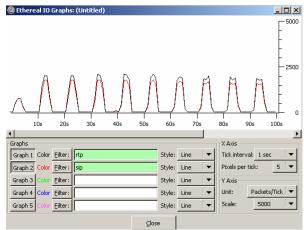


Figure 7. Spikes for IVR activation and call initiation

Figure 7 illustrates spikes for IVR activation and for call initiation, analogously to Figure 4 for processor load.

In order to obtain 250 calls per second we used two Pentium(R) systems D CPU 2.8GHz 1GB RAM, NIC 1Gbps. The Asterisk server used only the G.711 codec without echo supprimation.

The Asterisk configuration is detailed below:

Extensions.conf [default] exten => 2005, 1, Ringing;//ring exten => 2005,2,Wait(10); //wait 10 sec exten => 2005,3,Playback(vm-nobodyavail); //after 10 sec IVR exten => 2005,4,Playback(vm-goodbye); exten  $\Rightarrow$  2005,5,Hangup; exten => 2006, 1, Dial(SIP/user1, 10);exten => 2006,2,Playback(vm-nobodyavail); exten => 2006,3,Playback(vm-goodbye); exten  $\Rightarrow$  2006,4,Hangup; exten => 2007, 1, Dial(SIP/user2, 10);exten => 2007,2,Playback(vm-nobodyavail); exten => 2007,3,Playback(vm-goodbye); exten  $\Rightarrow$  2007,4,Hangup; sip.conf [user1] username=user1 secret=1234 callerid=User1 type=friend context=default host=dynamic disallow=all allow=alaw [user2] username=user2

secret=1234 callerid=User2 type=friend context=default host=dynamic disallow=all allow=alaw

### 2.3. Performance for lower call rates

For relatively small numbers of simultaneous calls, the situation is completely different. For 10 calls / second the processor load is low, up to 3%.

Call-rate(length) Port Tota		reen otal-calls 270	Remote-1	Change Screen lost :5060(UDP)	
0 new calls during 1.000 s per 30 calls (limit 30)	F	1 ms scheduler resolution Peak was 30 calls, after 3 s			
0 Running, 125 Paused, 10 Woker 0 dead call msg (discarded) 3 open sockets		0 out-of-call msg (discarded)			
INUITE>	Messages 270	Retrans Ø	Timeout Ø	Unexpected-Msg	
100 <	270	Ø	0	р	
180 <	270	Ø	0	Ø	
183 <	й И	Ø	õ	õ	
200 < E-RTD1		õ	й	й	
ACK>	240	ø		-	
Pause [ Øms]	240			Ø	
BYE>	240	Ø	0	-	
200 <	240	Ø	Ø	0	
[+ - * /]: Adjust rate	[q]: So	ft exit	[p]: Pa	uuse traffic	

Figure 8. Server information for lower call rates

The two spikes are generated by recording the two software calls.

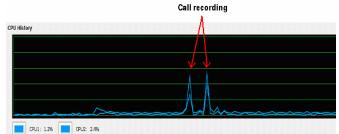


Figure 9. Spikes in processor load for lower call rates

The number of RTP packets is low, and they have the standard 214 byte size.

e	Ethereal: Summ	ary			
	File Name: Length: Format: Packet size limit:	1251894 b	pytes pdump, Etherea	s\user\Desktop\Test1_10cps\t1_10cps  , etc.)	
	Time First packet: Last packet: Elapsed:		0 14:06:13 0 14:07:35		
	Capture Inter(R) PRO/1000 PL Network Connection (Microsoft's Packet Schedule Dropped packets: unknown Capture Filter: not top port 3389				
	Display Display filter: Marked packets:	rtp 0			
	Traffic		Captured	Displayed	
	Between first and k Packets Avg. packets/sec Avg. packet size Bytes Avg. bytes/sec Avg. MBit/sec	ast packet	81,391 sec 4015 49,330 295,000 bytes 1187630 14591,584 0,117	63,562 sec 210 3,304 214,000 bytes 44940 707,028 0,006	
			1		

Figure 10. Test information for lower call rates

The system load does not increase significantly for a higher rate of calls per second, such as 50 or 100 calls/second. For 50 calls per second processor load reaches a maximum of 5%.

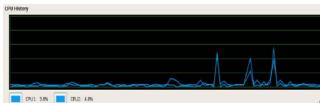


Figure 11. Processor load for 50 calls/second

For 100 calls / second the processor load reaches at most 7%.

CPU History	
	20
Λ Λ	10 -
and amarka amarka	0
GUL 126 002 196	

Figure 12. Processor load for 100 calls/second

# **2.4.** Performance for excessive call rates

If the limit of 250 calls/sec is exceeded, such as attempting to initiate 300 calls/second, certain SIP requests cannot be supported and this leads to retransmissions, while the calls are lost. For call rates of 300 calls/second and above, retransmissions are initiated after 12 seconds.

Call-rate(length) Por 300.0(0 ms)/1.000s 506			lls Remot		
108 new calls during 1. 900 calls (limit 900) 0 Running, 1156 Paused,		Peak was		olution , after 3	
3 dead call msg (discar 3 open sockets			call msg	(discarde	d>
INUITE>		sages Retrans	Timeou Ø	ıt Unexp	ected-Msg
100 <	103		ă	Ø	
180 <	1 / 1 / 2		0 0 0	ø	
183 <	0	ō	ø	0	
200 <	E-RTD1 593	Ø	Ø	Ø	
ACK>	593	0			
Pause [ Øms]	593			Ø	
BYE>	593	299	Ø		
200 <	256	0	Ø	Ø	
[+ - * /]: Adjust	rate [	[]: Soft exit	[p]:	Pause tr	affic

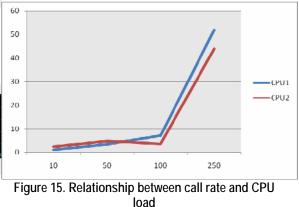
Figure 13. Server information for call rates of 300 calls/second and above

Start Time	2008-06-18	17:12:05:765	1213798325.765625	
Last Reset Time	2008-06-18	17:12:41:806	1213798361.806625	
Current Time	2008-06-18	17:12:42:803	1213798362.803625	
Counter Name	Periodic val	lue	Cumulative value	
Elapsed Time Call Rate	00:00:00:997 42.126 cps		00:00:37:038 80.431 cps	E
Incoming call created OutGoing call created	0 42		0 2979	10
Total Call created Current Call	900		2979	
Successful call Failed call	42 Ø		2073 6	ľ
Response Time 1 Call Length [+ - * /]: Adjust ;	1 00:00:10:448		00:00:11:529   00:00:12:347   [p]: Pause traffic	

Figure 14. Call failure for high call rates

# 3. Conclusions

There is a non-linear relationship between the call rate and the processor load for the Asterisk server in the given configuration.



Therefore, it is possible that security policies that require significant resources may be accommodated by a relatively slight decline in the call rate, avoiding the maximum levels.

### 4. References

[1] R. Rughiniş, and I. Herişanu, "Designing and Implementing an IP Telephony Architecture over a Heterogenous Infrastructure", Proceedings of the 15th International Conference on Control Systems and Computer Science (CSCS15), May 2005

[2] J. Albers et.al., An Analysis of Security Threats and Tools in SIP-Based VoIP Systems, A capstone paper submitted as partial fulfillment of the requirements of the degree of Masters of Interdisciplinary Telecommunications at the University of Colorado, Boulder, April 29, 2005, available at URL: <u>http://www.colorado.edu/policylab/Papers/voip\_security\_ capstone\_2005.pdf</u>.

V. Ionescu, F. Smaranda and E. Sofron, "VOIP: rent Security Issues", Proceedings of the workshop vices and Software Architectures, Infrastructures and ineering for enhancing EU Citizen's Quality of Life, 718, 2007, Bucharest, Romania, pp.91-96

B. Clayton, Securing Media Streams in an Asteriskd Environment and Evaluating the Resulting

ormance Cost, MS Thesis, Rhodes University, 2007, available at URL <u>http://eprints.ru.ac.za/851/01/Clayton-TR07-08.pdf</u>