TELKOMNIKA, Vol.13, No.2, June 2015, pp. 739 ~ 744
ISSN: 1693-6930, accredited A by DIKTI, Decree No: 58/DIKTI/Kep/2013
DOI: 10.12928/TELKOMNIKA.v13i2.1741

High Available VoIP Server Failover Mechanism in Wide Area Network

Hery Dian Septama*[1], Ardian Ulvan[2], Robert Bestak[3], Jiri Hlavacek[4]
[1,2] Department of Electrical Engineering, University of Lampung,
Jln. Soemantri Bojonegoro No. 1 Bandar Lampung 35145, Indonesia
[3,4] Department of Telecommunication Engineering, Czech Technical University,
Technicka 2, 166 27 Prague, Czech Republic
*Corresponding author, e-mail: hery@eng.unila.ac.id[1], ardian.ulvan@eng.unila.ac.id[2],
robert.bestak@fel.cvut.cz[3], hlavaji1@fel.cvut.cz[4]

Abstract

Voice over Internet Protocol (VoIP) is Internet Protocol (IP) based technology for the next generation network. VoIP has been develoved to overcome future telephony demand. However, some issues i.e. how to provide the services availability and reliability equally to circuit based telephony, need to overcome before VoIP replaced the circuit based telephony. Live migration in virtualization environment is used to provide high availability service, but the failover mechanism over Wide Area Network (WAN) need to be improved. In this paper, the ability of virtualization failover mechanism over WAN and voice quality of VoIP service in High Availability system is studied and examined. Both objective analysis by using quality of services (QoS) attributes is conducted as well as the subjective analysis using Mean Opinion Score (MOS). The work utilizes Xen® Hypervisor with modified Remus extensions to provide the High Availability environment with GRE tunnelling and newtwork virtualization. Remus approach using checkpoint based is deployed to copy the primary server to the backup server. A range of 40ms – 900ms has been applied as time interval of checkpoint. The results show that the the failover downtime is 1.4 s and mean jitter is 9,98 ms, packet loss 3,12% and MOS 3.61 for Remus 400ms checkpoint. MOS with different checkpoint time interval is also presented

Keywords: *VoIP, High Availability, Virtualization, Tunneling, WAN*

1. Introduction

Traditional telephony carriers using circuit based switching has been designed as voice carrier. It is a very good technology in early communication since users only need voice service. User's needs in communication technology are then increase and varying, such as voice connection, video call, email, instant messaging, web browsing and others multimedia services. These needs, however are not suitable to be implemented using traditional circuit based switching. Thus, packet based switching using Internet Protocol (IP) become more attractive to fulfill this requirement.

Telecommunication technology is also improved, in which the IP based technology had been chosen as next generation network. *Voice over Internet Protocol* (VoIP) introduced to overcome future telephony demand. Public or private networks have been utilised as the backbone of VoIP service. The utilisation of established internet network is one of VoIP advantage, therefore telecommunication providers do not need to build another special network to run VoIP service. However, these rapid changes are having problems, hence various researches have been done to make sure IP telephony could provide availability and reliability equally to circuit based telephony.

One of the most significant challenges in next generation network is how to provide high carrier-grade for VoIP to overcome the public telephony availability and fault tolerant system. Availability means the ability of server to be in state to perform required services that client requested at any given time interval. Telephony industries demand the availability that have to reach 99.999 percent of the time, well known as five nine rules, which means those networks must have the maximum of 5 minutes downtime over a year [1]. VoIP as new emerging technology in telephony uses non reliable protocol since IP is firstly developed for packet delivering. However, to be able to replace the traditional telephony, this availability requirement should be fullfilled.

Received December 8, 2014; Revised March 14, 2015; Accepted April 10, 2015

The work on voice quality for VoIP has been done in several occasions. It has been reported in [2] that the transmission of VoIP can generate network congestion due to weak supervision of the traffic incoming packet, quieng and scheduling. The congestion affected several quality of service (QoS) parameters e.g, unstable voice packet delivery, packet jitter, packet loss and echo, etc. To cope with this issue, the authors introduced a new queuing scheduling algorithm based on combined PQ algorithm and fuzzy logic. The proposed algorithm classified differentiated packet from incoming packet and also reduced recursive loop and starving as occurred in normal PQ algorithm. At the end it reduced delay in VoIP network. In addition, the authors in [3] used node disjoint multipath routing protocol to provide fault-tolerance system and high QoS to alert packets caused by unusual and critical events. Moreover, the FELGossiping protocol is also deployed to route the routine data packets, update the neighbors table and detect the failed nodes. The results showed that the network layer can achieve short average end to end delay and very low packet loss rate for alert packets.

Recent developments in the field of virtualization have led to a renewed interest in how to take advantages of virtualized technology in order to support high availability and fault tolerant system. Virtualized environment offer high availability and or fault tolerant for server using continuous live migration of virtual machine between primary server and backup server. Nowadays, virtualization technique mostly used by enterprise for hardware efficiencies. Therefore, the high available service in virtualization is also studied. Live migration in virtualization is able to replicate running server into its backup server. When failures occur the backup server took over the service without disruption.

The work in [4] studied high available ability in Local Area Network (LAN) using Xen® Hypervisor. It is an x86 virtual machine monitor that allows multiple commodity operating systems to share conventional hardware in a safe and resource managed fashion, but without sacrificing either performance or functionality. This is achieved by providing an idealized virtual machine abstraction to which operating system such as Linux, BSD and Windows XP, can be ported with minimal effort. By allowing 100 operating systems to run on a single server in the experimental system, Xen® Hypervisor system reduced the associated costs by two orders of magnitude. Furthermore, by turning the setup and configuration of each OS into a software concern, it facilitated much smaller granularity timescales of hosting.

The application transparent solution using continuous live migration in LAN is presented in [5]. A novel system called Remus was exploited for retrofitting high availability onto existing software running on commodity hardware by using virtualization to encapsulate a protected virtual machine.The frequent whole system checkpoints was performed to asynchronous replicate the state of a single speculative executing virtual machine. However, since this approach is not developed for real time service, it generated very high jitter, latency and packet loss. Classification of packet proposed by paper [6] is used in order to decrease those issues. Work in [7] expand the research to simulate the real server load by using call generator. This work showed the measurement of the objective voice quality analysis using the QoS attributes i.e., jitter, delay and packet loss and also utilises the subjective analysis using Mean Opinion Score (MOS).

In order to provide scalability and better service, providers of VoIP server may not locate its server only in one location of data center. Even though establishing several servers in several locations is highly recommended, consequently continuous live migration over the wide area environment is required. Wide area characteristics e.g., lack of bandwidth, higher delay, jitter and packet loss cause the replication of links more unreliable and become new challenges to overcome. The failover mechanism of Xen with Remus extensions utilise the gratuitous Address Resolution Protocol (ARP) broadcast. However, this mechanism is not suitable for Wide Area Network (WAN) since ARP broadcast is distributed only on its own local networks of origin The impact of WAN failover for server live migration was explained in this paper [8]. This paper also proposed tunneling and the dynamic Dynamic Name System (DNS) for wide area packet re-direction, when the backup server will have a new IP address after failover. Seamless live migration over Metropolitan Area Network (MAN) and Wide Area Network (WAN) is also conducted iny this paper [9]. However, those papers scenario only studied the original live migration instead of continuous live migration. There is no best solution right now since a lot of researcher still working on their enhancement. The paper in [10] provided work in the enhancement of continuous live migration to study the problem of continuous live migration over

WAN environment. The works proposed using Border Gateway Protocol (BGP) update to announce the new path to the server after failover.

In this paper, the failover mechanism in WAN using GRE tunneling and network virtualization is proposed. Downtime when the failover occurs and the impact to the voice quality is investigated and examined.

2. High Available VoIP Challenges

Availability is the ability of the server to be able to serve the client request at any given time interval. Circuit based telephony is designed to deliver high availability, reliable connection and high quality of voice. Since VoIP relies on IP, packetization and encoding process may caused high delay, jitter and packet loss. Therefore, to provide high available and fault tolerant VoIP service to met the telephony standard is challenging.

The general model of high available VoIP service is depicted in figure 1. One server acts as primary server that serve the clients request. The primary server states are frequently copied to the backup server. Backup server took over the job to serve the client request if the primary server fail with the last primary state copies with minimal downtime well known as fail over mechanism.

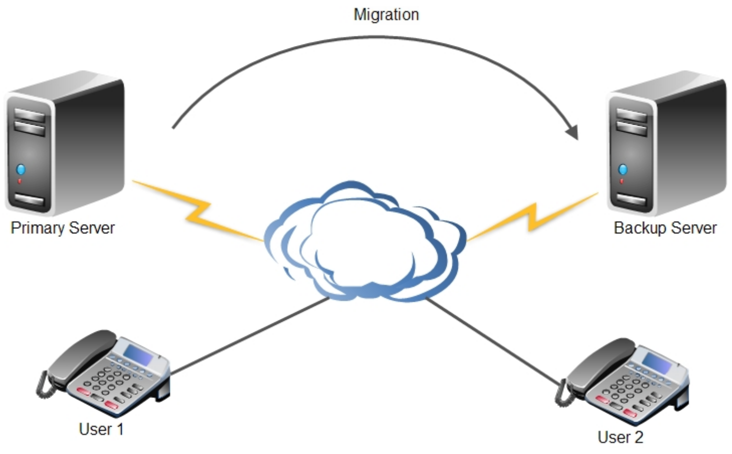

Figure 1. High availability model of VoIP Server

Virtualization technology offers not only hardware efficiencies but also high availability solution by continuous live migration of its running virtual machine between the primary server and backup server. Live migration using Xen is enhanced by the work in [5] to provide a solution for transparent continuous live migration using Remus. Since Remus is a lower layer availability protection, therefore it is not required to modify any application that running on top of it. The Remus work by sent copies of all primary machine states to the paused virtual machine at the backup server. Whenever failures occur on the primary server then the paused virtual machine at the backup server is resumed with the last checkpointing states therefore downtime of the service is minimized.

This failover is transparent from user point of view, backup machine sent gratuitous ARP broadcast to its LAN segment to announce the position of virtual machines. This gratuitous ARP notified the client to redirect the packet request to the new backup machine since virtual machine on the primary server has been failed. However, this failover mechanism is not acceptable in a wide area environment. The gratuitous ARP broadcast may not received by client in different network i.e., WAN connection consists of numbers different network. Therefore, the failover mechanism to guarantee the minimal downtime in wide area environment should be improved to meet WAN requirements. In this paper the tunneling mechanism and network virtualization to overcome the problem is considered.

3. Simulation and Results

In this paper, the proposed idea of failover mechanims over WAN and the impact of high available service in VoIP calls quality is studied and examined. Packets during VoIP call using High Available VoIP service is captured for further objective voice quality analysis using both the QoS attribute i.e., jitter, delay and packet loss, and the subjective analysis using Mean Opinion Score (MOS). The ITU – T standard in [11] describes parameters for good telephony communication i.e., the average MOS 4.4, delay < 250 ms, jitter < 30 ms and packet loss < 5 %.

Tunneling is a common way that used in WAN to provide a transparent network link between two or more devices in different network segment or remote area. Therefore, by creating the tunnel between primary to the backup server and also the client, it is ensuring the migration and access traffic transparent. IP tunneling also enabled virtual machine to keep the same IP address. Our proposed idea, as depicted on figure 2, shows a high level model of failover model in WAN. The works propose tunnel switching to redirect the network packet when failover occurs. Only 1 Generic Routing Encapsulation (GRE) tunnel or access link between client and physical server is active/up, depends on where the protected virtual machine is relied on.

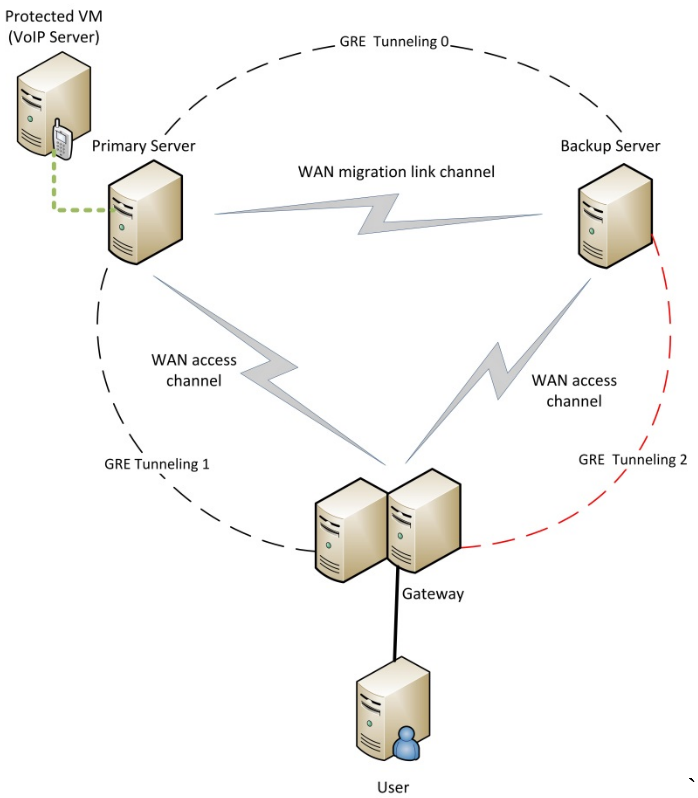

Figure 2. High level of test bed architecture.

If the protected virtual machine is running on the primary server the GRE Tunnel 1 is activated. This GRE Tunnel 1 enabled client to access the server while the Remus process is running through GRE Tunnel 0 or WAN migration channel. The measurement on this paper is done by loading up the High Available VoIP server that run in primary server with calls using VoIP call generator with minimum CPU load of 50%, representing the real server work load. Then, another call is made using SIP Phone for several minutes and the packet is sniffed to capture the data packets for further objective and subjective voice quality analyses. All call is using SIP signalling protocols and G.711 codec since it is widely used by VoIP systems.

After 1 minutes call, virtual machine on the primary server is shutting down to trigger the failover. When failover occured, backup machine sent a notification to the client to redirect the traffic to the backup host machine. This notification drops the GRE Tunnel 1 and GRE Tunnel 2 are established to guide the traffic into the backup machine to minimize the downtime. The voice call should not dropped since our system is high available but the voice will be interupted as long as the server downtime. All captured packet is analyzed using packet sniffer for further qualitative analysis and using perceptual voice quality measurement for subjective analysis.

In this result, jitter calculation is considered to present the impact of Remus continuous live migration in VoIP communications. Real-time Transport Protocol (RTP) is a protocol defined by IETF RFC 3550 [12] that is responsible to provide real time data delivery service. Jitter is calculated using interarrival jitter (J) and mean deviation of the difference (D) defined for pairs packet as shown on equation 1.

$$J(i) = J(i-1) + \frac{(|D(i-1,i)| - J(i-1))}{16} \quad (1)$$

Figure 3 shows the impact in High Availability VoIP service. The delay, jitter and packet loss are still acceptable. Maximum jitter is 17,02 ms, the mean jitter only 9,98 ms, packet loss also only 3,12%. Further analysis of subjective attribute using MOS is estimated using voice quality analyser. This analyser works by sending the recorded voice during the call and compared the echoed voice with original voice and estimates the MOS. This work also proof that user defined checkpointing interval has a significant impact in voice quality but does not impact the downtime interval, the result is shown in table 1. According to the result, the most suitable checkpoint interval is every 400 – 600 ms.

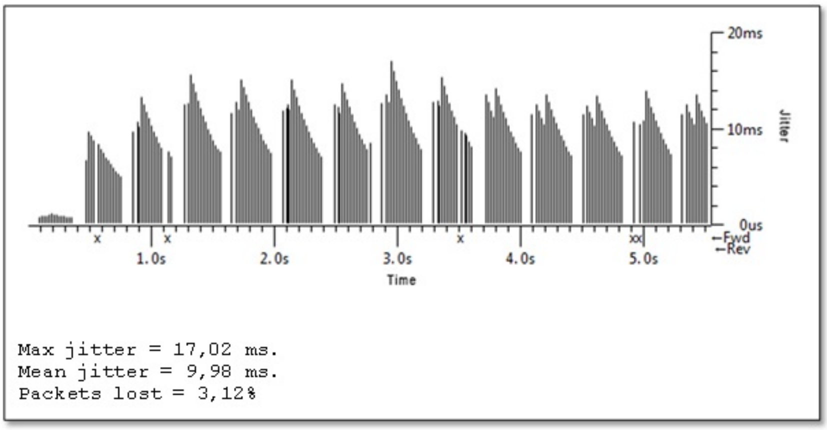

Figure 3. Jitter level in modified Remus HA system

Table 1. Checkpoint interval impact

No	Remus checkpoint interval (ms)	Downtime (s)	MOS
1	40	1.3	2.73
2	70	1.3	2.84
3	100	1.2	2.95
4	150	1.4	3.12
5	200	1.3	3.23
6	300	1.4	3.40
7	400	1.4	3.42
8	500	1.3	3.42
9	600	1.5	3.49
10	700	1.5	3.48
11	800	1.4	3.48

4. Conclussion and Future Works

This work shows that tunnelling and virtual network is promising for failover mechanism. The result show that the downtime is still comparable to the gratitous ARP. The impact of high available VoIP Server in WAN to the jitter, delay, packet loss and MOS of voice quality is still acceptable according to the guidelines of ITU-T. The results shows that the mean jitter is 9,98 ms, packet loss 3,12% and MOS 3.42 for Remus 400ms checkpoint. Further work should consider the optimization of packet compression due to network characteristics i.e., high delay, jitter and packet loss in WAN

References

[1] Gray J, Siewiorek DP. High-availability Computer Systems. *IEEE Computer Magazines.* 1991; 24(9): 39–48, ISSN: 0018-9162.

[2] Suardinata, Abu BK. A Fuzzy Logic Classification of Incoming Packet for VoIP. *TELKOMNIKA Telecommunication Computing Electronics and Control.* 2010; 8(2): 165–174.

[3] Mekki K, Zouinkhi A, Abdelkrim MN. Fault-tolerant and QoS Network Layer for Security Management. *TELKOMNIKA Telecommunication Computing Electronics and Control.* 2013; 11(2): 363–372.

[4] Barham P, Dragovic B, Fraser K, Hand S, Harris T, Ho A, Neugebauer R, Pratt I, Warfield A. *Xen and The Art of Virtualization.* Proceedings of the nineteenth ACM Symposium on Operating Systems Principles. 2003: 164–177.

[5] Cully B, Lefebvre G, Meyer D, Feeley M, Hutchinson N, Warfield A. *Remus: High Availability via Asynchronous Virtual Machine Replication.* Proceedings of the 5th USENIX Symposium on Networked Systems Design and Implementation. 2008: 161–174.

[6] Hlavacek J, Bestak R. *Live Replication of Virtualized VoIP Servers*. The Eighth International Multi-Conference on Computing in the Global Information Technology (ICCGI). 2013.

[7] Septama HD, Ulvan A, Hlavacek JB. *Voice Quality of VoIP in High Availability Environment*. International Conference on Electrical Engineering and Computer Science. 2014

[8] Bradford R, Kotsovinos E, Feldmann A, Schi¨oberg H. *Live Wide-Area Migration of Virtual Machines Including Local Persistent State*. VEE '07 Proceedings of the 3rd international conference on Virtual execution environments. 2007.

[9] Travostino F, Daspit P, Gommans L, Jog C, de Laat C, Mambretti J, Monga I, van Oudenaardec B, Raghunatha S, Wange P. Seamless live migration of virtual machines over the MAN/WAN. *Future Generation Computer Systems 22*. 2006.

[10] Rajagopalan S., Cully B, O'Connor R, Warfield A. *SecondSite: disaster tolerance as a service*. VEE '12 Proceedings of the 8th ACM SIGPLAN/SIGOPS conference on Virtual Execution Environments. 2012.

[11] One-way transmission time. ITU.T Recommendation G.114. International Telecommunication Union Telecommunication Standardization Sector (ITU.T). 2003.

[12] Schulzrinne H, Casner S, Frederick R, Jacobson V. RTP: A Transport Protocol for Real-Time Applications. IETF RFC 3550. Internet Engineering Task Force (IETF). 2003.