Comprehending SIP server overload control methodologies - A literature survey

Abdullah Akbar[1], S. Mahaboob Basha[2], Syed Abdul Sattar[3]

[1] Research Scholar, Department of C. S. E., JNTUH, Telangana, India
[2] Principal, Al Habeeb College of Engineering and Technology, Telangana, India
[3] Dean of Academics, Royal Institute of Technology and Science, Telangana, India

---***---

Abstract - *Session Initiation Protocol is very popular mechanism widely used in Voice over Internet Protocol call centers and other applications where voice, data, video and multimedia transmission is incorporated. The predominant role of SIP servers are sometimes experiencing the overload due to the heavy traffic of messages on the SIP servers and lead to collapse. The dramatic increase of message transmission to the SIP server from the User Agents will crease more load and the capacity of SIP servers can't tolerate to handle the calls or messages. To overcome this problem of overload a control mechanism has been developed by many predominant research scholars. Different kinds of models have introduced to control the high load on the servers and to handle the calls or messages from User agents in the form of inbound and outbound calls. The present paper is focusing on the best methods of overload control employing different latest techniques and methods from recent published research works.*

Key Words: SIP (Session Initiation Protocol), Overload Control

1. Introduction

Session Initiation Protocol (SIP) is predominantly used in VoIP Networks as a signaling communication protocol for initiating, modifying, controlling and terminating the multimedia sessions over Internet Protocol based networks. The multimedia sessions are rich with voice calls, text messages, Video streams and other multimedia formats transmitted between two user agents through SIP servers. The inbound calls and outbound calls are predominantly performed in SIP based telephony network services. VoIP based telephony network services are introduced recently in the place of Public Switched Telephone Networks (PSTN). VoIP signaling communications are comparatively very economic pricing and serves the same purpose as PSTN networks provides. The economic pricing has given more popularity and wide spread usage in the segment of small and medium sized entrepreneurs to operate international call centers with

the implementation of SIP servers to handle inbound and outbound calls.

The one and only disadvantage in managing and maintaining the SIP servers is overload problem when the more calls or messages are received than the hardware capacity of the SIP servers. This situation leads to collapse of the servers.
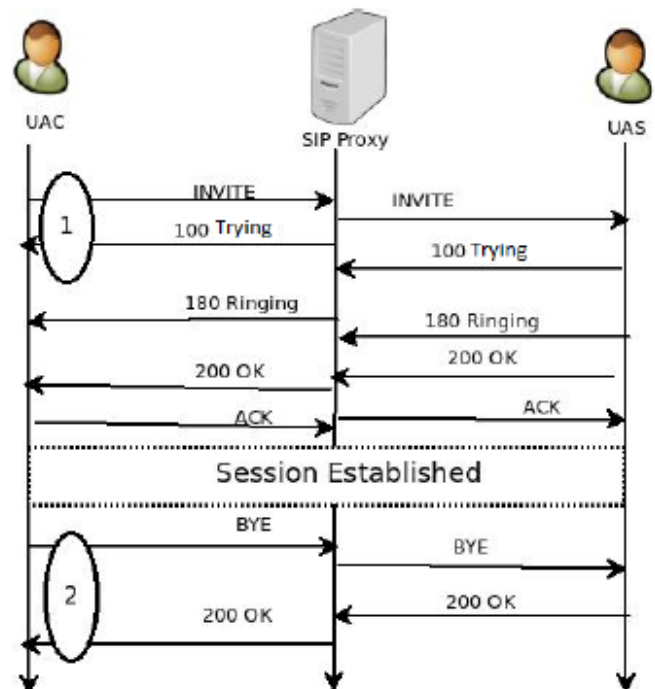

Figure 1 Basic Functionality of SIP Servers

The session initiation protocol server will be in between the client A and client B. The basic functionality of SiP Server is to forward the request to the clients around the server in the network. The messages or calls are transmitted between the clients through SIP servers. The main functionality of SIP server is to manage, termite or initiate the calls and messages to the clients. When the more messages are transmitted between the clients or another SIP network to the clients of local networks can increase the traffic. At this juncture it is difficult to manage the calls or messages in between the clients and leads to collapse of SIP servers.

SIP server load has to be balanced to escape from collapsible situation. The need of research has been developed to find the suitable method to manage and maintain the SIP servers without collapsing while the inbound and outbound calls are functioning. The real time situation has been taken by many research scholars and developed several algorithms and methods to control the load on SIP Servers.

Predominantly the algorithms and methods with controlling techniques have been divided into two major categories. These can be distinguished as local and distributed techniques to control the load on SIP servers. The local methods and techniques are ancient whereas the distributed techniques are modern and more effective. In the following section SIP session overload controlling techniques are discussed.

2 Literature Survey

Session initiation Protocol servers are widely used in VoIP networks and needs a mechanism to overcome the problem of overload in transmitting the messages. Several algorithms have been developed to suppress the overload problem of SIP servers.

Load balancing processes and fail over management tactics to use in web servers has been matured in research literature. Session Initiation Protocol is quite unique with respect to the HTTP standards, there is less centrality for sessions in HTTP while SIP dispatcher ought to keep record of the sessions so as to course messages of same session to the same intermediary / proxy server. For giving high accessibility in web servers; for the most part IP address takeover and MAC address takeover plans are utilized. Y. Hong, C. Huang and so on recognized two expansive SIP over-burden control method components: Load reducing and Load balancing approach. Load reducing so as to diminish incoming call solicitations to keep away from over-burdening the server, on the other hand load balancing just as divides the traffic movement among nearby SIP servers to dodge over-burden in SIP systems [7]. Load balancing algorithms for SIP server groups are tagged as static and dynamic. Cisco's proprietary architecture primarily employs static load balancing taking into account DNS SRV and Naming Authority Pointer records. Load dispatcher chooses the server in view of priority and weight fields of these records. Kundan Singh and Henning Schulzrinne introduced web server repetition methods to SIP telephony. They have portrayed different load sharing routines in view of DNS, SIP identifier, redundant servers with special IP address and Network Address Translation on a two-stage SIP server setup. They have likewise considered fall fail over strategies in view of DNS, clients, database replication and IP address takeover [11]. Wright et.al concentrated on SIP call executions on multi center

frameworks [12]. Hongbo Jiang and Arun Iyengar and others presented three novel element load balancing methodologies for SIP Server farms. These calculations depend on Session Aware Request Assignment method. Every calculation takes a shot at functionality by joining learning of SIP, recognition of calls and transactions and on-the-fly appraisals of server burden. Transaction Least Work Left outflanks the other remaining algorithms in measures of throughput and reaction time [13]. Luca De Cicco and Giuseppe Cofano and so on proposed a novel nearby over-burden controller utilizing control theoretic procedure called extremum looking / seeking. They had basically executed in Kamailio server yet does not control the over-burden in whole SIP system [10]. Jianxin Liao and Jinzhu Wang et.al portray a dispersed end-to-end SIP overload control system edge servers of SIP systems. They have given hypothetical investigation and broad simulations on NS-2 test system to demonstrate the control framework [14]. Ahmadreza Montazerolghaem and Mohammad Hossien Yaghmaee Moghaddam actualized window based versatile overload control in Asterisk Proxy servers [5].

2.1 Rate based overload control algorithm

In these Abdullah Akbar et.al [2015] has been regarded as the best solution for SIP overload problem. The solution has been developed in the form of rate based overload control method. This control method is rich with overload control algorithms. The implementation of congestion control algorithm is implemented on Hop-by-hop and end-to-end process of requests and sending messages from the clients. Implementation of overload control algorithm will understand the rate of priority of the message or call from each client and the least rate based call can be dropped and manage the load on the server. This type of mechanism has effectively and efficiently takes control over the server and skips the collapse situation with the overload problem. The dropped calls are processed and re send to the destinated client without message lost. This rate based overload controlling method is regarded as the latest and fastest methods among the other methods developed so far in the world.

2.2 Two Stage Secure Dynamic Load Balancing Architecture

G. Vebbuka et.al has introduced a novel load balancing architecture to SIP server clusters. In this architecture active least call SIP Server algorithm is playing a vital role to manage the load on the server. This architecture has facilitated the main objectives of load balancing like congestion avoidance, improved response times, throughput, resource utilization, reducing server faults, scalability and protection of SIP call from DoS attacks. This architecture has dynamically controlled the overload and

provided the robust security for SIP Servers. The specific architecture with ALC_Server method is regarded as the better method to Least Session Method to measure various parameters like throughput, CPU Utilization and response time. The specific architecture is rich with secure first tier honey wall architecture.

2.3 Local Overload Control

The scientific book named emerging research in computing, information Communication and application has described the SIP server over controlling mechanism with Local overload control with a mathematical model for an ideal rate-based control algorithm. In this mechanism the upper bound to the goodput achievement has been obtained. The mathematical model has successfully calculated the overload of the server and arranged the messages or calls in a queue system. It also developed a set point to guarantee a target response time for CPU utilization with a target. As soon as the traffic reaches the set point automatically it reduce the traffic and drops the messages. The local overload control method is the ancient and successfully delivered the optimal goodput function to the SIP servers. The mathematical calculation developed in this method is distinguished the CPU load for INVITE messages and RECEIVE or BYE messages. Based on the rate of these messages the withdrawal of calls or messages can be performed by the method. To demonstrate the results Kamailio open source SIP proxy has been used.

3 Other Related work on SIP Servers

The overload control methods have been developed by many research scholars these have been published in many books like issued in networks research and application, Internet of thins smart spaces and next generation methods, book on overload control methods – SIP. In the book published on Principles, systems and applications of IP Telecommunications, the author Henning Schulzrinne, Radu State and Saverio Niccolini [2015] has emphasized on some methods to avoid the overload on SIP servers. In this book a set of advanced telephony applications have been described. A java based programming framework has been introduced dedicately support the high level communication mechanism. The book has illustrated the supporting mechanism for SIP servers to compete with increased load within the existing infrastructures and entities.

The book emerging research in computing information, communication and Applications has depicted the overload control algorithm to increase the throughput and reduce the call set-up delay and retransmission rate. To demonstrate the mechanism an open source SIP Proxy server has been replicated by Kamailio server and demonstrated the results effectively -N. R. Shetty et.al [2015].

4 Conclusion

SIP servers are playing predominant role in the functionality of VoIP networks. VoIP Networks are playing predominant role in the multimedia transmission calls for international calling and substituted PSTN networks. The overload traffic problems have degraded the performance of SIP servers. The overload control methods are distinguished as local and distributed. The distributed overload control methods are latest and fastest. The proposed paper has illustrated the predominant and effective methods to control the overload problems of SIP servers with previous research articles. The books related to overload control methods have been verified and depicted in this paper. The paper has successfully illustrated different overload control methods for SIP servers.

5 Future Scope of Study

The future scope of the study is limited to perform the experimental comparative analysis for overload control methods developed in previous research papers. The future scope of the study should depict the comparative analysis results of the control methods for overload problems for SIP servers in a controlled unbiased environment.

6 References

6.1 Journals

[1] Abdullah Akbar, S. Mahaboob Basha and Syed Abdul Sattar [2015] A Novel Rate Based Overload Control Method for SIP Servers published in Springer New Delhi Heidelberg New York Dordrecht London © Springer India 2016

[2] Abdullah Akbar, S. Mahaboob Basha, and Syed Abdul Sattar [2015] A Cooperative Overload Control Method for A Cooperative Overload Control Method for SIP Servers published in International Conference on Communication and Signal Processing, April 2-4, 2015, India ISBN 978-1-4799-8080-2

[3] Abdullah Akbar, S. Mahaboob Basha and Syed Abdul Sattar [2015] A Comparative Study on Load balancing algorithms for SIP servers published in 379 CRC India for 2016 papers

[4] Bikash Upadhyay, Atul Mishra and Shraddha Upadhyay [2013] AIPC: Mathematical Model for SIP Server Overload Control published in INTERNATIONAL JOURNAL OF COMPUTER ENGINEERING & TECHNOLOGY (IJCET) Volume 5, Issue 1, January (2014), pp. 128-140

[5] Ahmad reza Montazerolghaem, Mohammad Hossein Yaghmaee [2013] SIP OVERLOAD CONTROL TESTBED: DESIGN,BUILDING AND EVALUATION published in International Journal of Ambient Systems and Applications (IJASA) Vol.1, No.2, June 2013

[6] Pavel O. Abaev, Yuliya V. Gaidamaka and Alexander V. Pechinkin [2012] SIMULATION OF OVERLOAD CONTROL IN SIP SERVER NETWORKS published in ECMS, The European Council for Modelling and Simulation Digital Library [2012]

[7] Yang Hong, Changcheng Huang, James Yan [2012] A Comparative Study of SIP Overload Control Algorithms published in Cornell University Library (CUL)

[8] Kiran Kumar Guduru and Dr. Usha J [2014] Overload Detection and Controlling Techniques in SIP Servers A Survey IRACST – International Journal of Computer Networks and Wireless Communications (IJCNWC), ISSN: 2250-3501 Vol.4, No3, June 2014

[9] G. Vennila* and MSK. Manikandan [2014] Two Stage Secure Dynamic Load Balancing Architecture for SIP Server Clusters published in JOURNAL OF Engineering Science and Technology Review Received 7 July 2014; Accepted 1 August 2014.

[10]   Luca De Cicco, Giuseppe Cofano and Saverio Mascolo [2015] Local SIP Overload Control: Controller Design and Optimization by Extremum Seeking published by IEEE TRANSACTIONS ON CONTROL OF NETWORK SYSTEMS, VOL. 2, NO. 3, SEPTEMBER 2015

[11]   Kundan Singh and Henning Schulzrinne "Failover and Load Sharing in SIP Telephony", Journal of Computer Communications, vol. 35(5), pp. 927-942, 2007

[12]   C.Wright. Nahum D.Wood, J.Tracey and E.Hu,"SIP server performance on multicore systems,"IBM J.Res. Dev., vol.54, no. 1,Feb.2010

[13]   Hongbo Jiang, ArunIyengar, Erich Nahum, Wolfgang Segmuller, Asser N. Tantawiand Charles P.Wright, "Design, Implementation, and Performance of a Load Balancer for SIP Server Clusters" , IEEE/ACM Trans Networking, Vol. 20, No. 4, pp.1190-1202, Aug 2012

[14]   Jianxin Liao, Jinzhu Wang, Tonghong Lib, Jing Wang, Jingyu Wang, Xiaomin Zhu,"A distributed end-to-end overload control mechanism for networks of SIP servers",Science Direct, Computer Networks Volume 56, Issue 12, pp 2847-2868,Aug 2012

6.2 Books

[1] N. R. Shetty, N.H. Prasad, N. Nalini [2015] Emerging Research in Computing, Information, Communication and Applications: ERCICA 2015, Volume 2 published by Springer, 21-Aug-2015 - Technology & Engineering - 600 pages

[2] Vassilis Tsaoussidis, Andreas Kassler, Yevgeni Koucheryavy, Abdelhamid Mellouk [2013] Wired/Wireless Internet Communication: 11th International Conference, WWIC 2013, St. Petersburg, Russia, June 5-7, 2013. Proceedings published by Springer, 29-May-2013 - Computers - 277 pages

[3] Alexander Dudin, Valentina Klimenok, Gennadiy Tsarenkov, Sergey Dudin [2012] Modern Probabilistic Methods for Analysis of Telecommunication Networks: Belarusian Winter Workshops in Queueing Theory, BWWQT 2013, Minsk, Belarus, January 28-31, 2013. Proceedings published in Springer, 06-Dec-2012 - Computers - 217 pages

[4] Jerzy Swiątek, Adam Grzech, Jerzy Świątek, Jakub M. Tomczak [213] Advances in Systems Science: Proceedings of the International Conference on Systems Science 2013 (ICSS 2013) published by Springer Science & Business Media, 13-Aug-2013 - Computers - 809 pages

[5] Henning Schulzrinne, Radu State, Saverio Niccolini [2008] Principles, Systems and Applications of IP Telecommunications. Services and Security for Next Generation Networks: Second International Conference, IPTComm 2008, Heidelberg, Germany, July 1-2, 2008. Revised Selected Papers Springer, 23-Oct-2008 - Computers - 343 pages

[6] Q. Ashton Acton[2013]Issues in Networks Research and Application: 2013 Edition 01-May-2013 - Computers - 1087 pages published by Scholarly Editions

[7] Sergey Balandin, Sergey Andreev, Yevgeni Koucheryavy [2013] Internet of Things, Smart Spaces, and Next Generation Networking: 13th International Conference, NEW2AN 2013, and 6th Conference, ruSMART 2013, St. Petersburg, Russia, August 28-30, 2013. Proceedings published by Springer, 23-Aug-2013 - Computers - 440 pages

7 BIOGRAPHIES

Abdullah Akbar is a research scholar at the Department of Computer Science Engineering. JNTUH, Hyderabad, Telangana, India. (akbar.jntuphd@gmail.com)

Shaik Mahaboob Basha is the Principal at Al Habeeb College of Engineering and Technology, R.R District, Telangana, India. (smbasha.phd@gmail.com)

Syed Abdul Sattar is the Dean of Academics at Royal Institute of Technology and Science, R.R District, Telangana, India. (syedabdulsattar1965@gmail.com)