PERFORMANCE ANALYSIS OF STRATEGIES FOR CONTROLLING OVERLOAD AND IMPROVING SCALABILITY IN SIP BASED NETWORKS

GARIMA MISHRA

BHARTI SCHOOL OF TELECOMMUNICATION TECHNOLOGY AND MANAGEMENT
INDIAN INSTITUTE OF TECHNOLOGY DELHI

JULY 2016

PERFORMANCE ANALYSIS OF STRATEGIES FOR CONTROLLING OVERLOAD AND IMPROVING SCALABILITY IN SIP BASED NETWORKS

by

GARIMA MISHRA

BHARTI SCHOOL OF TELECOMMUNICATION TECHNOLOGY AND MANAGEMENT

Submitted
in fulfillment of the requirements of the degree of

Doctor of Philosophy
to the

INDIAN INSTITUTE OF TECHNOLOGY DELHI

JULY 2016

To my parents, husband and son.

Certificate

This is to certify that the thesis entitled **"Performance analysis of strategies for controlling overload and improving scalability in SIP based networks"** being submitted by **Mrs. Garima Mishra** to the Bharti School of Telecommunication Technology and Management, Indian Institute of Technology Delhi, for the award of the degree of **Doctor of Philosophy** is the record of the bonafide research work carried out by her under our supervision. In our opinion, the thesis has reached the standards fulfilling the requirements of the regulations relating to the degree.

The results contained in this thesis have not been submitted either in part or in full to any other University or Institute for the award of any degree or diploma.

Dr. Subrat Kar
Professor
Department of Electrical Engineering
and
Bharti School of Telecommunication
Technology & Management
Indian Institute of Technology Delhi

Dr. S Dharmaraja
Professor
Department of Mathematics
and
Bharti School of Telecommunication
Technology & Management
Indian Institute of Technology Delhi

Date:
Place: New Delhi

Acknowledgments

First of all, I would like to express my deepest gratitude to all those who extended their valuable support to me so I could complete my work. To start with, I would extend my prayers to the most generous, merciful, the almighty God for keeping his utmost blessings on me.

I am deeply grateful to my thesis supervisors Prof. Subrat Kar and Prof. S. Dharmaraja for their invaluable guidance. Prof. Subrat Kar is the epitome of goodness , sincerity and work ethics. His sincere advice, positive approach, constructive criticism and high encouraging power kept me going even when the going was difficult. I would also like to express my sincere gratitude to Prof. S. Dharmaraja for his support, advice, comments, guidance that have benefited me in the completion and success of this study. I have learnt many valuable lessons from him such as discipline, organization, dedication which have not only helped me in my work but will always be helpful in my career ahead. It is hard to express my feelings in words.

I take this opportunity to thank my SRC committee Prof. H. M. Gupta, Prof. Huzur Saran and Prof. Swades De for their support and valuable suggestions.

I would also extend my sincere thanks to my parents who sacrifice their sleep and comfort to ensure I accomplish my dreams. They are my best admirers and loving critics that makes me feel like hero even before I could achieve any big feet in the world. Thanks again for their unending love, motivation and warmth.

I am obligated by the huge support provided by my husband, Mr. Ritesh Verma. His selfless care that was huge motivation for me to keep going cannot be expressed in words. It also gives me immense pleasure in acknowledging the understanding and love extended to me by my parents-in-law. One of the best experiences that I lived through in this eventful period was the birth of my son Darsh, who provided an abundance of happiness and added a joyful dimension to my life. I could not spend as much time with him as I would have liked to and now hope to compensate for all the missed time.

I extend my heartful thanks to Ms. Charu Tyagi, Ms. Binish Fatima and Ms. Payal

Sinha Nath. They always gave me a feeling of being at home by making those tough moments lighter with their sense of humor when I was lost under pressure. Their smiles with generosity gave me a reason to fall back on them for any kind of support.

In the end, I would also like to thank my colleagues from the research scholar lab. They were generally a very happy and joyful group. The conducive atmosphere created by them really increased my productivity to great extent.

Garima Mishra

Abstract

Recent collapses of Session Initiation Protocol (SIP) servers in the carrier networks indicate two potential problems of SIP: (1) the built-in SIP overload control mechanism cannot handle overload conditions effectively, and (2) the current SIP design does not easily scale up to large network sizes. In order to help carriers prevent widespread SIP network failures effectively, this thesis presents a systematic investigation of current state-of-the-art overload algorithms and SIP-based group communication frameworks adopted by SIPPING, XCON and MMUSIC working groups of IETF. To achieve this goal, first, we propose an analytical model for the re-designed SIP proxy server to control overload locally and develop an algorithm to control overload hop-by-hop. Then a packet aggregation and compression mechanism is provided which improves the capability of single conferencing servers to handle large number of concurrent users.

SIP has a retransmission mechanism to maintain the reliability for its real time transmission. But these real time transmissions cause overload in the server and creates redundant messages. A feedback response code 503 (Service Unavailable) is sent from a SIP server once it experiences overload. It is a built-in functionality of SIP which provides a limited overload control, thus making SIP servers susceptible to overload. This limited overload control mechanism oscillates signaling traffic and shifts it between their neighbouring servers and make overload condition worse (RFC 5390). In this thesis, local overload mechanism is implemented within a SIP proxy server without requiring cooperation between servers. The local overload control is achieved by redesigning the SIP proxy server as a multi-server cut-off priority queue model. We design SIP proxy server by separating signaling traffic in two different classes (`Invites` and `Non-Invites`) and creating a cut-off priority queueing model. In order to study

the impact of overloaded SIP proxy server on throughput, the entire end-to-end SIP signaling flow is modeled between a sending and a receiving SIP servers. We propose a complete end-to-end SIP signaling flow for session establishment between two end users using Generalized Stochastic Petri Nets (GSPN) technique.

Next, we study the **hop-by-hop overload control** mechanism by extending the previously developed SIP proxy model for timed-out packets which cause retransmissions. SIP retransmission mechanism with timeout is modeled as a queueing system with impatient customers. Using this model, the effect of unnecessary retransmissions is studied and delay distribution and loss probability (Pb_loss) is calculated. The proposed analytical model is verified with simulations which demonstrate that the inclusion of timeout gives better delay performance. Using Pb_loss, an algorithm is developed to control the overload in hop-by-hop transaction as described in RFC 6357 [1]. The simulation results for the proposed model with overload control and the standard SIP as per RFC 3261 [2] are compared. The results demonstrate that the server utilization factor of an overloaded server always remains less than one and hence avoids system collapse. Further, the redundant messages in the system are reduced by 30% as compared with the standard SIP network.

SIP signaling protocol is also used for providing group communication in IP-based next generation network. The conferencing framework proposed by Internet Engineering Task Force (IETF) working groups has limitation for large scale networks due to their control over entire conference by a single central server. In order to overcome this limitation, we propose multiple SIP packet aggregation and compression of common header part of each aggregated packet process. This mechanism increases the bandwidth utilization and processing capability of conference server. A queueing model is presented to describe the process of packet aggregation and compression at the aggregator node. Using the proposed queueing model, the steady state probability distribution of the queue length is derived and the average packet delay is obtained. We show the trade-off between the overhead gain and delay. Based on analytical results obtained, the maximum frame size is optimized for which the average delay of a packet is minimum. The packet aggregation and compression mechanism is implemented using NS2 simulator. The key session setup performance measures discussed in **ITU E.721** are obtained which particularly focus on the SIP protocol requirements. The proposed queueing model is further studied under Min(N, T) policy. We consider two control parameters queue length threshold N and timer T to suppress the increase in delay due to the long frame formation itself. These parameters are optimized such that we get the maximum bandwidth utilization and minimum delay.

We may also note, as a interesting observation based on the above results, that if multi-party games are played on the Next Generation Networks (NGN), it would be better to aggregate and compress packets at the edge of the network. A gaming API with this provision would perform better than one without it.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

3GPP	Generation Partnership Project
AS	Application Server
CSCF	Call/Session Control Function
CTMC	Continuous Time Markov Chain
DA	Destination-based Aggregation
DES	Discrete Event Simulation
FIFO	First In First Out
GSPN	Generalized Stochastic Petri Nets
I-CSCF	Intermediate-Call/Session Control Function
IETF	Internet Engineering Task Force
IMS	IP Multimedia Subsystem
IP	Internet Protocol
LST	Laplace Stieltjes Transform
MFS	Maximum Frame Size
OC	Overload Control
P2P	Peer2Peer
P-CSCF	Proxy-Call/Session Control Function
PGF	Probability Generating Function
QBD	Quasi Birth Death
QoS	Quality of Service
RTP	Real-time Transport Protocol
RTT	Round Trip time
S-CSCF	Server-Call/Session Control Function
SDP	Session Description Protocol
SIP	Session Initiation Protocol

TCP	Transmission Control Protocol
UA	User Agent
UAC	User Agent Client
UAS	User Agent Server
UE	User Equipment
UDP	User Datagram Protocol