

Simulation-based Optimization of Signaling Procedures in IP Multimedia Subsystem

Jasmina Barakovic Husic, Alisa Hidic, Mesud Hadzialic, and Sabina Barakovic

Abstract—This paper presents a simulation-based optimization of signaling procedures in Internet protocol Multimedia Subsystem (IMS). The aim is to improve the performance of signaling procedures by applying an algorithm for Session Initiation Protocol (SIP) message classification and prioritization that is proposed in our previous work. This three-priority level classification of SIP messages is implemented in Network Simulator version 2 (ns-2). Its effectiveness is verified through the simulation-based study of SIP signaling procedures under different conditions. The simulation results are analyzed in terms of Registration Request Delay (RRD), Session Request Delay (SRD), and Session Disconnect Delay (SDD). These SIP performance metrics are improved if IMS is configured to process SIP messages using the proposed algorithm. Differentiated handling of SIP messages reduces the overload in IMS and thereby improves the overall Quality of Service (QoS). This encourages our next step in research activity, i.e., implementation and incorporation of SIP message classification and prioritization algorithm into experimental environment.

Keywords—classification, optimization, prioritization, SIP signaling, simulation

I. INTRODUCTION

THE Internet protocol Multimedia Subsystem (IMS) supports the development of next generation services. This causes the rapid growth in new services, subscribers and devices that increase the signaling volume. The growing amount of signaling may result in congestion and impact the Quality of Service (QoS) [1]. The IMS procedures used for QoS negotiation and signaling are based on Session Initiation Protocol (SIP). These signaling procedures play an important role in affecting the overall Quality of Experience (QoE) [2]. In this regard, there is a need for an optimization

This research is supported by the Federal Ministry of Education and Science in Bosnia and Herzegovina under the project “Signalling Information Management to Improve Quality of Next Generation Services”.

J. Barakovic Husic is with the Faculty of Electrical Engineering, University of Sarajevo, Zmaja od Bosne bb, 71000 Sarajevo, Bosnia and Herzegovina (phone: +387 33 250 700; fax: +387 33 250 725; e-mail: jbarakovic@etf.unsa.ba).

A. Hidic is with the Faculty of Electrical Engineering, University of Sarajevo, Zmaja od Bosne bb, 71000 Sarajevo, Bosnia and Herzegovina (e-mail: ah15310@etf.unsa.ba).

M. Hadzialic is with the Faculty of Electrical Engineering, University of Sarajevo, Zmaja od Bosne bb, 71000 Sarajevo, Bosnia and Herzegovina (e-mail: mhadzialic@etf.unsa.ba).

S. Barakovic is with the Ministry of Security of Bosnia and Herzegovina, Trg BiH 1, 71000 Sarajevo, Bosnia and Herzegovina (e-mail: barakovic.sabina@gmail.com).

of SIP signaling procedures.

In a large IMS network with millions of potential user agents, several thousand or more SIP messages may be processed by individual SIP proxy servers [3]. Conventional SIP proxy servers are configured to process SIP messages using First In First Out (FIFO) scheduling. However, FIFO scheduling may have several disadvantages: (1) it may be inefficient, (2) it does not enable service differentiation, (3) it does not permit differentiation based on the classes of service, and (4) it exposes the network to possible instability because of SIP messages that are allowed to loopback through SIP proxy servers.

In the view of these disadvantages, the SIP message classification and prioritization algorithm is proposed in our previous work [4]. Although this algorithm is primarily intended to improve the performance of IMS in high-load and overload conditions, it may be also considered as an extension of our previously published approach [5], which is based on prioritizing of signaling information transmission. According to this approach, signaling service class is given the highest priority over all other user service classes. Although this approach is signaling protocol independent, it is discussed in the context of SIP. Therefore, our previously proposed algorithm further extends this approach by classifying SIP messages into three priority classes.

This paper presents an implementation of this SIP message classification and prioritization algorithm in Network Simulator version 2 (ns-2). The aim is to analyze the impact of implemented algorithm on the SIP performance metrics and compare the results with those obtained using the conventional FIFO scheduling. The intention is to perform the simulation-based optimization of SIP signaling procedures, especially under high-load or overload conditions, in order to improve their performance as an important factor that contributes to the QoE.

The rest of the paper is organized as follows. Section II summarizes the related work. Section III gives an informal and a formal description of algorithm for SIP message classification and prioritization, which is implemented in ns-2. Section IV considers the algorithm’s impact on SIP performance metrics. It discusses the obtained results, together with their analysis to show that the conclusions are warranted. Section V concludes the paper and outlines open issues for future work.

II. RELATED WORK

While the performance evaluation of IMS is a subject of emerging research activities, there are many related works regarding the SIP server overload control. Two broad categories of SIP overload control mechanisms are identified in [6]: load balancing approach and load reducing approach. Load balancing approach tries to avoid the overload by distributing the traffic load equally among the local SIP servers. Load reducing approach tries to prevent the overload collapse by reducing the traffic load in the whole SIP network. This approach differentiates three categories of mechanisms for SIP overload control: priority-based, push-back, and retransmission-based. Load balancing mechanisms have been deployed in operator networks, while other three types of load reducing mechanisms are in the stage of research proposals.

The problem of SIP server overload control is the subject of many research activities. As a result, a wide range of mechanisms and algorithms have been developed in order to solve this problem. In this respect, three novel load-balancing mechanisms have been introduced in [7]. Each mechanism combines knowledge of the SIP, dynamic estimates of server load, and Session-Aware Request Assignment (SARA). The proposed mechanisms provide finer-grained load balancing resulting in throughput and response-time improvements.

An optimized algorithm for SIP server overload control that randomly makes the decision of acceptance or rejection of every SIP message is proposed in [8]. This algorithm is based on the calculation of the queue length and the reject probability of SIP messages. The simulation results show that the algorithm better meet the demand on SIP signaling network under overload conditions.

A distributed and adaptive window-based overload control algorithm is proposed in [9]. This algorithm controls the amount of calls that are forwarded to a downstream SIP server in an attempt to prevent it from being overloaded, and does not rely on explicit feedback. This algorithm performs better both in terms of call setup delay and throughput than a commonly used overload control algorithm that is based on maintaining Central Processing Unit (CPU) utilization.

A backpressure-based SIP overload control mechanism called Bassoon is proposed in [10]. It consists of two parts: (1) optimal load balancing algorithm that ensures full utilization of available of available network resources, and (2) end-to-end load control algorithm that regulate traffic at the edge of the network. The Bassoon effectively controls overload in SIP networks and outperforms existing schemes in terms of goodput, fairness and responsiveness.

The implementation and comparison of Adaptive Rate Control (ARC) and Support Vector Machine (SVM)-based algorithms are described in [11]. The ARC is a queue delay-based algorithm. The SVM-based algorithm is based on integration of four input data: call establishment delay, queuing delay, SIP 100 Trying status code delay, and database response time. The comparison shows that the SVM-based algorithm outperforms ARC algorithm in terms of goodput.

Most of the SIP overload control mechanisms are focused on User Datagram Protocol (UDP), although Transmission Control Protocol (TCP) is more suitable for the transport of SIP messages [12]. Therefore, a novel mechanism that effectively uses TCP flow control to aid application level SIP overload control has been developed. Other experiments indicate that throughput with SIP-over-TCP exhibits similar overload collapse behavior as that with SIP-over-UDP [13].

Priority-based overload control mechanisms aim to mitigate the overload by rejecting the SIP messages with low priority [6]. Prioritization may be performed by using different SIP message header fields [14]. In addition, four approaches to SIP message prioritization are proposed in [15]. A SIP messages scheduling mechanism that is applied on service broker is presented in [16]. This approach balances the load on application servers and enhances the overall QoS. Furthermore, it is proposed an automatic originator regulation of IMS multiple traffic by stateless signal prioritization based on the types, message order and retransmission of SIP messages within each call session [17]. It is shown that best prioritization is to give a higher priority to a message type appearing at a later stage in each session and a lower priority to a retransmitted message. Moreover, light-weight messages like instant messaging are assigned higher priority than those generated by voice calls.

Considering this, the SIP message classification and prioritization algorithm is proposed in our previous work [4]. This algorithm is discussed in detail in Section III.

III. SIP MESSAGE CLASSIFICATION AND PRIORITIZATION ALGORITHM

This paper presents the ns-2 implementation of our algorithm for SIP message classification and prioritization. The decision to use the ns-2 is based on comparative analysis of several simulators [18]. The simulators are compared in terms of modelling capabilities, credibility of simulation models and results, extendibility, usability, and cost of licenses. The analysis shows that commercial OPNET Modeler provides the largest support for IMS simulation [19]. However, ns-2 is chosen because it is free and open-source simulator [20] that provides IMS functionality by adding an independently developed SIP module [21]. This SIP module is based on ns-2.27 version.

A. Informal Algorithm Description

The SIP message classification and prioritization algorithm involves two modes of operation: (1) normal mode, wherein the SIP messages are processed using FIFO scheduling, and (2) priority mode, wherein the Priority Queuing (PQ) scheduling is used for three-priority level classification of SIP messages. The normal mode of operation is switched to the priority mode when congestion is detected. The congestion is determined by exceeding the predefined queue length. In priority mode of operation, the packet's content is checked, and packets are put into three-priority classes according to SIP message type.

The class 1 includes SIP messages that terminate the communication session, such as BYE or CANCEL, and those that appear at the later stage in each communication session, such as ACK or 2xx status codes. This class of SIP messages is given the highest priority in the overload conditions. The class 2 includes lightweight SIP messages such as REGISTER, MESSAGE, PUBLISH, NOTIFY, and SUBSCRIBE. These types of SIP messages are less delay sensitive and have lower processing time than INVITE messages [3]. The retransmitted messages are also involved into class 2. A recent studies show that the presence service can account for 50% or more of the total signaling traffic that IMS network handles. Thus, the number of NOTIFY messages is several times larger than the number of other SIP messages. Therefore, this class of SIP messages is assigned medium priority. The class 3 includes the SIP messages that establish the communication session such as INVITE, and those that provide provisional responses, such as 1xx status codes. The class 3 messages are given the lowest priority in the overload conditions [4].

B. Formal Algorithm Description

To implement the previously described algorithm, a modification of existing ns-2.27 source code is done. This modification refers to the adjustment of class *Queue* in order to enable two modes of operations, which are required to discuss advantages and disadvantages of the algorithm. The modification of class *Queue* entails the modification of class *DropTail*, which implements FIFO scheduling. This class is derived from class *Queue*. The class *Queue* is child class of class *Connector*. This class is derived from class *NsObject*, which is the base class for all network objects in ns-2. This is shown on Fig. 1, which represents a class diagram of the SIP message classification and prioritization algorithm.

Fig. 2 and Fig. 3 present sequential diagrams of the proposed algorithm that illustrate the sequence of messages between objects in an interaction in case of normal mode and priority mode of operation.

In the process of SIP message receiving, an internal message *CongestionCheck* is sent to a *bufferManager*. The *bufferManager* checks the queue length. If the queue length is less than congestion threshold, *bufferManager* allocates the memory for one queue and sends an internal message *NoCongestion*. If the queue length is greater than congestion threshold, *bufferManager* creates three queues for each priority class and returns an internal message *Congested*. When *NoCongestion* message is received, the SIP message is forwarded into the created queue. However, when *Congested* message is received, an appropriate priority is assigned to SIP message, which is further placed into corresponding queue.

In the process of SIP message forwarding, the *bufferManager* checks the mode of operation. In normal mode of operation, the packets are served by FIFO scheduling. In priority mode of operation, the PQ is used to forward packets that carry SIP message. The packets in the highest priority queue are forwarded first to the output interface. If there are no packets in the highest priority queue, the packets from the medium priority queue are

sent next. Failing that, the lowest priority queue is inspected and packets from that queue are forwarded to output interface.

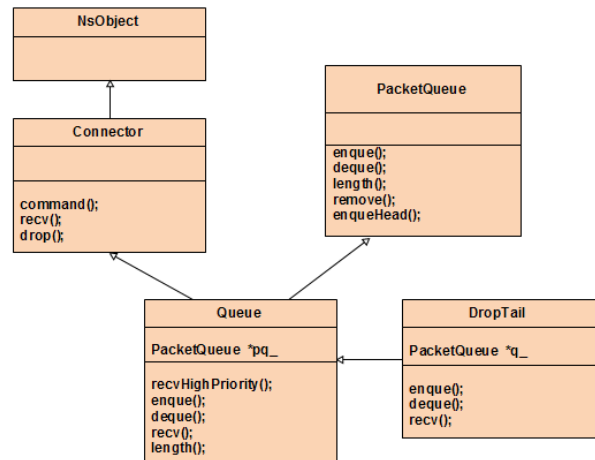


Fig. 1. Class diagram of the algorithm for SIP message classification and prioritization.

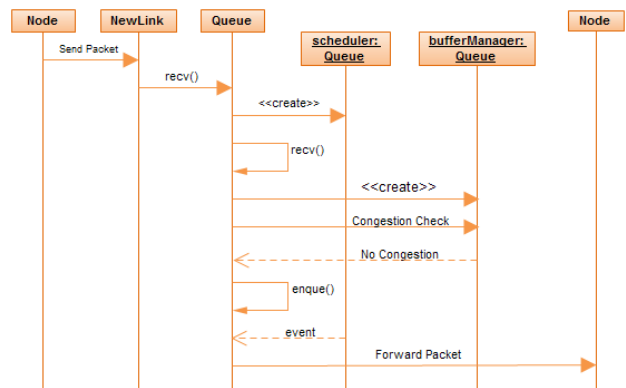


Fig. 2. Sequential diagram of the algorithm for SIP message classification and prioritization in normal mode of operation.

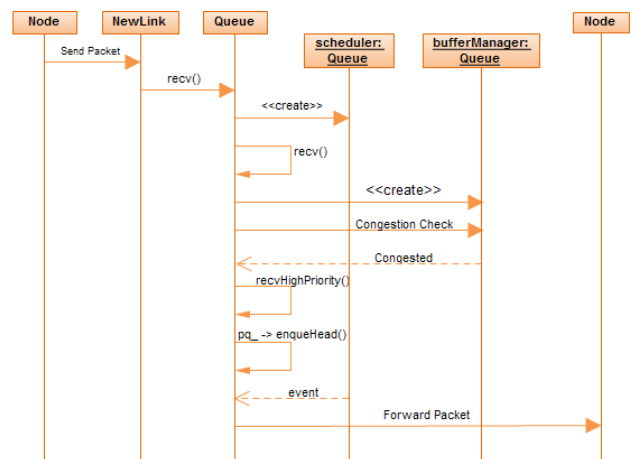


Fig. 3. Sequential diagram of the algorithm for SIP message classification and prioritization in priority mode of operation.

IV. SIMULATION OF SIP MESSAGE CLASSIFICATION AND PRIORITIZATION ALGORITHM

To verify the effectiveness of SIP message classification and prioritization algorithm, the simulation study is performed.

A. Simulation Setup and Environment

The simulations are based on two different scenarios. The simulation scenarios differ in SIP message scheduling algorithm. In Scenario 1, SIP messages are processed using conventional FIFO scheduling. In Scenario 2, SIP messages are processed using proposed algorithm, which is based on FIFO/PQ scheduling.

The simulations are performed on the simple network topology consisting of two boundary routers and one interior router. Boundary routers are connected to SIP domain consisting of SIP proxy servers and SIP user agents. Every SIP domain includes 150 SIP user agents that are used for generating background traffic. The links between routers are dimensioned to implement simple network configuration. The links capacities are configured to 1 Mbps. The delay of all links is set to 10 milliseconds and the queue lengths equal to 100 packets.

The network is loaded by different number of SIP messages generated and exchanged during the signaling procedures. Three types of SIP signaling procedures are considered: registration, session establishment and session termination. Different number of simultaneous SIP signaling procedures is initiated in order to generate the background traffic. This number is in the range from 0 to 900 for the purpose of measuring the SIP performance metrics.

B. Simulation Results and Discussion

The simulation results are analyzed in terms of Registration Request Delay (RRD), Session Request Delay (SRD), and Session Disconnect Delay (SDD). These SIP performance metrics are defined in Request for Comments (RFC) 6076. The simulation results for these SIP performance metrics are shown on Fig. 4. They are not discussed in comparison with results published in related works [2], [22] because the measurements are performed under different conditions and environment. However, the simulation results are analyzed with the aim to show the impact of the implemented algorithm on the SIP performance metrics.

Fig 4(a) shows the comparative analysis of RRD when FIFO and FIFO/PQPQ scheduling is used. It is noticed that the RRD increases in both scenarios with increasing the number of simultaneous SIP signaling procedures. The RRD values are identical in both scenarios until the network is loaded by 60 simultaneous SIP signaling procedures. For a larger number of simultaneous SIP signaling procedures, the RRD grows slower when FIFO/PQ scheduling is used. In this case, the RRD is approximately 150 milliseconds lower in overload conditions. This can be explained by that fact that the FIFO is used in normal load conditions, while the PQ is activated when the overload is detected. Moreover, the REGISTER requests are put in class 2 that has a medium priority in high-load and overload conditions.

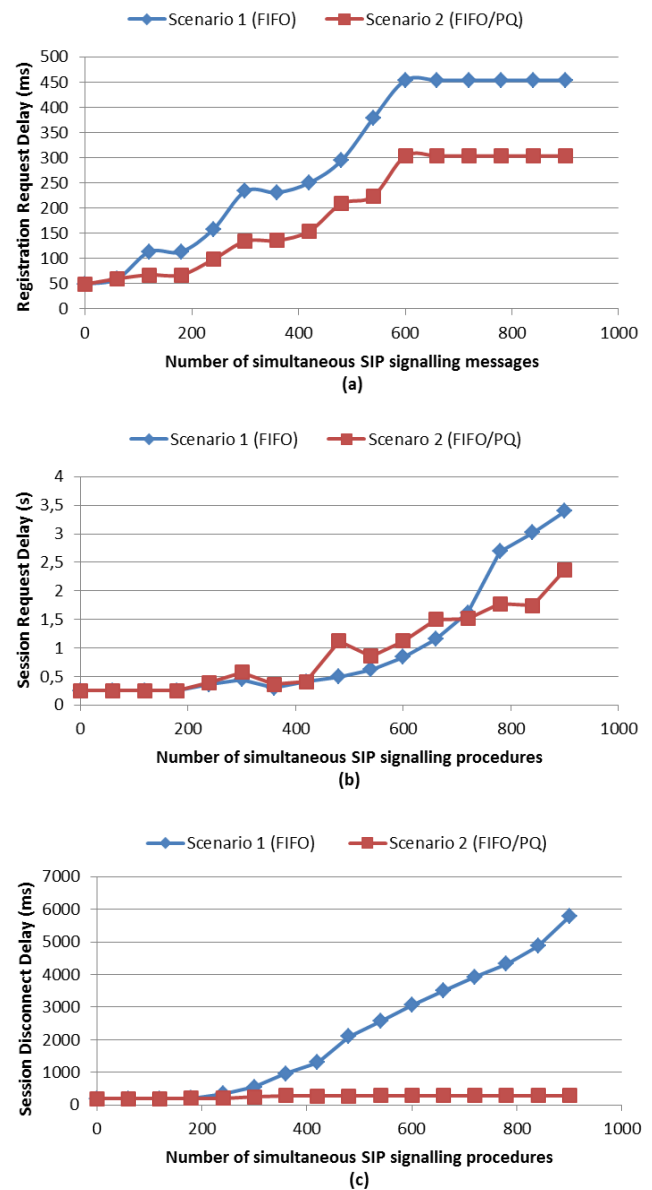


Fig. 4. Simulation results: (a) Registration Request Delay (RRD); (b) Session Request Delay (SRD); (c) Session Disconnect Delay (SDD).

Fig. 4(b) shows the comparison of SRD values when FIFO and FIFO/PQ scheduling is used. The SRD values are identical in both scenarios until the network is loaded by 180 simultaneous SIP signaling procedures. The increasing number of SIP signaling procedures leads to the faster deterioration of SRD if FIFO/PQ scheduling is used. This can be explained by the fact that the PQ scheduling is activated in high-load or overload conditions giving the lowest priority to INVITE messages. However, when network is loaded by more than 720 simultaneous SIP signaling procedures, the FIFO/PQ scheduling achieves the better SRD values. This is the consequence of giving a higher priority to SIP messages that terminate the existing sessions and thereby reduce the overload. Moreover, this leads to faster processing of INVITE requests and reducing of SRD values, consequently.

Fig. 4(c) compares the SDD values obtained when FIFO and FIFO/PQ scheduling are used. The SDD values are identical in both scenarios until the network is loaded by more than 120 simultaneous SIP signaling procedures. This can be explained by the fact that PQ is not activated in normal load conditions. Therefore, better results are achieved in high-load or overload conditions if FIFO/PQ scheduling is used. The insignificant increase of SDD value is noticed until the network is loaded by 480 simultaneous SIP signaling procedures. Then the SDD value becomes constant and equals to 282 milliseconds. This is the consequence of putting the BYE requests into the class 1, which is given the highest priority in high-load and overload conditions.

V. CONCLUSION

This paper presents the simulation-based optimization of SIP signaling procedures in IMS. The optimization is achieved by implementing SIP message classification and prioritization algorithm that is proposed in our previous work. This algorithm is based on three-priority classification of SIP messages. The main idea is to assign the highest priority to SIP messages that may serve to reduce the demand on the SIP proxy server. Giving a higher priority to SIP messages that terminate the existing sessions rather than those used to open the new sessions reduces the overload in IMS and thereby improves the overall QoS.

This SIP message classification and prioritization algorithm is implemented in ns-2 simulator. Its effectiveness is verified through the simulation-based optimization of SIP signaling procedures under different conditions. The simulation results are analyzed in terms of RRD, SRD, and SDD. These SIP performance metrics deteriorate with the increasing network load. However, better results are achieved if our algorithm is used instead of conventional approach. In this case, SIP signaling procedures have a shorter duration time, which improves the QoS. Therefore, the simulation results encourage the development and deployment of SIP message classification and prioritization algorithm in the experimental environment. This is going to be a starting point for our next research activities.

REFERENCES

- [1] J. Barakovic Husic, S. Barakovic, and H. Bajric. (2012, September) Evolution of Signaling Information Transmission. *ISRN Communications and Networking*. [Online]. 2012. Available: <http://www.hindawi.com/ismn/communications.networking/2012/70591/0/>.
- [2] J. Barakovic Husic, H. Bajric, E. Nekovic, S. Barakovic, "Basic Telephone SIP End-to-End Performance Metrics," in *Proc. 9th Int. Symp. On Telecommun. (BIHTEL 2012)*, Sarajevo, Oct. 2012, pp. 1-6.
- [3] V. S. Abhayawardhana and R. Babbage, "A Traffic Model for the IP Multimedia Subsystem (IMS)," in *Proc. 65th IEEE Veh. Technology Conf., (VTC Spring 2007)*, Dublin, Apr. 2007, pp. 783-787.
- [4] J. Barakovic Husic, E. Perenda, M. Hadzialic, S. Barakovic, "Performance Modelling and Optimization of IP Multimedia Subsystem," in *Proc. UKSim -AMSS 7th European Modelling Symposium on Mathematical Modelling and Computer Simulation (EMS2013)*, Manchester, Nov. 2013.
- [5] J. Baraković, H. Bajrić, M. Kos, S. Baraković, and A. Husić. (2011, September). Prioritizing Signaling Information Transmission in Next Generation Network. *J. of Comput. Networks and Commun.* [Online]. 2011. Available: <http://www.hindawi.com/journals/jcnc/2011/470264/>.
- [6] Y. Hong, C. Huang, and J. Yan, "A Comparative Study of SIP Overload Control Algorithms," in *Network and Traffic Engineering in Emerging Distributed Computing Applications*, IGI Global, 2012, pp. 1-20.
- [7] H. Jiang et al., "Design, Implementation, and Performance of a Load Balancer for SIP Server Clusters," *IEEE/ACM Trans. on Networking*, vol. 20, no. 4, pp. 1190-1202, Aug. 2012.
- [8] J. Yang, F. Huang, and S. Gou, "An Optimized Algorithm for Overload Control of SIP Signalling Network," in *Proc. 5th Int. Conf. on Wireless Commun., Networking and Mobile Computing (WiCom'09)*, Beijing, Sept. 2009, pp. 1-4.
- [9] M. Homayouni, M. Jahanbakhsh, V. Azhari, and A. Akbari, "Overload Control in SIP Servers: Evaluation and Improvement," in *Proc. 17th Int. Conf. on Telecommun.*, Doha, Apr. 2010, pp. 666-672.
- [10] Y. Wang, "SIP Overload Control: A Backpressure-based Approach," in *Proc. ACM SIGCOMM 2010 Conf.*, New Delhi, Sept. 2010, pp. 399-400.
- [11] Z. Chentouf, "SIP Overload Control Using Automatic Classification," in *Proc. Saudi Int. Electron., Commun., and Photonics Conf. (SIEPCP 2011)*, Riyadh, Apr. 2011, pp. 1-6.
- [12] C. Shen and H. Schulzrinne, "On TCP-based SIP Server Overload Control," in *Proc. Principles, Syst. and Applicat. of IP Telecommun. (IPTComm 2010)*, Munich, Aug. 2010, pp. 71-83.
- [13] V. Hilt and I. Widjaja, "Controlling Overload in Networks of SIP Servers," in *Proc. IEEE Int. Conf. on Network Protocols (ICNP 2008)*, Orlando, Oct. 2008, pp. 83-93.
- [14] A. Acharya, X. Wang, and C. Wright, "A programmable message classification engine for session initiation protocol (SIP)," in *Proc. 3rd ACM/IEEE Symp. on Architecture for Networking and Commun. Syst. (ANCS'07)*, Orlando, 2007, pp. 185-194.
- [15] M. Cortes, A. Walid, and N. Nithi, "Improving Performance of IMS Networks with Intelligent Message Processing," in *Proc. Int. Conf. on Intell. Networks (ICIN 2007)*, Bordeaux, Oct. 2007.
- [16] Z. S. Khan et al., "Ordering and Organization of SIP Requests within IMS Core by Using Load Awareness and Priority," *Int. J. of Recent Trends in Eng.*, vol. 2, pp. 24-26, 2009.
- [17] W. Zhu, A. K. A. Hamid, Y. Kawahara, T. Asami, and Y. Murata, "Automatic Originator Regulation of IMS Multiple Traffic by Stateless Signaling Prioritization," in *Proc. Global Commun. Conf. (GLOBECOM 2012)*, Anaheim, Dec. 2012, pp. 2856-2851.
- [18] L. Begg, W. Liu, K. Pawlikowski, S. Perera, and H. Sirisena, "Survey of Simulators of Next Generation Networks for Studying Service Availability and Resilience," Univ. of Canterbury, Christchurch, New Zealand, Tech. Rep. TR-COSC, Feb. 2006.
- [19] T. Peter, G. Marek, F. Tomas, G. Radoslav, and L. Michal, "Simulation of IMS Using Current Simulators," in *Proc. 50th Int. Symp. ELMAR 2008*, Zadar, Sept. 2008, pp. 545-548.
- [20] Network Simulator version 2. [Online]. Available: http://nslam.isi.edu/nslam/index.php/User_Information.
- [21] R. Prior. ns-2 network simulator extensions. [Online]. Available: <http://www.dcc.fc.up.pt/~rprior/ns/>.
- [22] M. Voznak and J. Rozhon, "Methodology for SIP Infrastructure Performance Testing," *WSEAS Transactions on Computers*, vol. 9, no. 9, pp. 1012-1021, Sept. 2010.