

Service Continuity Support in Self-Organizing IMS Networks

Christian Makaya, Ashutosh Dutta, Subir Das,
Dana Chee, F. Joe Lin
Telcordia Technologies, Inc.
Piscataway, NJ, USA

Satoshi Komorita, Hidetoshi Yokota
KDDI R&D Labs
Saitama, Japan

Henning Schulzrinne, Columbia University, NY, USA

Abstract— With the increasing interest in deploying 4G/LTE networks, IMS has a potential to be deployed in a wide scale in order to support mobile Internet and value-added services over next-generation networks. Moreover, the effort to create an operator-controlled signaling infrastructure using IP-based protocols has resulted in a large number of functional components and interactions between core networks elements. Thus, the carriers are trying to explore alternative ways to deploy IMS that will allow them to manage their network in a cost effective manner while offering rich communications services. One of such approaches is self-organization of IMS (SOIMS). The self-organizing IMS can enable the IMS functional components to adapt dynamically based on the features like network load, number of users, node failures and available system resources. This paper proposes different mechanisms to handle self-organizing IMS with usage of load balancing paradigm. The proposed solution enables topology hiding, IMS nodes failure recovery, session continuity support, and IMS scalability.

Keywords—Self-organizing IMS, load balancing, session continuity, scalability.

I. INTRODUCTION

The IMS (IP Multimedia Subsystem) architecture was designed to provide IP-based voice and multimedia services independent of access network technologies. Furthermore, IMS is a unified platform that enables convergence of wired and mobile networks. However, IMS is a complex architecture and does not scale very well. Several functional elements have been defined but how these functional elements should be combined or implemented are not specified. For these reasons, it is important to reduce complexity and cost of IMS. The self-organizing networks (SON) concept has been widely studied for example in mobile ad hoc networks. SON can be defined as a set of functionalities that enable the automation of network operations by minimizing manual intervention or configuration. This way, SON could help network operators to reduce operations expense (OPEX) by automating their network operation. The SON concept can be subdivided in three main topics: self-configuring, self-operating and self-optimization. Furthermore, SON can be implemented based on centralized method or distributed method.

The *Self-Organizing IMS* (SOIMS) is defined as the IMS environment that can dynamically self-configure, self-optimize and self-recover its components and roles due to

different network situation such as nodes failure, capacity or resources limit [1]. It can be based on ones of the two models: centralized or distributed. In the centralized model, the *Master Node* or controller should maintain a database with operator policy and state information for all IMS nodes under its control. The database is updated when a new node notifies the Master Node about its capabilities, when a specific IMS role has been assigned to this node, and when a node changes roles due to higher load or failure.

Two approaches of SOIMS can be considered: UE (User Equipment) involved and non-involved methods. In UE involved method, the UE is aware of IMS node changes in the core IMS network. Detailed mechanisms to handle this scenario are described in [1]. With the UE involved method, when any change happens in IMS networks, the UEs are notified. This will induce a higher signaling traffic and scalability issues when the number of UEs increases [13]. In wireless environment where radio resources are limited, these issues are critical. To avoid these issues, change in IMS networks should be transparent to UEs. In other words, the UEs should not be notified about IMS node failure or reconfiguration and it implies that the P-CSCFs IP address appears to UEs unchanged for example when P-CSCF fails. This approach is called UE non-involved. In case of UE non-involved method the UE is completely unaware of the IMS node changes in the core IMS network. With this method, no additional message will be sent to UEs over the access link and UEs also don't take action after re-configuration of IMS network. This method is the main purpose and contribution of this paper.

To allow support of the UE non-involved method, a solution based on *Load Balancing* (LB) concept is proposed in this paper. The load balancing is a technique used to distribute network load across different network components in order to get optimal resource utilization, scalability, high availability and flexibility [9]. The proposed solution satisfies these LB features while handling support of session continuity in self-organizing IMS networks, topology hiding, and distributed IMS functions in an efficient manner.

The rest of the paper is organized as follows. Section II provides an overview of related work. Section III presents the self-organizing IMS architecture (SOIMS). More details on SOIMS can be found in [1]. Section IV describes the load balancing concept. Section V describes the proposed

approach for the integration of SOIMS and LB paradigm. Finally, Section VI concludes the paper.

II. RELATED WORK

With the emerging 3GPP LTE (Long Term Evolution) technology, SON is envisioned as a new model for next-generation operations and business support systems (OSS/BSS). The 3GPP [2] and Next Generation Mobile Networks (NGMN) Alliance [3] have standardized a set of capabilities known as SON to improve operation and maintenance (O&M). Since IMS is envisioned as a main component for the deployment of LTE or 4G wireless networks, for example, in order to provide real-time applications and rich communications services (RCS), it is important to extend SON concept to the core IMS networks rather than only to the LTE RAN and EPC (Evolved Packet Core). The concept of self-organizing IMS networks is relatively a new topic and has not been widely studied. It is important to understand that the self-organizing IMS is different than P2PSIP concept [1] where significant research has been done.

Bessis [5] describes performance analysis and benefits of running multiple SIP servers on the same host. That paper shows how to design the IMS networks in order to maximize IMS server co-location and explains which types of SIP calls can benefit from the co-location of IMS servers. Fabini et al. [6] describe an IMS configuration with respect to architecture and QoS aspects. It demonstrates the feasibility of an IMS system implementation within a single virtual device (all-in-one). In [7] a distributed IMS architecture has been proposed by representing network functional elements in Distributed Hash Tables (DHT) overlay networks. The main focus was to distribute S-CSCF functionalities by using an overlay network where these functionalities are merged in one node (called IMS DHT). A common issue with DHT overlay networks is related to the number of operations or message in DHT to retrieve information or data.

However, none of these papers have looked into methods of supporting self-organizing networking and SIP load balancing capabilities of IMS networks. Furthermore, these papers did not take into account reconfiguration of the functions or roles, for example due to nodes failure or for load balancing. The work presented in [1], namely UE involved method, is the first truly self-organization concept applied to IMS networks without using overlay networks.

This paper works around the concept of self-organization for IMS networks and the main focus is the UE non involved case. In particular, this paper develops the mechanisms and framework to facilitate self-organizing networks capabilities such as self-configuration, self-optimization and self-recovery or self-healing at the service layer with minimal manual intervention. This offers a significant savings on OPEX and CAPEX. The papers describes different functions that are needed to support the SON features of IMS that can be applicable to 4G/LTE networks or any communications technology using IMS as part of the whole system environment.

III. SELF-ORGANIZING IMS CONCEPTS

For sake of simplicity, we consider a centralized approach for SON requiring an anchor entity called Master Node, which maintains the operator policy and state information for all IMS nodes under its control. The Master Node database is dynamically updated when for example a new IMS node is up and notifies its capabilities to the Master Node, the IMS role assignment has been changed. To support self-organizing IMS (SOIMS) networks, negotiation or capabilities exchange and event notification protocols have been designed in a previous work and described in [1]. The major features of SOIMS are: IMS nodes discovery, IMS role request and assignment, and interaction protocol between IMS node and Master Node.

When the IMS role changes, the Master Node must notify the appropriate IMS nodes (e.g., HSS and S-CSCF when P-CSCF role change) and provide them the information about the new network's configuration [1]. To allow SOIMS networks, P-CSCF should be able to receive notification of role change from the Master Node and retrieve updated information (e.g., get information of the new S-CSCF) from the HSS. Similarly, the S-CSCF should be able to receive notification of role change from the Master Node, store UE's subscription and profile information, update HSS with the UE information after the role change.

IV. LOAD BALANCING CONCEPT

The load balancing concept and technology are widely used in order to distribute traffic among several back-end network components (e.g., servers, proxies). In fact, the Load Balancer (LB) is a device acting as a front-end node of different machines and it intercepts traffic destined to a given site (domain) and redirects the traffic to these network components. The main idea of LB is that the real servers/proxies appear as a virtual server on a single IP address, and the clients or users interact with the domain as if there were a single server without knowing all real servers.

Load balancing concept offers flexibility, high availability and scalability for network management. From the outside, the network domain is reachable through a virtual IP address (VIP). The LB spreads the traffic amongst several machines by using different methods and metrics, e.g., round-robin, least connection, source or destination address hashing, etc. With the usage of LB, any changes or failure on back-end machines are transparent to the client requesting services.

Several solutions for load balancing are available. However, these solutions are designed for non real-time applications such as web servers, DNS, and database. Session persistence with these solutions is based on cookies, client's IP address and port number. Moreover, these solutions focus mainly at IP layer. With the increasing penetration rate of real-time application such as VoIP, it is necessary to have load balancing support at SIP layer since IP load balancing cannot be used for SIP-based application. In fact, with IP load balancing technology, the LB is not able to add SIP headers such as Via, Route, and Record-Route in

the SIP messages. This information is required in SIP message to allow adequate routing and session activation.

V. PROPOSED SOLUTION

A. SIP Load Balancing

As described above, the UE non-involved method requires usage of the Load Balancer (LB) in order to hide any change in the core IMS network. In the proposed solution, the LB appears as a virtual P-CSCF to the UEs. In other words, from the UE's perspective, the LB is the P-CSCF and the UE sends all request or SIP messages to the LB's virtual IP address (VIP). Moreover, rather than establishing an IPSec security association (SA) between UEs and P-CSCF as specified by IMS standard, the IPSec SA is established between UEs and LB.

Figure 1 illustrates our proposed SIP-based load balancing framework for IMS networks. When the LB receives SIP messages from the UE, it forwards these messages to the selected P-CSCF after adding its own entries in the Via and Record-Route headers. The selection of P-CSCF is done by a scheduling algorithm defined in LB and by using information provided by the Master Node or HSS. There is no direct communication between UEs and the real P-CSCF since all communications are handled by the LB. In order to correctly route the SIP packets to P-CSCFs and maintain the session persistence, the LB needs to intercept the SIP packets and modify the headers accordingly. For these reasons, the LB works as a SIP Proxy and processes SIP messages received from/to UE and the real IMS node (e.g., P-CSCF). Having a SIP-aware LB, SIP headers (i.e., Via, Record-Route, Path, and Route) in SIP messages will be set appropriately by the LB.

B. Registration Procedure

When a UE initiates registration procedure, it sends a SIP REGISTER request to the LB. Since LB is acting as an outbound SIP proxy, it processes this message and adds itself in (the topmost) Via and Record-Route headers of SIP REGISTER message before forwarding the message to the selected P-CSCF. By adding its information to these headers, the LB will receive the response to the request. In the opposite direction, the LB removes these headers which contain its information before to send the message to the UE. The selection of P-CSCF is based on different scheduling algorithm such as: hash over Call-ID, hash over From URI, Round-Robin, etc. In our prototype, the selection and session persistence are based on two methods: hash over Call-ID and hash over From URI since they are more consistent and provide better session persistence. Figure 2 shows the SIP registration procedure. For sake of simplicity, only main registration steps are illustrated. For example, interaction with I-CSCF has been ignored.

The LB gets list of P-CSCF from the Master Node. Since the Master Node has knowledge of IMS nodes, getting the list of P-CSCFs from Master Node allows

reduction of signaling due to IMS node monitoring by LB. The remaining registration procedure process follows IMS specification. The LB retrieves security association parameters (SPIs), integrity (IK) and confidentiality keys (CK) from the P-CSCF.

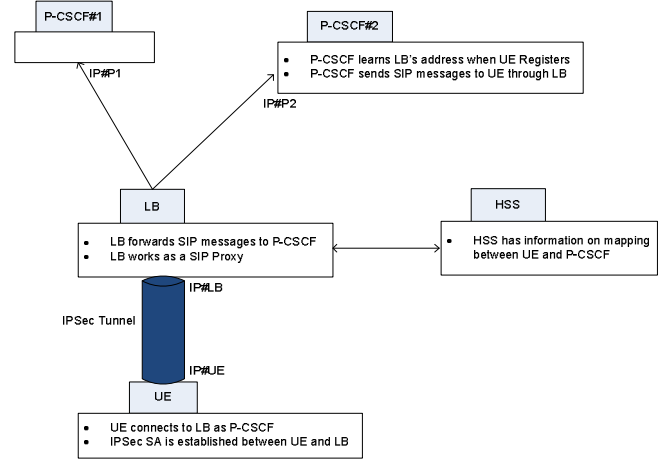


Figure 1: SIP Load Balancing in IMS architecture.

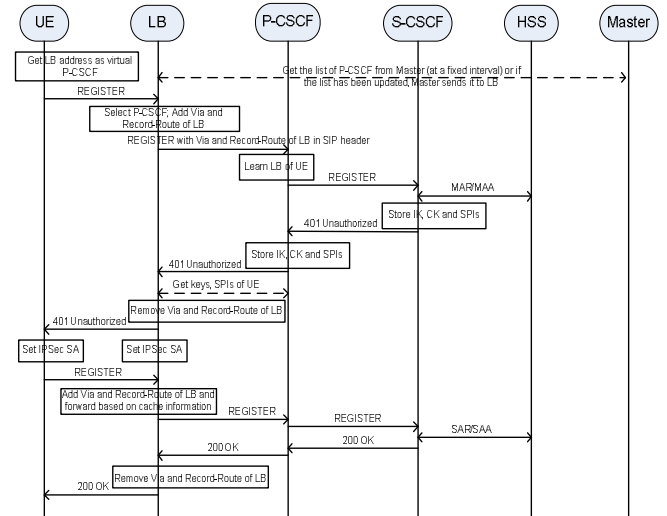


Figure 2: SIP Registration Procedure.

C. Session Setup Procedure

After a successful registration, a call or session setup can be initiated by the UE. The UE pre-loads the stored information of the outbound proxy (i.e., LB) into the Route header of SIP INVITE message before sending it out. The LB uses the same procedure as for REGISTER message to select the P-CSCF, removes its own entry from Route header, adds its own entry in Via and Record-Route headers and forwards the INVITE message to the selected P-CSCF. By adding its own entry in Record-Route header, all subsequent requests within the established SIP dialog will be routed through the LB. In order to guarantee session persistence, LB can use cached information about the previously selected P-CSCF for a given UE.

D. Load Balancer Failure Support

Since the LB is a main anchor point, it might become a single point of failure for the SOIMS architecture. To avoid the entire system from being out of service due to the LB failure, a backup of LB might be deployed (i.e., redundancy). Redundancy is always a desired feature in real deployment scenario. The heartbeat technique or Virtual Router Redundancy Protocol (VRRP) [3] can be used between the primary and the backup LB and they communicate periodically to inform each other they still alive. The primary and backup LBs are synchronized in order to share the ongoing session information (e.g., SIP dialog, list of P-CSCFs).

E. Session Continuity with P-CSCF Failover Support

Service or session continuity is one of the main requirements for real-time application, for example when the UE roams from one access network to another or in presence of IMS node failure. In this section, we will describe session continuity when IMS role change event (e.g., IMS node failure, role assignment) occurs during ongoing session. Let us assume that the P-CSCF's (e.g., P-CSCF#1 in Fig. 3) role changes due to failure, the LB will be notified by the Master Node about network configuration change or since LB gets list of P-CSCFs from the Master Node, it will discover role assignment change or failure.

When the Master Node detects failure of P-CSCF#1, it notifies the S-CSCF about this event with information about this new P-CSCF. Upon this notification, the S-CSCF can retrieve information of the new P-CSCF (e.g., P-CSCF#2) from the HSS. At the same time, the S-CSCF updates registration status (e.g., association and mapping) of LB and UE through P-CSCF#2. Then P-CSCF#2 can restore registration information and update mapping between LB and UE for subsequent SIP messages.

After restoration of registration information, the S-CSCF sends a message to P-CSCF#2 with information about media negotiation (i.e., SDP), new and old SIP Route. This information exchange allows restoration of the ongoing SIP session state in the S-CSCF and P-CSCF#2, and reconfiguration of IMS core network. When P-CSCF#2 completes the update of session information, it informs the LB and HSS about the new SIP Route. The old and new SIP Routes are stored in the LB and HSS to allow mapping of previously established SIP dialog with P-CSCF#2.

All of these changes are transparent to the UEs. In fact, the LB hides the change and reconfiguration of IMS core network and UEs have no direct communication with IMS components. Any subsequent SIP message sent by the UE will be set with the old SIP Route. It is LB's responsibility to change the old SIP Route (e.g., Service-Route) in SIP message to the new SIP Route before to forward this message to the first IMS entity (i.e., P-CSCF). Since any change in the core IMS network is transparent to UEs, with the proposed solution, there is no need for UEs to subscribe for events notification leading to minimal signaling overhead and network resources usage.

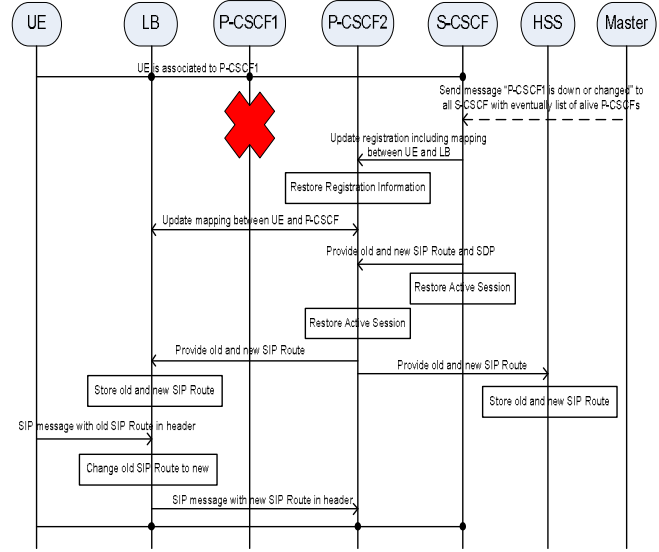


Figure 3: Session Continuity during P-CSCF Failure.

F. Session Continuity with S-CSCF Failover Support

The S-CSCF failover support is similar to P-CSCF failover as illustrated in Fig. 4. When the Master Node detects the failure of S-CSCF#1, it notifies all active P-CSCFs about this event and provides a list of available S-CSCFs. The P-CSCF updates registration information. When the new S-CSCF (e.g., S-CSCF#2 in Fig. 4) receives the notification from the P-CSCF, it restores the registration information associated to the UEs registered with the failed S-CSCF.

When S-CSCF#1 fails or changes, the P-CSCF transfers ongoing session information to S-CSCF#2 to allow session restoration and continuity. The P-CSCF notifies also the LB and HSS about this event by sending information about old and new SIP Route. The UE will continue to send any SIP message with the old SIP Route information to the LB. The LB will perform all required mapping with the new SIP Route for the ongoing SIP dialog.

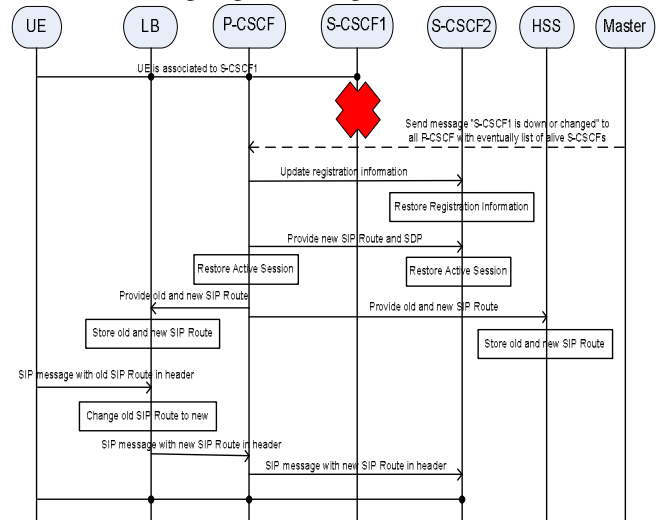


Figure 4: Session Continuity during S-CSCF Failure.

VI. PERFORMANCE EVALUATION

Figure 5 shows the test-bed architecture used for prototyping the proposed solution. IMS node in Fig. 5 refers to an IMS component (e.g., P-CSCF, I-CSCF, S-CSCF or HSS) running in one or different physical nodes. All components in the test-bed are Linux-based and there are two access networks and one home core network. The core network is equipped with all SIP-based IMS components: HSS, I-CSCF, S-CSCF and P-CSCF. The master node is located in the router that also acts as the DNS and DHCP server. The edge routers act as 3GPP's PDIF (Policy Decision and Information Function) [2] and DHCP relay agent. For LB, Open SIP Server (OpenSIPS) [10] has been extended with a new module in order to support the UE non-involved method. This module implements the scheduling algorithm and session persistence approaches based on hash over Call-ID and From URI. SIP Communicator [11], which has been extended to support 3GPP IMS requirements, is used as the IMS SIP client. For SIP traffic overload, SIPp and IMS Bench SIPp [12] have been used for testing.

Performance results for IMS network reconfiguration indicate that it takes less than one second to re-configure the respective IMS components functionalities by the Master Node while being transparent to UEs. This reconfiguration delay plays an important role in order to guarantee session continuity, IMS network availability, and call tracking (for billing purpose). During this reconfiguration delay, ongoing session is not interrupted. Moreover, since UEs are not aware of any reconfiguration event, there is not signaling over the wireless link due to IMS network reconfiguration. The proposed approach guarantees a 100% success rate for session completion ratio and effective registration attempt metrics [14].

VII. CONCLUSION

Self-organizing networks (SON) has been standardized by 3GPP and NGMN Alliance and is an important step to improve operation and maintenance (O&M). Since IMS is considered as a main component for the real deployment of 4G/LTE networks, self-organization of IMS networks must be considered also. In fact, a policy-based configuration and reconfiguration of IMS components will help operators to reduce the cost and complexities of current IMS networks. This paper has proposed a self-organizing, dynamic adaptation, and context aware networks capabilities for IMS networks. The proposed approach allows to hide any change in the core IMS networks to end-users (i.e., user equipment doesn't participate in the IMS network reconfiguration), minimize IMS networks reconfiguration delay, and reduces signaling overhead. The proposed solution has been prototyped and results show that session continuity can be guaranteed when IMS network topology changes due to

network entities failure or IMS role re-assignment. As a result, the IMS networks efficiency and scalability has been improved. Although, in this paper we have focused on P-CSCF and S-CSCF, the proposed solution can be applicable to other IMS core network elements such as I-CSCF, Application Servers, and Session Border Controller (SBC).

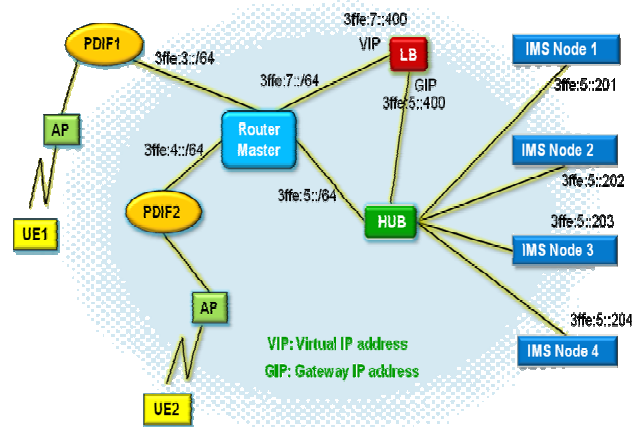


Figure 5: SOIMS-LB Testbed Architecture.

REFERENCES

- [1] A. Dutta *et al.*, "Self Organizing IP Multimedia Subsystem," In Proc. of 3rd IEEE Int'l Conf. on Internet Multimedia Systems Architecture and Application (IMSAA'09), Bangalore, India, Dec. 2009.
- [2] Third Generation Partnership Project (3GPP), www.3gpp.org
- [3] NGMN Alliance: www.ngmn.org
- [4] R. Hinden (Ed.): Virtual Router Redundancy Protocol (VRRP), IETF RFC 3768, April 2004.
- [5] T. Bessis, "Improving Performance and Reliability of IMS Network by Co-Locating IMS Servers," Bell Labs Technical Journal, Vol. 10, No. 4, pp. 167-178, 2006.
- [6] J. Fabini *et al.*, "IMS in a Bottle, Initial Experiences from an OpenSER-based Prototype Implementation of the 3GPP IP Multimedia Subsystem," in Proc. of the Int'l Conf. on Mobile Business (ICMB'06), Copenhagen, June 2006.
- [7] M. Matuszewski, M. Garcia-Martin, "A Distributed IP Multimedia Subsystem (IMS)," IEEE Int'l Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM'07), pp. 1-8, 2007.
- [8] M. Poikselka and G. Mayer, "The IMS: IP Multimedia Concepts and Services", 3rd Edition, ISBN: 978-0-470-72196-4, March 2009.
- [9] T. Bourke, "Server Load Balancing", O'Reilly, August 2001.
- [10] OpenSIPS: www.opensips.org
- [11] SIP Communicator: www.sip-communicator.org
- [12] SIPp: <http://sipp.sourceforge.net/index.html>
- [13] S. Komorita *et al.*: "Performance Evaluation of Service Continuity in Self-Organizing IP Multimedia Subsystem Networks", Proc. of WASET International Conference on Communication Technology (ICCT'10), pp. 118-123, Venice, Italy, Nov. 24-26, 2010.
- [14] D. Malas and A. Morton: "Basic Telephony SIP End-to-End Performance Metrics", IETF Internet Draft draft-ietf-pmol-sip-perf-metrics-05.txt, Sept. 2010.