
Self-Organizing IP Multimedia Subsystem

Ashutosh Dutta¹, Christian Makaya¹, Subir Das¹, Dana Chee¹, Joe Lin¹ Satoshi Komorita², Tsunehiko Chiba², Hidetoshi Yokota² and Henning Schulzrinne³

¹*Telcordia Technologies Inc., 1 Telcordia Drive, Piscataway, NJ 08854, USA;*
e-mail: ashutosh.dutta@ieee.org

²*KDDI R&D Laboratories, Japan*

³*Columbia University, New York, NY 10027, USA*

Abstract

While there have been tremendous efforts to develop the architecture and protocols to support advanced Internet-based services over 3G and 4G networks, IMS is far from being deployed in wide scale. Effort to create an operator controlled signaling infrastructure using IP-based protocols has resulted in a large number of functional components and interactions among those components. 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 the value-added services. One of such approaches is self-organization of IMS. The self-organizing IMS can enable the IMS functional components and corresponding nodes to adapt them dynamically based on the features like network load, number of users and available system resources. This chapter introduces such a self-organizing and adaptive IMS architecture, describes the advanced functions and demonstrates the initial results from the prototype test-bed. In particular, we show how all IMS functional components can be merged and split among different nodes as the network demand and environment change without disrupting the ongoing sessions or calls. Although it is too early to conclude the effectiveness of self-organizing IMS, initial results

Anand R. Prasad et al. (Eds.), Advances in Next Generation Services and Service Architectures, 123–147.

© 2011 River Publishers. All rights reserved.

are encouraging and it may provide additional incentives to the operators for network evolution.

Keywords: IMS, self-organization, load balancing, adaptivity, context aware.

7.1 Introduction

The current version of IMS (IP Multimedia Subsystem) has several shortcomings that may act as deterrent for wider deployment. Depending upon the architecture there may be a need to have one-to-one mapping between functional component and a physical node where each node is equipped only to perform a certain function. This would require the deployment of a large number of network nodes which in turn will make the network management and operation difficult, in particular, the environment where dynamic adaptation is required based on the functionalities, processing, network loads and node failures. One way to circumvent these issues is to allocate redundant network resources, but that will not help the service providers to achieve the goal of offering services at a lower cost. Alternatively, one can define new techniques that support dynamic adaptation with the nodes merging and splitting the IMS functional components. By node merging, we mean all the IMS functional components merge and operate on a single physical node; whereas, by node splitting we mean the functional components get distributed across different physical nodes. In its current form, IMS architecture and protocols do not have the mechanisms that can easily help IMS functionalities (e.g., P-CSCF, I-CSCF, S-CSCF) to easily migrate from one node to another. Thus, it needs additional features and mechanisms to support self-organizing capability. These additional features will enable the IMS network to adapt and distribute the functionalities based on the network condition and operational environment. Currently, IEEE NGSON (Next Generation Service Overlay Network) [9], ATIS SON [2] and ITU-T SG13 [18] are working on defining the requirements that would enable the IMS to support several features such as self-organization, context awareness, and peer-to-peer services and will help the service providers to better manage their IMS network while reducing the cost of operations.

This chapter proposes and develops methodologies that are needed to support several self-organizing features of IMS. These methodologies can provide additional flexibilities to the current IMS architecture and help service providers to better manage the IMS components in a dynamic en-

vironment where nodes or links are prone to failure and the system grows or sinks with the demand. This chapter also describes the laboratory prototype test-bed and provides some initial results to demonstrate the feasibility of such a system.

The rest of the chapter is organized as follows. We introduce related work and standards efforts in Section 7.2. Section 7.3 introduces the basic IMS architecture and then describes the self-organizing features and the associated functions that are needed to support this architecture. Section 7.4 describes the operations of different function nodes in the self-organizing architecture. Section 7.5 describes two possible deployment scenarios and the call flows associated with these scenarios. Section 7.6 describes the test-bed prototype for self-organizing test-bed and the load balancer. Finally, Section 7.7 concludes the chapter.

7.2 Related Work

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 P2P-SIP concept [12] where significant research results are available. Bessis [3] 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 a minimal optimal IMS configuration with respect to architecture and QoS aspects. A virtual IMS test-bed (i.e., any IMS component is assigned its own virtual host), with different domains has been setup on one physical machine. It demonstrates the feasibility of an IMS system implementation within a single device (all-in-one). Matus et al. [13] propose a distributed IMS architecture by representing network functional elements in Distributed Hash Tables (DHT) overlay networks. The main focus was to distribute S-CSCF, I-CSCF and HSS functionalities by using an overlay network where these functionalities are merged in one node (called IMS DHT). Manzalini et al. [11] describe a platform to provide automatic and situation-aware communication services. However, none of these papers have looked into methods of supporting self-organizing capabilities of IMS. Furthermore, these papers did not take into account reconfiguration of the functions when, for example nodes fail or for load balancing reasons. Most recently, the Next Generation Service Overlay Network (NGSON) [9], a standards activity launched by the IEEE Communications Society in March

2008, is specifying a framework for service overlay networks. This framework will include context-aware, dynamically adaptive, and self-organizing networking capabilities including advanced routing and forwarding schemes that are independent of underlying transport networks.

This chapter works around the basic concept of NGSON that is based on self-organizing IMS and has drawn some inspiration from [13]. In particular, this chapter develops the mechanisms and framework to facilitate self-organizing capabilities such as self-configuration, self-optimization and self-recovery at the service layer.

7.3 IMS Overview

At a functional level, IMS architecture primarily consists of several signaling entities such as P-CSCF (Proxy-Call Session Control Function), I-CSCF (Interrogating CSCF), S-CSCF (Serving CSCF), and HSS (Home Subscriber Server). Figure 7.1 shows one such functional architecture where P-CSCFs are distributed across the visited networks. In this specific architecture, there are four networks labeled as Home Network, Visited Network 1, Visited Network 2 and Visited Network 3. Two of the networks have CDMA2000 access while the third network supports either WiFi or WiMAX. Details of this architecture can be found in [5]. We provide a description of some of the functional components.

MN: MN is the mobile node that moves across the networks.

DHCP Servers: DHCP servers are configuration agents and help the mobiles with the configuration of network layer identifiers such as IP address, address of P-CSCF, and DNS servers. Each visited network may be equipped with a DHCP server.

HA: HA serves as the anchor point for a mobile in mobile IP environment and maintains the mapping between the mobile's home address and the new care-of-address that the mobile obtains in each network. This care-of-address is obtained either from a stateful DHCP server, or from a foreign agent or by means of stateless auto-configuration.

S-CSCF: The S-CSCF is the central node of the signaling plane. It is a SIP server and performs session control functions. It is always located in the home network. It can either query the HSS or the DNS server to locate the appropriate P-CSCF for outgoing intra-domain calls and the appropriate I-CSCF for inter-domain calls.

P-CSCF: The P-CSCF behaves as a SIP proxy and is the first outbound proxy for a mobile in the visited network. The P-CSCF routes REGISTER

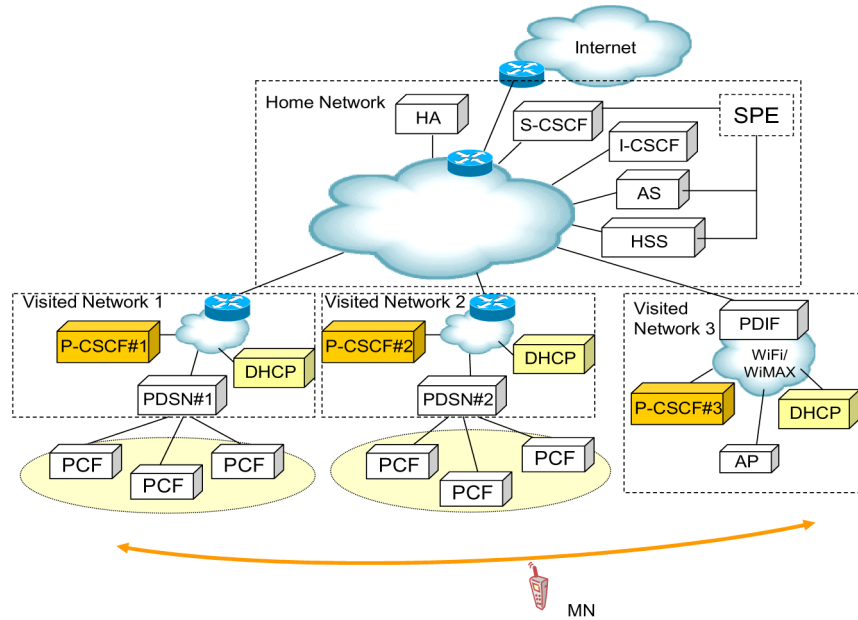


Figure 7.1 A distributed IMS architecture.

requests to the I-CSCF and caches the S-CSCF address so that it can route the rest of SIP signals directly to the S-CSCF.

I-CSCF: The I-CSCF is another SIP proxy that provides forwarding of messages to the correct S-CSCF, through HSS or DNS look-ups. I-CSCF also acts as an entity that hides information for inter-domain calls.

HSS: The HSS stores information about the subscribers, their addresses, and their services such as user account, contact URI of the user, address of P-CSCF for each mobile, E.164 number. The HSS is located in the home network and communicates with the S-CSCF.

Service Provisioning Environment (SPE): The SPE is implemented in combination with the HSS, but logically it is a separate component. It provides a mechanism to view the services that are deployed on the AS and system administrators can use it to provision services automatically for each user.

Application Server (AS): The AS sends messages to the HSS defining the applications deployed on the AS, along with the parameters needed to configure an instance of the service. This description is used by the SPE

to generate the system administrators screens for service configuration and provisioning.

IMS Node: Any physical node within an IMS infrastructure that is capable of running any of the IMS functions is defined as an IMS node.

7.3.1 Self-organizing IMS

Figure 7.2 shows a specific deployment scenario for the self-organizing IMS where nodes and links may fail resulting in IMS component reconfiguration. Figure 7.3 shows how an IMS node in Figure 7.2, termed as basic self-organizing node capable of running several IMS functions, can adapt itself based on the server load and network conditions. It is assumed that all the nodes are capable of taking on the roles of the IMS components, such as P-CSCF, S-CSCF, I-CSCF. These nodes can run one or multiple instances of different IMS functions. In this section, we introduce the functional behavior and interactions of different logical entities that constitute the self-organizing IMS architecture.

Self-organizing IMS can be based on one of the two modes: centralized or distributed. In the centralized mode, there is a master node that maintains a database with operator policy and state information for all nodes under its control. For example, this master node can be the HSS (Home Subscriber Server) as specified in [15]. The reason behind this choice could be that HSS is the master database of the cellular network and the availability and reliability of such database is much higher than ordinary nodes.

The master node database is updated when (i) a new node notifies the master node about its capabilities, (ii) a specific IMS role has been assigned to the node, and (iii) a node changes roles due to overloading or failure. An efficient policy-based mechanism should be defined in the master node to assign the functionalities or roles.

In the distributed mode of operation, the new node announces its presence through a multicast message. Existing IMS nodes reply to this message if they need to transfer some of their current functionalities. This may result in conflicts but it can be resolved by a simple rule such as, first in first out (FIFO) based on response message. If a node receives no responses, it should assume that there exists no other node and it takes over all roles. In either the centralized or distributed mode, negotiation or capabilities exchange and event notification protocols should be defined. In fact, the decision to assign a role to a new node will be based on its capabilities (e.g., CPU, load, processing power, memory). We focus on the centralized mode in this chapter.

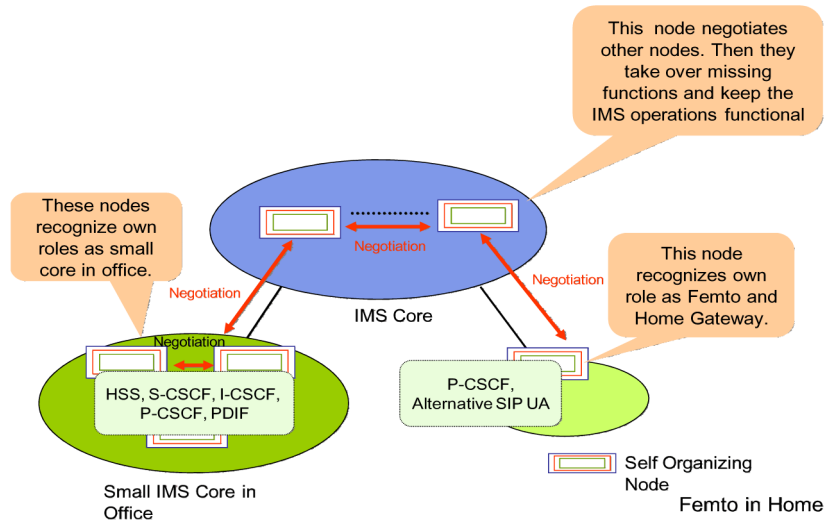


Figure 7.2 Applicability of self-organizing IMS.

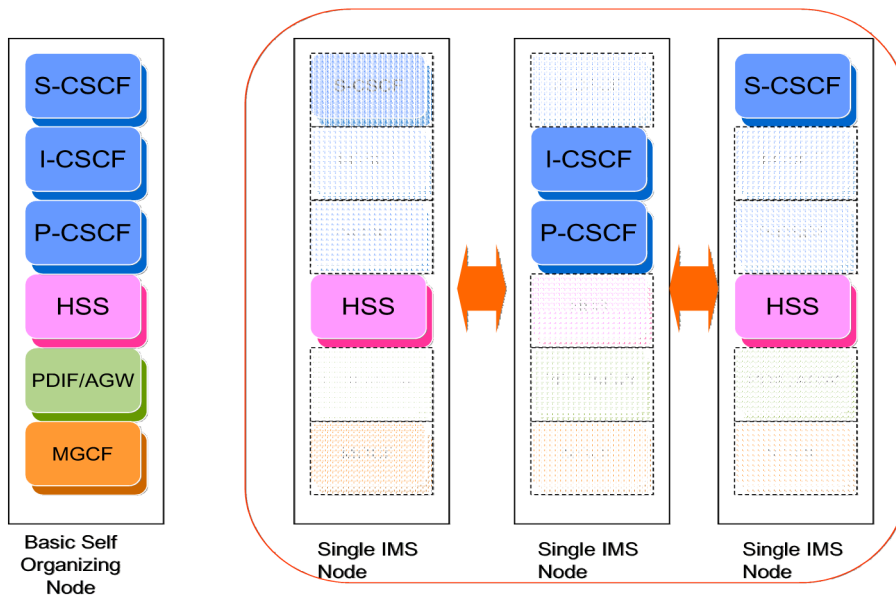


Figure 7.3 Self-configuration of IMS functions.

Some of these functionalities can be easily applied to a distributed mode as well. We next describe the important functions of self-organizing IMS.

7.3.2 Nodes Discovery Function

There are two kinds of node discovery function in Self-Organizing IMS: (1) a master node discovers the IMS nodes and their capability and (2) mobile node discovers the IMS node for registration. The master node discovers the other IMS nodes and their capability via a negotiation protocol. Mobile node on the other hand can use standardized mechanisms such as DHCP options [17] or DNS query [7] for IMS node discovery.

7.3.3 Role Request and Assignment

When an IMS node comes online, it requests a role or function from the master node by sending its capability. The capability may include IMS functionality such as P-CSCF, S-CSCF or I-CSCF, and other server specific capability such as processing power, memory. Based on this capability, the master node assigns a given role to the new node. The role assignment can happen due to load balancing, node failure and other node management purposes that can be governed by high level policies configured at the master node. Alternatively, it can be network event-based, such as link failures or sudden change in bandwidth. When such event occurs, master node notifies the appropriate nodes and the network will be auto-configured accordingly. The network auto-configuration can happen in several ways. For example, an IMS node can notify the mobile node about a role change provided the mobile node subscribes to an event after registration. This event could be defined as a function that indicates the change of role for an IMS node. SIP Event notification messages such as SUBSCRIBE and NOTIFY [16] may be used for this event change notification. The role assignment algorithm on the other hand can be implementation specific.

7.3.4 Protocol Interaction between Nodes

To allow deployment of self-organizing IMS networks, protocol interactions among different IMS nodes are necessary. These interactions are among IMS nodes and mobile nodes, and between the master node and IMS nodes.

7.3.5 Node Monitoring

To allow the master node to determine the functional behavior of all other nodes in the network, a periodic message (heartbeat) will be sent to all known nodes. If a response is not forthcoming within a specific time, the given node is marked as troubled or failed, and the master node will distribute its functionality to the remaining nodes dynamically.

7.4 Self-organizing IMS Node Functions

In this section, we describe the functionality supported by each node involved in self-organizing IMS network.

7.4.1 Operations of the Master Node

The master node is the main component of the centralized approach. In addition to assigning roles to nodes, it has other functions depending on the type of event. As an example, in Figure 7.4, if the P-CSCF (e.g., P1) role changes, the master node notifies all S-CSCF nodes and provides them the information about the new P-CSCF. When the S-CSCF receives this notification, it establishes a list of mobile nodes assigned to this P-CSCF and notifies them by way of the new P-CSCF (e.g., P2) (Receive Role Change). Mobile node (MN) and User Equipment (UE) are used interchangeably and have the same meaning in this chapter. The notification message sent by S-CSCF to the mobile nodes contains information about the new P-CSCF. Upon receipt of notification from the S-CSCF, mobile nodes re-register to the new P-CSCF and re-subscribe to event state change for future changes.

On the other hand, if the S-CSCF role changes, the master node notifies all P-CSCF followed by notification of mobile nodes. The notification message includes information about the new S-CSCF. Upon receipt of this notification message, the mobile node re-registers and subscribes to event state change. To handle load balancing, the master node may request each of the IMS components to support a given percentage of mobile nodes previously attached to the IMS node with role change functionality.

7.4.2 Operations of Mobile Node

The operation of the mobile node is based on SUBSCRIBE and NOTIFY methods. The mobile node should be able to receive and process the notification message (NOTIFY) received either from S-CSCF when P-CSCF

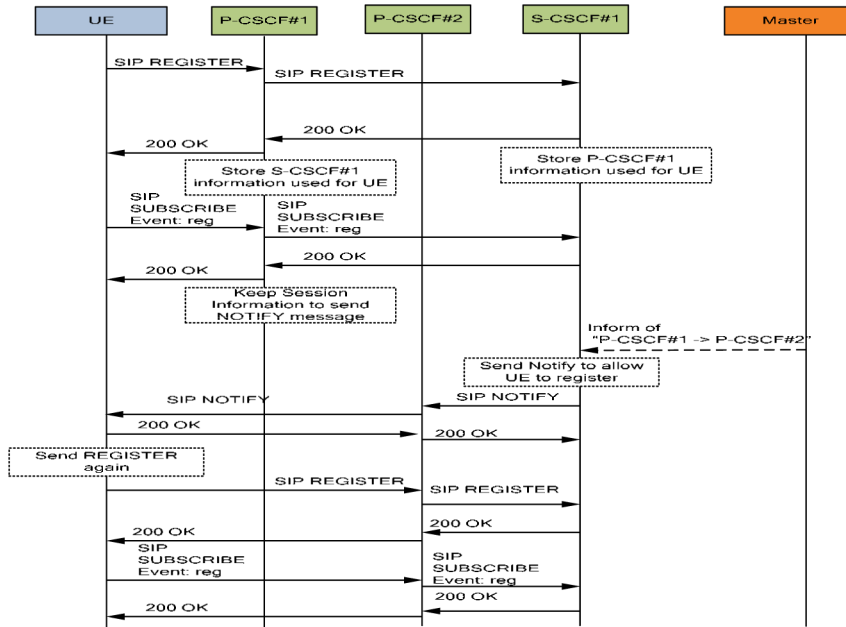


Figure 7.4 Role of master node during P-CSCF change.

role changes or from P-CSCF if S-CSCF role changes. Upon receipt of this NOTIFY message, the mobile node uses such information to register and subscribe again to event change. Mobile node’s ongoing session including the QoS of the media is not affected since the access network that the mobile node is attached to is not changing. However, if the mobile node needs to establish a new session, it should include the Replace-header field in the INVITE message or use re-INVITE message.

7.4.3 Operation of the P-CSCF

To allow self-organizing IMS networks, P-CSCF components should be able to support the following additional operations:

- Receive notification of role change from the master node;
- Process (proxy) and store mobile node’s subscription information during SUBSCRIBE operation;
- Send NOTIFY message with information about the new S-CSCF;

- Retrieve all mobile nodes registered to S-CSCF upon receipt of event change from the master node;
- Process mobile node’s re-registration request after a role change event.

7.4.4 Operation of the S-CSCF

To allow self-organizing IMS networks, S-CSCF components should be able to support the following additional operations:

- Receive notification of role change from the master node;
- Store mobile node’s subscription and profile information;
- Send NOTIFY message with information about the new P-CSCF;
- Retrieve all mobile nodes registered to the P-CSCF upon receipt of event change from the master node;
- Update HSS with the mobile node information after the role change.

7.5 Self-organizing Deployment Scenarios

In this section, we first illustrate a few migration scenarios of functional components for self-organizing IMS network and then describe the associated call flows and proof-of-concept from a prototype test-bed. Ideally, self-organizing IMS should be able to support all kinds of reconfiguration in the network as shown in Figure 7.5. For example, Scenario 1 in Figure 7.5 shows that all

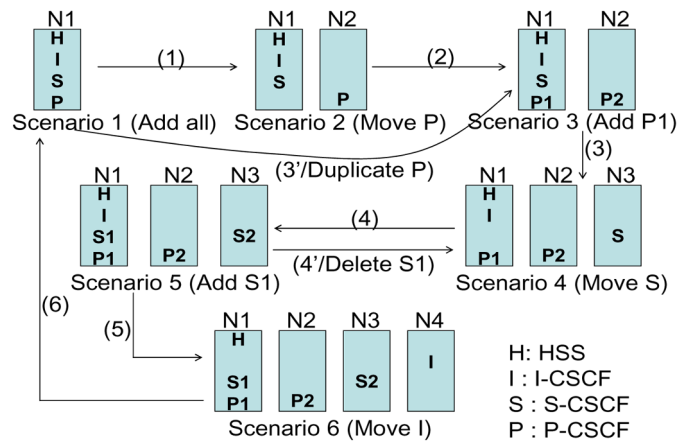


Figure 7.5 Possible re-configuration scenarios for IMS nodes.

the IMS components are running in one node, N1. Scenario 2 shows that P-CSCF functionality is migrated to node N2. Scenario 3 corresponds to a case when the P-CSCF functionality gets split into two nodes, P1 and P2. Scenario 4 depicts when the S-CSCF functionality is relocated to a different node. Scenario 5 corresponds to a case when the S-CSCF functionality gets split into two nodes S1 and S2. Scenario 6 shows I-CSCF's functionality getting migrated to a new node, and finally, the transition step (6) shows the merging functionality where all the functions merge with a single node. It is important to note that our goal is not to affect the end users (whether they have an ongoing session or initiate a new session) due to these changes in the configuration within the core of the network. In Figure 7.5, H stands for HSS, I-CSCF is termed as I, S-CSCF and P-CSCF are denoted as S and P, respectively.

We discuss two types of self-configuration scenarios: Scenario (i) where mobile node is aware of reconfiguration and participates during self-organizing process and Scenario (ii) where mobile node is not aware of self-organizing IMS and does not participate during the reconfiguration process.

7.5.1 Self-organizing with UE Involved Case

In this section, we illustrate several self-organizing IMS scenarios that could be applicable to both UE involved and UE non-involved cases.

Figure 7.6 shows the call flows for Scenario 1, where all SIP-based IMS components (i.e., HSS, S-CSCF, I-CSCF, and P-CSCF) are deployed in the same node. In this scenario, mobile node 1 (MN1) and mobile node 2 (MN2) register with Node 1 (N1) and a call session is established between them. To allow deployment of self-organizing IMS networks, after registration, mobile nodes must subscribe to S-CSCF and P-CSCF for the role change event.

We use MESSAGE method for any communication between the mobile node and the master node. Initial capabilities exchange between the IMS node and master node and role assignment functions are carried out using MESSAGE method. We show below a sample MESSAGE method [1] showing "Role Request" and "Role Assignment". As part of role request, the IMS node offers its capabilities in terms of CPU, memory, and network load. Role Query response is not shown. Figure 7.7 shows the call flows for Scenario 2 as depicted in Figure 7.5, which corresponds to the case when the P-CSCF functionality is relocated in a new node (i.e., Node 2 (N2)) while HSS, S-CSCF, and I-CSCF remain in Node 1. In this scenario, when Node 2 comes

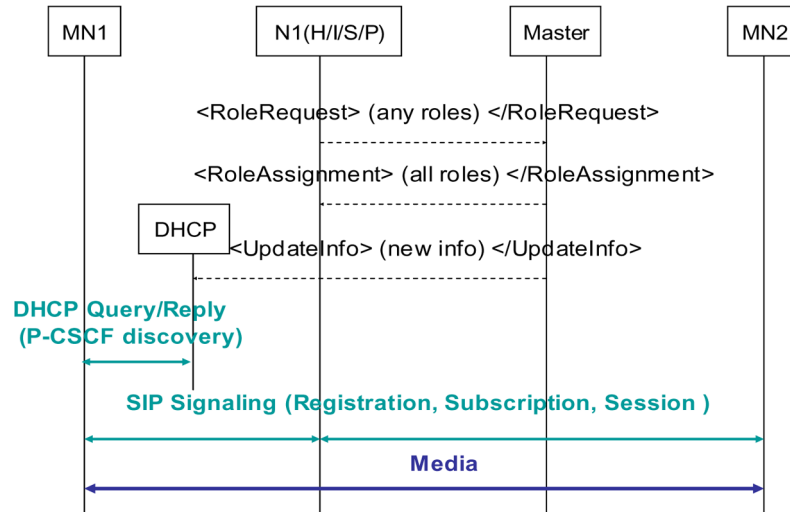


Figure 7.6 Call flow for Scenario 1.

online, it exchanges its capability information with the master node and it is found that Node 2 has the ability of serving as P-CSCF.

Hence, the master node decides to transfer or move the P-CSCF functionality to Node 2 which acts now as P-CSCF for both mobile node 1 and mobile node 2. The procedure is described as follows:

- The old IMS Node (N1) is providing the IMS functionalities for the Mobile Node.
- When a new IMS Node (N2) comes online, it sends a request to the master node for role assignment. According to the capabilities provided in the role request response message, the master node decides the role assignment to N2 governed by the operator policy. The master node provides information about the current IMS entity if the role has been already assigned to other IMS nodes.
- Upon deciding on the role assignment, the master node notifies the S-CSCF (N1/S) about role assignment changes and provides information about the new IMS Node (N2/P). The S-CSCF will notify all mobile nodes registered to the previous P-CSCF.
- The new IMS Node (N2/P) retrieves information (state or context) from the old IMS Node (N1).

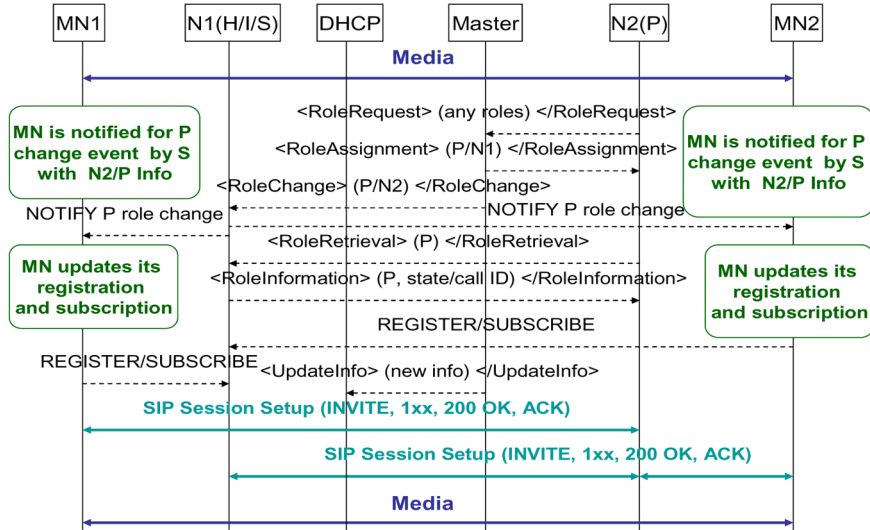


Figure 7.7 Call flow for P-CSCF relocation (Scenario 2).

- In order to allow new mobile nodes to discover the correct P-CSCF, the master node updates the DHCP server configuration.

Scenario 3 corresponds to the situation when P-CSCF functionality gets split among two nodes (P-CSCF1 in Node 1 and P-CSCF2 in Node 2). In this scenario, the master node decides to assign P-CSCF role to Node 1 while earlier P-CSCF role is located in Node 2 (from Scenario 2). Mobile node 1 is assigned to Node 1 as its P-CSCF (i.e., P-CSCF1) while mobile node 2 is assigned to Node 2 as its P-CSCF (i.e., P-CSCF2). After these steps, a call session is established between mobile node 1 and mobile node 2. Note that the P-CSCF functionality splits happen after the initial registration of mobile node 1 and mobile node 2. In other words, mobile node 1 and mobile node 2 were associated to the same P-CSCF (Node 2).

Figure 7.8 shows the call flows for Scenario 4 that corresponds to the situation when S-CSCF functionality is relocated in a new node (i.e., Node 3). In this scenario, when Node 3 comes online, it exchanges its capability information with the master node and it is found that Node 3 is suitable for serving as S-CSCF. Then, S-CSCF functionality of Node 1 is transferred to Node 3 and Node 1 (i.e., P-CSCF, HSS or I-CSCF) is notified for this change.

Transition step (6) depicts a situation when IMS main nodes (P-CSCF, S-CSCF) functionalities move back to one node due to failure of other nodes

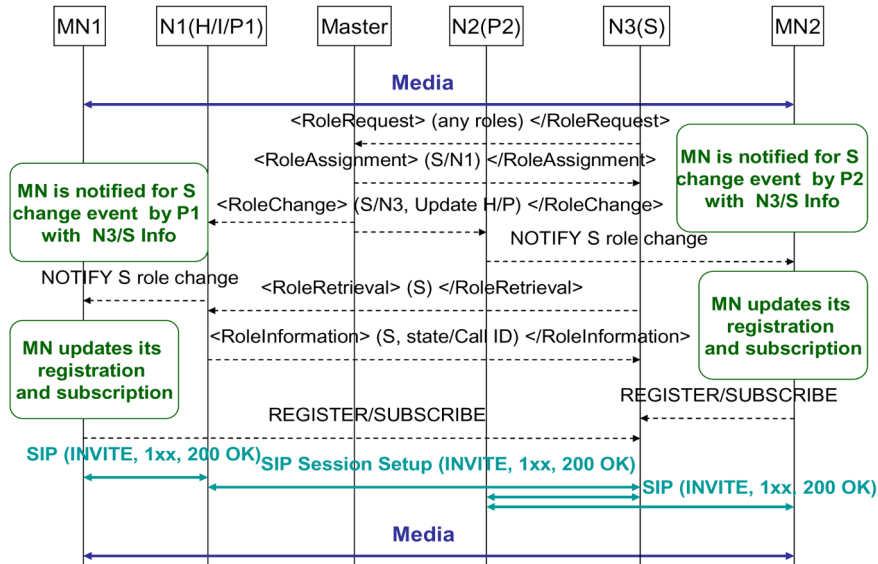


Figure 7.8 Call flow for S-CSCF relocation (Scenario 4).

or load balancing reasons. The master node detects this state change or event and restores all functionalities previously supported by each IMS node to Node 1. Mobile node 1 and mobile node 2 re-register with Node 1 again. Before making such merging decision in one node, the master node needs to consider several factors such as time to restore the functionalities and service disruption.

7.5.2 UE Non-involved Scenarios

In this section, we describe the techniques needed to support UE non-involved scenario. UE non-involved method requires usage of the Load Balancer (LB) in order to hide any change in the core IMS network. With the increasing penetration rate of real-time application such as VoIP, it is necessary to have load balancing support at SIP level 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.

In the proposed solution, the LB appears as a virtual P-CSCF to UEs. In other words, from the UE's perspective, the LB is the P-CSCF and

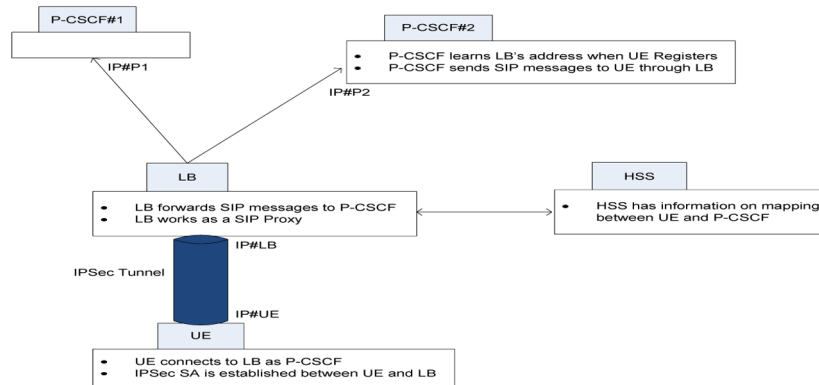


Figure 7.9 SIP load balancing in IMS architecture.

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 7.9 illustrates our proposed SIP-based load balancing procedure for IMS networks. When the LB receives SIP messages from the UE, it forwards these messages to the selected P-CSCF. 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 and Route) in SIP messages will be set appropriately by the LB.

7.5.2.1 Registration Procedure

When 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 the Via header, the LB will receive the response to the request. In the opposite direction, the LB removes the Via

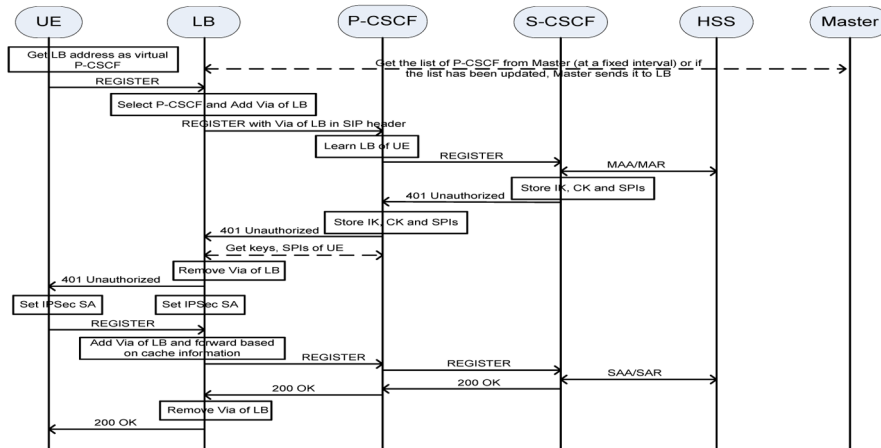


Figure 7.10 SIP registration procedure.

header which contains its information before to send the message to the UE. The selection of P-CSCF is based on different scheduling algorithms, such as hash over Call-ID, hash over From URI, or Round-Robin. 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 7.10 shows the SIP registration procedure for UE non-involved case. For sake of simplicity, only main registration steps are illustrated. For example, interaction with I-CSCF has been ignored. The LB gets a list of P-CSCF from the master node. Since the master node has knowledge of IMS nodes, getting the list of P-CSCFs from the master node allows a reduction of signaling due to IMS node monitoring by LB. The remaining registration procedure process follows the IMS specification. The LB retrieves security association parameters (SPis), integrity (IK) and confidentiality keys (CK) from the P-CSCF.

With all security credentials, the UE will compute the response to the challenge and establish an IPsec SA with the LB before it can send the second SIP REGISTER message. The LB then forwards this message to the previously selected P-CSCF based on the session persistence approach. Similarly, to the processing of initial registration message, the LB adds/removes its information in the Via header.

7.5.2.2 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 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 on the previous selected P-CSCF for a given UE.

7.5.2.3 Load Balancer Failure Support

Since the load balancer is a main anchor point, it might become a single point of failure for the SOIMS architecture. To avoid the whole system from being out of service due to the LB failure, a backup of LB might be deployed for redundancy. The heartbeat technique or Virtual Router Redundancy Protocol (VRRP) [8] can be used between the primary and the backup LB and they communicate periodically to inform each other that they are 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).

7.5.2.4 Session Continuity with P-CSCF Failover Support

Service provisioning 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 Figure 7.11) 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. 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 nego-

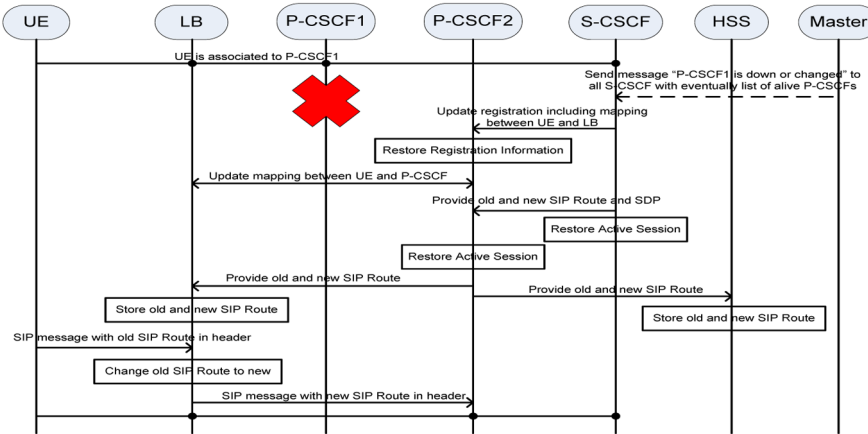


Figure 7.11 Session continuity during P-CSCF failure.

tiation (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 sent with the old SIP Route. It is the 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. In other words, there is no need to exchange SIP SUBSCRIBE/NOTIFY messages, leading to minimal signaling overhead and network resources usage.

7.5.2.5 Session Continuity with S-CSCF Failover Support

The S-CSCF failover support is similar to P-CSCF failover as illustrated in Figure 7.12. 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 Figure 7.4) receives the notification from the P-CSCF, it re-

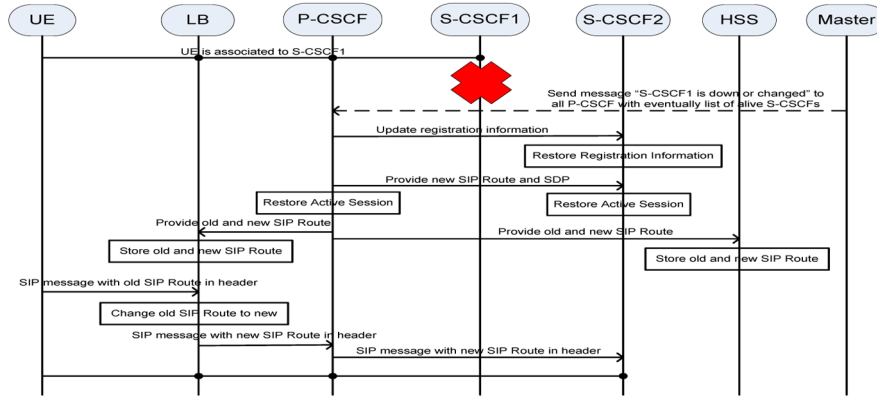


Figure 7.12 Session continuity during S-CSCF failure.

stores 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.

7.6 Implementation and Test-bed Prototype

We implemented both UE (mobile) involved and UE (mobile) non-involved cases to realize the self-organizing IMS. UE involved case requires that the mobile is involved in signaling exchange during reconfiguration process. UE non-involved case does not require mobile’s involvement during network node reconfiguration, network entities such as load balancer and SIP proxies help re-registering the mobile when the network components change their roles.

Figure 7.13 shows the test-bed architecture used for prototyping the proposed solution. IMS node in Figure 7.13 refers to an IMS component (e.g., P-CSCF, I-CSCF, S-CSCF or HSS) running in one or different physical nodes. The home 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

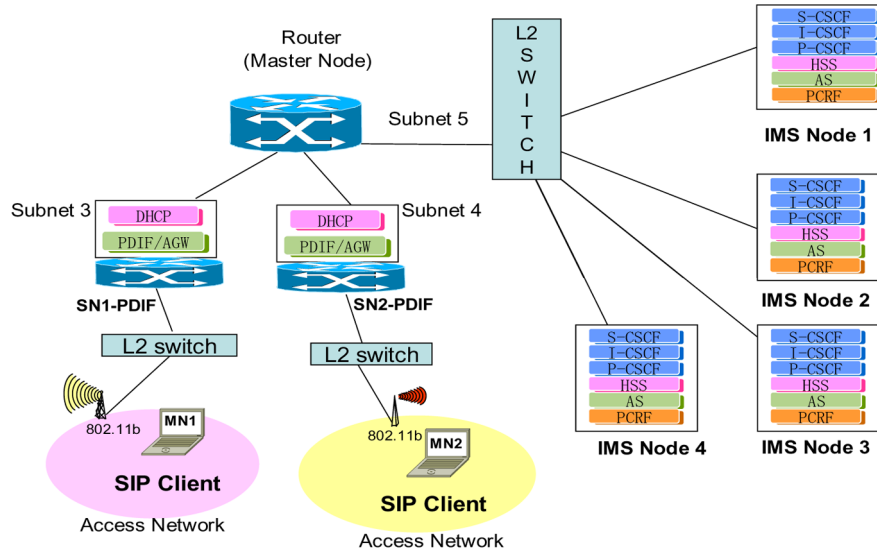


Figure 7.13 Self-organizing IMS test-bed.

PDIF (Policy Decision and Information Function) and DHCP relay agent. SIP stack used in the test-bed is based on NIST implementation [10]. Mobile nodes use SIP user agent based on SIP Communicator [4] to communicate with the IMS nodes. We use “dibbler” as a DHCP client on the mobile. P-CSCF discovery procedure is based on DHCP as specified by IMS specification. We have used XML-based query-response mechanism to obtain the required information from the master node and for role assignment. Mobile nodes communicate with the DHCP server that resides in the router via the DHCP relay agent in the edge router in order to obtain the IP address and discover SIP servers.

Figure 7.14 shows the screen shots of different operations in the test-bed. As shown in Figure 7.14, green illustrates that the specific functionality (e.g., P-CSCF, S-CSCF, I-CSCF) in the current node is active. Red indicates that the nodes are inactive. It also shows how initially all the IMS functions are running in Node 1 but some changes in network condition migrate some of these functionalities to Node 2 (e.g., P-CSCF) and Node 3 (e.g., S-CSCF). Preliminary performance results for server reconfiguration indicate that it takes less than one second to re-assign the respective server functionalities by the master node. This reconfiguration time plays an important role in determining the extent of service interruption for the new calls.

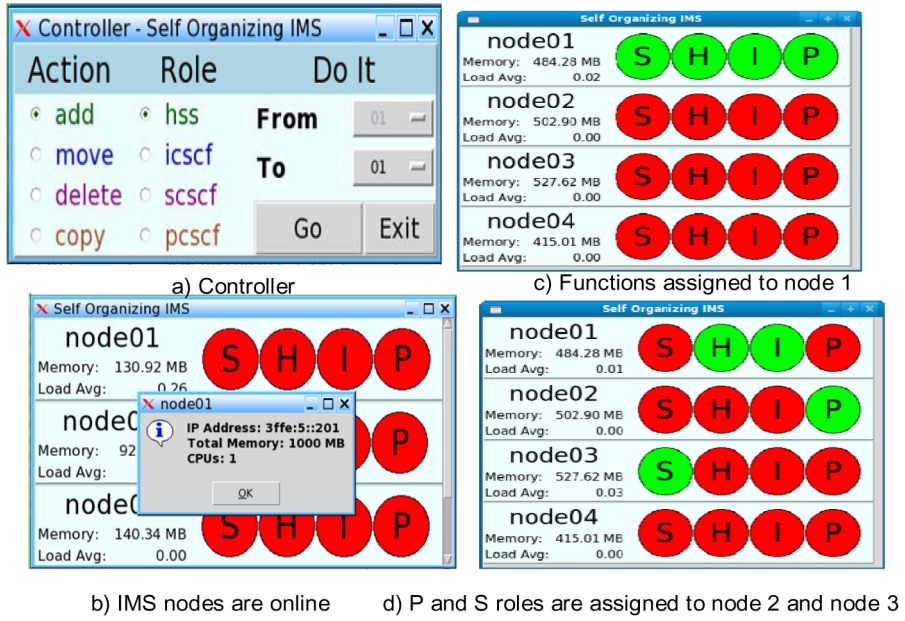


Figure 7.14 Screen shots from self-organizing IMS.

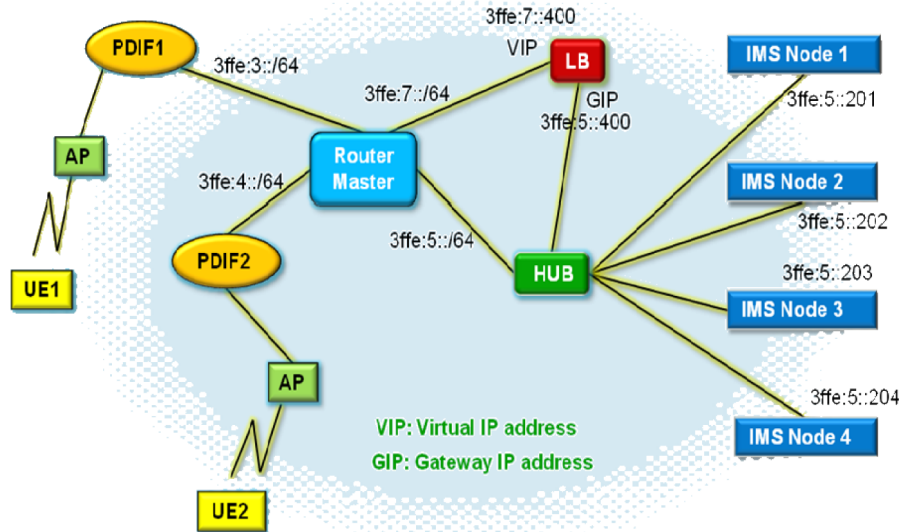


Figure 7.15 Self-organizing IMS with a load balancer.

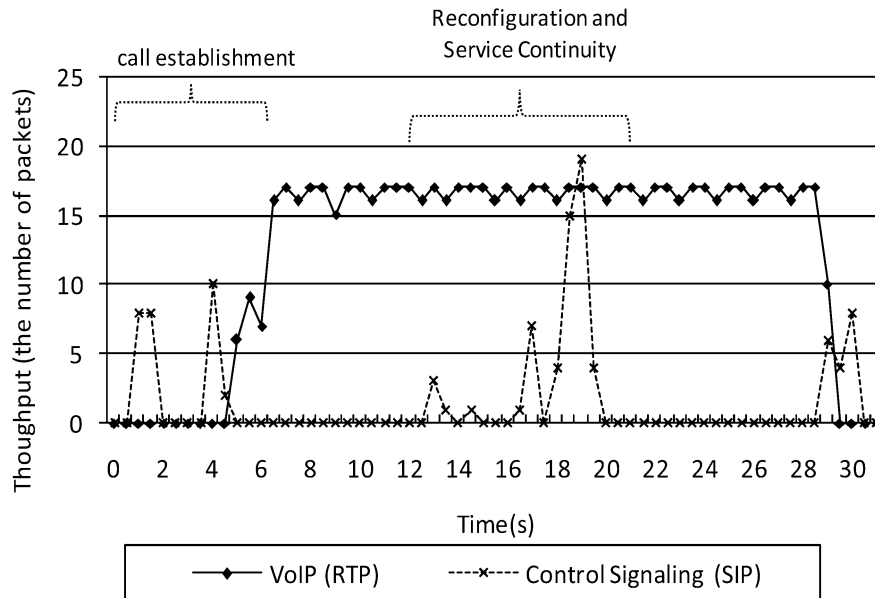


Figure 7.16 SIP control signal during P-CSCF change.

Figure 7.15 shows how the load balancer (LB) is used for UE non-involved scenario. During this reconfiguration delay, ongoing session is not interrupted. For LB, Open SIP Server (OpenSIPS) [14] 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, which has been extended to support 3GPP IMS requirements, is used as the IMS SIP client. We used SIPP [19] to generate SIP traffic load. Figure 7.16 shows the amount of additional control signals (e.g., SIP) generated when P-CSCF is changed although there is no interruption in media traffic (e.g., RTP).

7.7 Conclusions

We believe that self-organizing IMS is an important step towards the successful deployment of IMS networks. A policy-based configuration and reconfiguration of IMS components will help operators reduce the cost and complexities of current networks. While there are standards groups such as

IEEE NGSON, ATIS SON and ITU-SG13 that are currently working on the requirements and architecture of self-organizing, context-aware networks, this chapter goes one level beyond by realizing the concept of self-organizing IMS in a laboratory environment. Our approach is self-adaptive in the sense that we preserve the basic IMS node functionalities with network and operational environment change.

Self-organizing networks (SONs) have been standardized by NGMN alliance and 3GPP 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 chapter has proposed a self-organizing and context aware approach to IMS networks. The proposed approach allows to hide any change in the core IMS networks to end-users (i.e., user equipment or agent does not 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 failure of network entities or IMS role re-assignment with minimal control overhead. As a result, the IMS network's efficiency and scalability have 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 (e.g., I-CSCF, Application Servers).

References

- [1] Session initiation protocol (SIP) extension for instant messaging. RFC 3428, Internet Engineering Task Force, December 2002.
- [2] ATIS SON, <http://www.atis.org/son>.
- [3] T. Bessis. Improving performance and reliability of an IMS network by co-locating IMS servers. *Bell Labs Technical Journal*, 10(4):167–178, 2006.
- [4] SIP Communicator, <http://sip-communicator.org/>.
- [5] A. Dutta, K. Manousakis, S. Das, T. Chiba, and H. Schulzrinne. Mobility testbed for 3GPP2-based multimedia domain networks. *IEEE Communications Magazine*, 45(7):118, July 2007.
- [6] J. Fabini, P. Reichl, A. Poropatich, R. Huber, and N. Jordan. "IMS in a Bottle": Initial experiences from an OpenSER-based prototype implementation of the 3GPP IP multimedia subsystem. In *Proceedings of International Conference on Mobile Business, ICMB'06*, pages 13–13, 2006.

- [7] A. Gulbrandsen, P. Vixie, and L. Esibov. A DNS RR for specifying the location of services (DNS SRV). RFC 2782, Internet Engineering Task Force, February 2000.
- [8] Ed. Hinden. Virtual router redundancy protocol (VRRP). RFC 3768, Internet Engineering Task Force, April 2004.
- [9] IEEE NGSON, <http://grouper.ieee.org/groups/ngson>.
- [10] NIST implementation, <https://jain.sip.dev.java/>.
- [11] A. Manzalini and F. Zambonelli. Towards autonomic and situation-aware communication services: The CASCADAS vision. In *Proceedings of IEEE Workshop on Distributed Intelligent Systems: Collective Intelligence and Its Applications, DIS 2006*, pages 383–388. IEEE, 2006.
- [12] E. Marocco, A. Manzalini, M. Sampò, and G. Canal. Interworking between P2PSIP overlays and IMS networks – Scenarios and technical solutions.
- [13] M. Matuszewski and M.A. Garcia-Martin. A distributed IP multimedia subsystem (IMS). In *Proceedings of IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks, WoWMoM 2007*, pages 1–8, 2007.
- [14] OpenSIPS, <http://www.opensips.org/>.
- [15] Third Generation Partnership Project, <http://www.3gpp.org/>.
- [16] Adam Roach. Session initiation protocol (SIP)-Specific event notification. RFC 3265, Internet Engineering Task Force, June 2002.
- [17] Henning Schulzrinne. Dynamic host configuration protocol (DHCP-for-IPv4) option for session initiation protocol (SIP) servers. RFC 3361, Internet Engineering Task Force, August 2002.
- [18] ITU-T SG13, <http://www.itu.int/itu-t/studygroups/com13/index.asp>.
- [19] SIPP, <http://sipp.sourceforge.net/index.html/>.