A Flow-based Traffic Model for SIP Messages in IMS

Jie Xiao, Changcheng Huang and James Yan
Department of Systems and Computer Engineering, Carleton University, Ottawa, Canada
{jiexiao, huang}@sce.carleton.ca, jim.yan@sympatico.ca

Abstract— The IP Multimedia Subsystem (IMS) defined by the 3rd Generation Partnership Project (3GPP) and 3GPP2 provides a platform for the provision of multimedia services with quality of service (QoS). In addition, this service architecture allows third-party vendors to create advanced multimedia and multisession services across wireless and wireline network access. The Session Initiation Protocol (SIP) supports the signaling and session management functions of these services; therefore, the SIP performance is critical to the services' quality of experience. Thus, in order to conduct an SIP performance evaluation, an efficient yet representative model for SIP signaling traffic is needed. In this article, we provide an in-depth flow analysis of a number of SIP session procedures defined in IMS and quantify the SIP signaling traffic at *flow* level. By utilizing the signaling *flow* analysis, the workload of servers can be predicted with a simple mathematical calculation. The complex correlation structure of the workloads across different signaling servers is naturally captured by the flow concept we introduced. This model also allows for flexibility when expanding the SIP session procedures in IMS networks. According to the simulations that we carried out using OPNET, the model we proposed is proven to be acceptable.

***Index Terms*— IP Multimedia Subsystem, SIP Server, SIP signaling Traffic, Traffic Model**

I. INTRODUCTION

The IP Multimedia Subsystem (IMS), standardized by the 3rd Generation Partnership Project (3GPP) and 3GPP2 [1], is envisioned as the next-generation IP-based multimedia communication system that integrates data, speech, and video network technology, and it also covers wireless and wireline networks. Furthermore, it provides a service control platform that allows mobile users to access new multimedia and multisession applications across fixed and mobile terminals [2, 3, and 4]. Through standardized service creation interfaces, IMS allows the development of new multimedia and multi-session applications. Also, it enables users to set up multiple services very easily in a single session or multiple synchronized sessions. At the same time, IMS is appealing to all types of service providers, given that it allows service providers to charge according to different services.

For session control, IMS uses Session Initiation Protocol (SIP), which is defined in [5]. SIP as an application layer control protocol is an Internet Engineering Task Force (IETF) standard for multimedia conferencing over IP [6]. In addition, SIP lies at the core of the IMS architecture, and plays the role of session establishment, modification, and termination between two or more end points; this occurs mostly between end-user and Call/Session Control Function (CSCF) server, or between the two CSCF [1] servers. Figure 1 illustrates a simplified IMS core network architecture. This paper only focuses on the network entities illustrated. IMS based networks consist of distinct CSCF servers and Home Subscriber Server (HSS). There are three types of CSCF servers, the Proxy-, Interrogating- and Server-Call/Session Control Function servers (P-CSCF, I-CSCF, S-CSCF, respectively [1]).

P-CSCF as an outbound/inbound SIP proxy server interfaces with the User Equipment (UE) for core network service access. I-CSCF is a SIP server that is located at the edge of the administrative domain and performs the routing function. S-CSCF is a SIP server that is the central node of the signaling plane. It acts as a registrar with the responsibilities of both the UE registration as well as session control. Furthermore, HSS, as a master database, maintains user profiles. As shown in Figure 1, P-CSCF is located in the access network while I-, S-CSCF, and HSS are located in the home network.

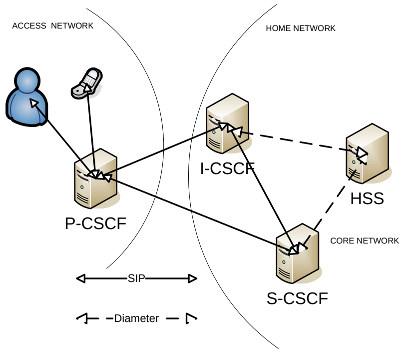

Figure 1. Simplified IMS core network architecture

Signaling traffic is considered to be another important type of network traffic other than media traffic. End users may experience a delay due to the congestion of the signaling network. Moreover, while the number of subscribers and their demands on new services is increasing, service providers need a model in order to predict the impact of new services on servers when the new services are introduced to market. By having such a model, service providers can predict the network availability as well as maintain the stability of the system, and thereby increasing their revenue potential. As a result, establishing a SIP signaling traffic model is necessary for performance evaluation. However, in the literature, only a few studies on IMS signaling traffic model are published. Moreover, these publications offered limited coverage of the session procedures in IMS, and only focus on one single network, namely, home network. In this paper， we will do a careful and more comprehensive signaling traffic analysis that covers 20 session procedures in 6 routing scenarios. In addition, we will introduce the flow concept into the quantification of SIP signaling traffic, and propose a flow-based signaling traffic model for IMS.

Most of IMS-related research work currently has concentrated on IMS architecture and SIP protocol development [4], network performance evaluation under varying network parameters [3][8][9], and the Quality of Service (QoS) issue [10]. Besides that, [2] proposed a traffic model that is limited to the signaling traffic created to and from the HSS in only 3 procedures, which are Registration, Session Setup and Presence watch subscription in home network. The paper also lacks network implementation and simulation. In [11], the authors have done SIP signaling delay and characterized SIP server workload on the Yahoo case study. The traffic in the network is assumed to follow an M/G/1 model.

The rest of this paper is organized as follows: Section II covers 6 defined routing scenarios and 20 IMS session procedures; Section III analyzes signaling traffic by applying the *flow* concept and formulates our traffic model; Section IV introduces the implementation of the traffic model and gives the results from the simulations; Conclusion and future work are presented in Section V.

II. IMS SIP SESSION PROCEDURES AND ROUTING SCENARIOS

A. *IMS Session Procedures*

IMS defines a large number of procedures for various signaling functions in [12]. Generally speaking, there are five different categories of procedures. They are Registration and De-registration, Session Initiation, Session Termination, Session Failure, and Session Redirection. Each category is further divided into several cases as shown in Table 1. In the last column of the table, the likely frequency of each procedure within its category is provided. The frequencies are based on our analysis and they are the assumption we made for the network simulation, discussed in Section IV.

Both new and old end users are required to register in the IMS network before any session is initiated. After successful registration, the session initiation can be performed by users' requests. Once the basic session has been set up completely, the media traffic can flow between the end users. The de-registration procedure may be triggered by mobiles or the network under some situations. A session initiated by the user may fail due to an error detected in the servers; however, the decision to redirect a session to a different destination may be made for different reasons in the establishment of the session. At the end of a session, the session termination procedure allows the session to be released.

B. *Call Scenarios*

An end-to-end session establishment is clearly defined by IMS. It is achieved by successfully establishing a call between two end users. As two parties, the end users can belong to and lie in two identical/different network operators [13]. All possible call scenarios between two parties are shown in Figure 2. It illustrates four types of parties, H_1, H_2, V_1, and V_2. Each party acts either as a caller/originator or a callee/terminator. Any of the parties can call another party to establish an end-to-end session. Therefore, all possible call combinations between these 4 parties form 10 call scenarios as shown in Figure 2.

Table 1 IMS session procedure

		Session Procedures	*Frequency*
Session Initiation	1	Basic Session Setup	75%
	2	Re-invite for new codec, without I-CSCF	10%
	3	Re-invite for server codec	10%
	4	Re-invite, failure happen	5%
Registration	5	Registration, user not registered	54%
	6	Re-registration, user registered	20%
De-registration	7	Mobile initiated	10%
	8	Network initiated, registration timeout	10%
	9	Network initiated by HSS, Administration	3%
	10	Network initiated, service platform	3%
Session Termination	11	Mobile terminal initiated Session release	50%
	12	Network initiated session release P-CSCF initiated	50%
Session Failure	13	Failure in session abandon, origination procedure	40%
	14	Failure in obtaining resource, origination procedure	40%
	15	Failure in termination procedure	10%
	16	Rejection by termination procedure	10%
Session Redirection	17	Initiated by S-CSCF to CS-domain	40%
	18	Initiated by S-CSCF to IM CN subsystem	40%
	19	Initiated by P-CSCF	10%
	20	Initiated by UE	10%

For example, in Figure 2, Link 1 represents a complete end-to-end session between a non-roaming user (H_1) and one roaming user (V_1); the two users are subscribers of different network operators and both lie currently in Network 1. The possible call scenarios for this session are:

1. H1 user as originator (H1,O) →Call→ V1 user as terminator (V1,T)
2. V1 user as originator (V1,O) →Call→ H1 user as terminator (H1,T)

In this paper, we only focus on the traffic created to and from the Network 1, since home network performance is our concern. Moreover, it is noted that the traffic generated from H_2 as originating procedure (H2, O) and terminating procedure (H2, T), will be considered to be unrelated to Network 1 traffic. Then, the 4 parties, H_1, V_1, V_2, and H_2, form 9 call scenarios, which are considered in our analysis.

Any call scenario can establish an end-to-end session and we have 20 session procedures identified in Table 1. Thus, we have to analyze the signaling traffic involved in the 20 session procedures for the 9 call scenarios individually; this is a huge task. As a result, we introduce routing scenarios to help us reduce the complexity of analysis of IMS signaling traffic.

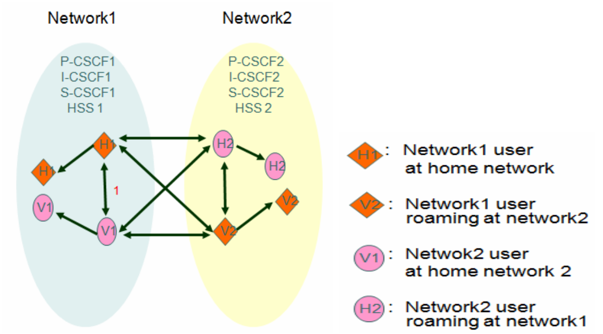

Figure 2. End-to-end session flows in 2 networks

C. Routing Scenarios

When a signaling message is routed through the signaling network, the procedures as defined in IMS can be divided into two parts: the originator part and the terminator part. The procedures related to the originator are called originating procedures while the procedures related to the terminator are called terminator procedures. Therefore, an end-to-end call is a concatenation of the originating procedures and the terminating procedures.

For the 9 call scenarios discussed above, either the originator or the terminator must be in the home Network 1, and then the routing scenarios can be classified below:

- Originating routing scenarios:(H_1, O), (V_1, O), (V_2, O)
- Terminating routing scenarios: (H_1, T), (V_1, T), (V_2, T)

As mentioned in the previous section, Network 1 performance is our concern. In order to perform the signaling flow analysis within Network 1, we need to identify which servers are logically located in Network 1 and carry the signaling traffic for a given routing scenario. To get this information, we analyzed all procedures as defined in IMS; the results are shown in Table 2. In the last column of the table, the proportionate amount for each routing scenario is provided based on our analysis and they are the assumptions made for the network simulation, which we will discuss in section IV.

Two examples are shown in Figure 3. The first one refers to the cases (V_1, O) and (V_1, T), where only P-CSCF 1 is involved in Network 1. The second one is the cases (V_2, O) and (V_2, T), where I-CSCF 1, S-CSCF 1, and HSS 1 are involved in Network 1.

As we mentioned earlier, the routing of an end-to-end call scenario is the concatenation of the corresponding, originating, and terminating routing scenarios. One example is shown in Figure 4. The example shows the servers involved in the call scenario, where (H1, O) calls (V2, T). It can be observed that P-CSCF 1 is involved once in the path of the originating routing scenario, while I-CSCF 1 and S-CSCF 1 are involved twice in the end-to-end call scenario: once in the path of originating routing scenario and once in the terminating routing scenario. The HSS-1 server is only in the terminating routing scenario. The detail of message flows is provided in [1].

It should be noted that although signaling procedures are grouped into routing scenarios as discussed above, not all servers in a routing scenario will appear in all the signaling procedures in the group. For example, the I-CSCF 1 server in Figure 4 may not be involved in some of the procedures in the group; we use dashed lines in the figure to indicate this. Clearly, this makes the estimation of traffic load on a particular server more complex. This is one of the major reasons why we introduce the flow concept.

In summary, out of the 9 call scenarios, we identified 6 routing scenarios. With each routing scenario, 20 session procedures are defined.

Table 2 Network 1 server(s) that involved in routing scenarios

Routing Scenario	**Server(s) in home Network 1**	**Amount (%)**
(H1,O)	P-CSCF 1, I-CSCF 1, S-CSCF 1, HSS 1	50
(H1,T)	P-CSCF 1, I-CSCF 1, S-CSCF 1, HSS 1	20
(V1,O)	P-CSCF 1	10
(V1,T)	P-CSCF 1	5
(V2,O)	I-CSCF 1, S-CSCF 1, HSS 1	10
(V2,T)	I-CSCF 1, S-CSCF 1, HSS 1	5

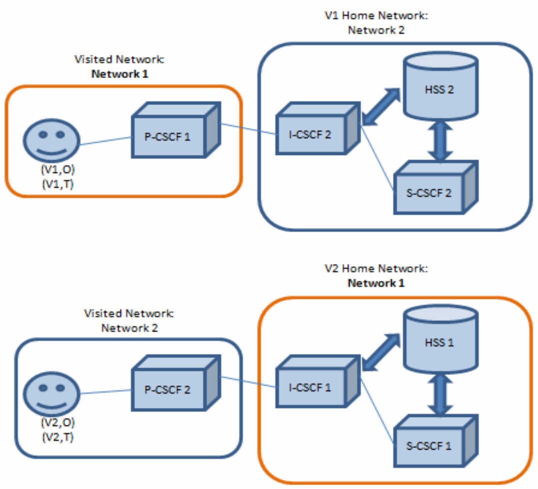

Figure 3. IMS servers involved in different networks according to 4 routing scenarios and servers in orange boxes are logically located in Network 1

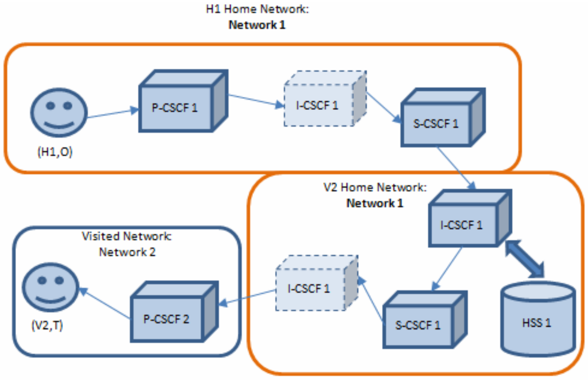

Figure 4. IMS end-to-end session flows where (H1, O) calls (V2, T)

III. IMS TRAFFIC MODEL

A. *Quantification of SIP signaling traffic at flow-level*

As discussed above, the IMS system is extremely complex due to the combination of various call scenarios and the large number of session procedures. It is a challenging task to engineer all kinds of IMS servers in order to satisfy the real-time call performance requirements. To solve this problem, we propose a flow-based approach that maps various combinations of routing scenarios and session procedures to the limited number of signaling flows so that the load of each IMS server can be easily estimated. The flow concept tries to capture the causal relationship among a sequence of messages so that the loads at different servers can be correlated as they occur in the real network.

A signaling *flow* is defined to be the aggregation of message sequences that follows the same path in a network of IMS servers. The statistics of a signaling flow include both flow path and flow volume. The path of a flow identifies the servers that a specific sequence of messages will traverse and the order that these servers will be traversed, while the specific sequence of messages is established by the causal structure among these sequences: one message triggers another one.

The flow volume represents the mean number of messages of the specific flow passing a node for a given unit of time. The key characteristic of a flow is that its volume stays constant across all the servers the flow traverses. Therefore, the flow concept captures the correlation structure among different servers due to various message sequences.

Next, we will show the methods of extracting signaling flows from a typical IMS session procedure, namely, Basic Session Setup in the home network, at originating routing scenario (H1, O). As shown in Figure 5, since P-CSCF 1 and S-CSCF 1 servers are located in the designated home network, the signaling traffic traversing through these two servers is taken into account. We divide all the signaling messages in this session into three signaling flows. Messages sequences (INVITE, 183), (PRACK, 200), (UPDATE, 200), and (PRACK, 200) follow the same path. Furthermore, they are noticed as one response followed by one request, so that the correlation between P-CSCF 1 and S-CSCF 1 servers load can be captured, namely 1:1. They are treated as one signaling flow, which is called flow 1, and its flow volume is 8 messages per dialogue. Flow 1 is a round trip flow, while the other two flows are one-way trip in nature; the details are listed in Table 3. Following the same approach, we have analyzed the 20 session procedures with the 6 routing scenarios as discussed above. In the end, 17 flows have been identified as shown in Table 4. Also shown in Figure 6 are the volumes of the flows per session procedure in one routing scenario, (H1, O). We will call the table as matrix X_1. The matrices for the remaining 5 routing scenarios, $X_2,\ldots, X_6$, are not listed here due to space limitation.

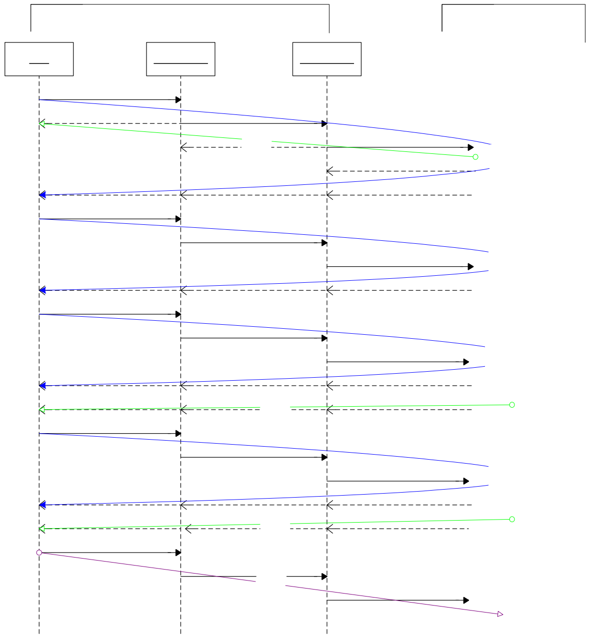

Figure 5. IMS basic session setup procedure with signaling flow analysis, in (H_1, O)

Table 3 Signaling flow summery for basic session setup in (H_1, O)

Signaling *Flows*	Message Sequences	*Flow* Volume	*Flow* Path
1	(INVITE, 183), (PRACK, 200), (UPDATE, 200), (PRACK, 200)	8	→P1 → S1 → --- → S1 → P1 →
2	100,180,200	3	→ S1 → P1 →
3	ACK	1	→ P1 →S1 →

Table 4 Summary of 17 flows

Signaling Flows	**Flow Path**
1	→P1 → S1 → --- → S1 → P1 →
2	→ S1 → P1 →
3	→ P1 →S1 →
4	→I1 → S1 → P1 →---→ P1 → S1 → I1 →
5	→ S1 → P1 →---→ P1 → S1 →
6	→ P1 → S1 → I1 →
7	--- S1 ---
8	→I1 → S1 →---→ S1 → I1 →
9	→ S1 → I1 →
10	--- P1 ---
11	→ P1 → I1 → S1 → I1 → P1 →
12	→ I1 → S1 → I1 →
13	→ I1 → S1 → P1 →
14	→ I1 →S1 →
15	→ HSS → I1
16	→ HSS → S1
17	→ S1 → HSS1

B. *SIP Signaling Flow Analysis*

As discussed in the previous section, every signaling flow transverses different servers according to the flow path. We use a matrix to identify the relationship between the flows and the servers. We denote this matrix as *I*, shown in Figure 7. If a flow traverses a server twice during a round trip, the corresponding value will be 1. If it only traverses a server once, the values will be 1/2, such as both *flow 11* and flow 12 at S node . If a flow does not traverse a server at all, the corresponding value will be 0.

$$X_1 = \begin{array}{l|ccccccccccccccccc|}
 & \mathit{Flow}1 & 2 & 3 & 4 & 5 & \cdots & \cdots & \cdots & \cdots & & 11 & \cdots & \cdots & & 15 & 16 & 17 \\
\mathit{Procedure}1 & 8 & 3 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
2 & 8 & 3 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
3 & 2 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
4 & 2 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 2 & 2 & 0 \\
\vdots & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 1 & 2 & 0 \\
\vdots & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 1 & 1 & 0 \\
\vdots & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
\vdots & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
\vdots & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 2 \\
\vdots & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\vdots & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\vdots & 10 & 3 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\vdots & 8 & 4 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\vdots & 2 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\vdots & 8 & 3 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\vdots & 2 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
18 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
19 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
20 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}$$

Figure 6. Matrix of procedures and flows in (H1, O)

$$I = \begin{array}{l|cccc|}
 & P & I & S & HSS \\
\mathit{Flow}1 & 1 & 0 & 1 & 0 \\
2 & 1 & 0 & 1 & 0 \\
3 & 1 & 0 & 1 & 0 \\
4 & 1 & 1 & 1 & 0 \\
5 & 1 & 0 & 1 & 0 \\
6 & 1 & 1 & 1 & 0 \\
7 & 0 & 0 & 1 & 0 \\
8 & 0 & 1 & 1 & 0 \\
9 & 0 & 1 & 1 & 0 \\
10 & 1 & 0 & 0 & 0 \\
11 & 1 & 1 & 1/2 & 0 \\
12 & 0 & 1 & 1/2 & 0 \\
13 & 1 & 1 & 1 & 0 \\
14 & 0 & 1 & 1 & 0 \\
15 & 0 & 1 & 0 & 1 \\
16 & 0 & 0 & 1 & 1 \\
17 & 0 & 0 & 1 & 1
\end{array}
\qquad
S_1 = \begin{array}{l|cccc|}
 & P & I & S & HSS \\
\mathit{Procedure}1 & 12 & 0 & 12 & 0 \\
2 & 12 & 0 & 12 & 0 \\
3 & 3 & 0 & 3 & 0 \\
4 & 3 & 0 & 3 & 0 \\
5 & 4 & 6 & 4 & 4 \\
\vdots & 2 & 3 & 3 & 3 \\
\vdots & 2 & 3 & 2 & 2 \\
\vdots & 0 & 0 & 1 & 1 \\
\vdots & 2 & 0 & 2 & 1 \\
\vdots & 2 & 0 & 2 & 1 \\
\vdots & 2 & 0 & 2 & 0 \\
\vdots & 2 & 0 & 2 & 0 \\
\vdots & 14 & 0 & 14 & 0 \\
\vdots & 14 & 0 & 14 & 0 \\
\vdots & 4 & 0 & 4 & 0 \\
\vdots & 12 & 0 & 12 & 0 \\
\vdots & 4 & 0 & 4 & 0 \\
18 & 0 & 0 & 0 & 0 \\
19 & 0 & 0 & 0 & 0 \\
20 & 0 & 0 & 0 & 0
\end{array}$$

Figure 7. Matrices of flows and servers, procedures and servers workload

C. *Characterization of Server Workload*

Through the quantitative analysis from previous section, the matrix $S_1=X_1I$ represents the load carried by each server as generated per session procedure for the (H1, O) routing scenario. The result is shown in Figure 7.

$$\begin{array}{cccccccccccccccccccccc}
\mathit{procedure} & & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & & & 19 & 20 \\
 & T_1 = [& 5 & 5 & 5 & 5 & 4 & 4 & 4 & 4 & 4 & 4 & 3 & 3 & 2 & 2 & 2 & 2 & 1 & 1 & 1 & 1]
\end{array}$$

Figure 8. Example of average arrival rates matrix in (H1, O)

Furthermore, let T_1 be the row vector representing the average arrival rates of the session procedures for the routing scenario (H1, O). An example is shown in Figure 8. Then

$$L_I=T_IS_I=T_IX_II \tag{1}$$

represents the loads carried by each server for the (H1, O) routing scenario in terms of messages per unit time.

Considering the fact that there are 6 routing scenarios, the total average loads of all the servers can be calculated as:

$$\text{Л} = \sum_{r=1}^{6} L_r \tag{2}$$

where *Л* is a row vector denoting the total loads of all severs in terms of messages per unit time. Let λ_i denote the i^{th} element of the row vector *Л* and μ_i denote the mean service rate of server *i*, then the utilization of server *i* is $\rho_i= \lambda_i /\mu_i$.

D. *Example of predicting server workload by introducing new application*

IMS supports a wide range of IP-based services over wireline and wireless networks. It allows the third-party vendors to develop new applications for both operators and end-users. In order to guarantee the stability of a network system, it is important for service providers to predict the impact on servers when introducing new applications. Now, let us have an example that shows how to utilize the signaling *flow* analysis to predict the impact on different servers when a new application is introduced into the market. Now, we assume a new application called ABC.

The detailed procedures for predicting server workloads are as follows:

1. Find the session procedures involved when end-users request this service, as shown in Table 5. Those session procedures may contain some of the 20 session procedures discussed above.

   If new session procedure(s) is/are introduced, an approach similar to the one above can be used to identify flows.
2. On the basis of all the signalling *flows* we obtained from the involved session procedures, the two matrices *X* and *I* are created. Estimate the arrival rates matrix *T* based on user statistics and then apply Equations (1) and (2).

Table 5 Example of predicting the server load by introducing a new application

New Application	**Involved Session Procedures**
ABC	1. ABC Session Setup
	2. Re-invite for new codec, without I-CSCF
	5. Registration, user not registered
	6. Re-registration, user registered
	18. Session direction Initiated by S-CSCF to IM CN subsystem
	11.Mobile terminal initiated Session release
	A new Session Procedure: *Flow 1, 2, 3, 18*

IV. NETWORK SIMULATION MODEL & RESULTS

The purpose of this section is to evaluate the behavior of the proposed traffic model in IMS networks with given traffic parameters such as arrival rates of signaling traffic into the networks.

A. Network Model

Figure 9 illustrates the test bench we deployed in OPNET Modeler 14.5. It is a simplified IMS core network that composes of main IMS entities. It effectively realizes the session interactions that allow messages to be exchanged in the network. The network topology is comprised of 4 types of SIP servers: P-CSCF, I-CSCF, S-CSCF, and HSS, one single unit for User Equipments (UEs) and 6 traffic generators for the 6 routing scenarios. For simplicity, we deployed two parallel load balancing servers for every type of SIP server. The network traffic is generated from the 6 traffic generators, and each individual traffic generator generates the traffic according to 20 session procedures. The arrival rates of the 20 session requests for one network domain can be manually configured. It also allows us to change all input parameters and monitor the output results. In order to facilitate the message exchanges among the individual entities, a center switch is used to connect the servers and traffic generators. The center switch emulates an IP network in real life and messages are routed as defined in IMS.

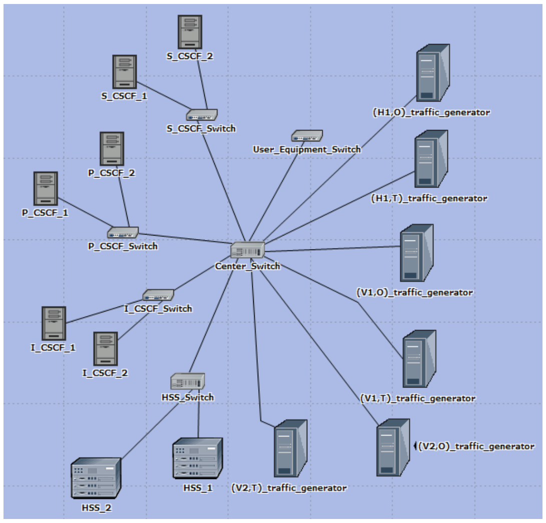

Figure 9. Test Bench of IMS core network in OPNET

B. Network Assumptions

In the simulation, we studied the case of high traffic load incurred by real-time multimedia applications. We assume that there are 0.36 million (M) subscribers. Every subscriber generates an average of two session requests per hour. Consequently, the total arrival rate of session requests equals to 200 session requests per second. The last column of Table 1 provides the assumptions made for the distribution of sessions for all the requested sessions. In a real network, the distribution can be calculated based on the statistics of user requests. Besides that, in the last column of Table 2, we made the assumptions for the distribution of routing scenarios. Based on these two distributions, we can estimate the matrix *T*.

Furthermore, we have assumed the M/M/1 queuing model for the network servers, and the mean service rate for each request arrived at servers follows Poisson distribution and is set to be 10^4 messages per second.

C. Simulation Results

The simulation results consist of server utilizations and mean server queuing delays. As mentioned earlier, the service of every two parallel servers is equally distributed. The utilization values of these two servers remain at the same level; therefore, we will collect one of them. Then, the simulation is set up and has duration of 1 hour. Figure 10 demonstrates the instantaneous utilizations of the 4 servers recorded every 0.1 second in one hour of simulation time. The results calculated from equation (2) are illustrated in Figure 10 as straight lines, which match the simulation results. It is seen that the utilization values in Figure 10 maintain the same level along the one hour simulation for all 4 servers, and it proves that the behavior of the proposed traffic model is consistent with simulation results.

Figure 10. Server utilization recorded every 0.1 second for one hour

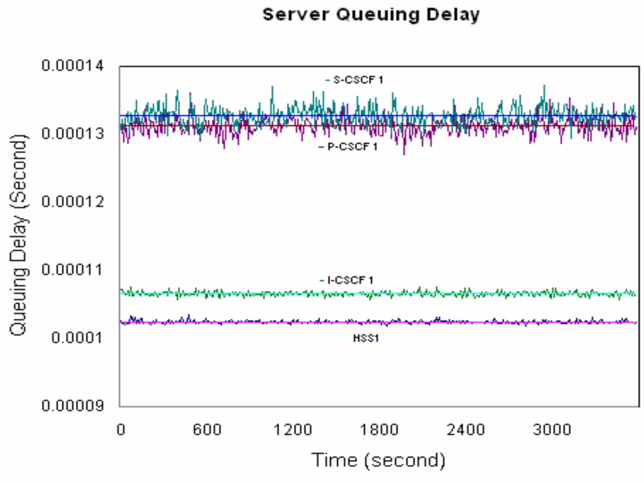

Figure 11. Average server queuing delay over 1 hour

Figure 11 shows the server queuing delays recorded by every second observation instance over 1 hour. We know the mean server queuing delay can be calculated as:

$$D = \frac{1}{\mu - \lambda} = \frac{1}{\mu - \mu\rho} \qquad (3)$$

where μ is the mean service rate (message/sec), λ is the arrival rate for one server (message/sec), and ρ is the server utilization.

Take S-CSCF 1 server as an example:

$$D_1 = \frac{1}{\mu - \mu\rho} = \frac{1}{10{,}000 * (1 - 0.247)} = 0.0001328 \text{ sec}$$

Then, the mean service rates for 4 servers are illustrated in Figure 11 as straight lines, which match the simulation data from the same figure.

V. Conclusion And Future Work

In this paper, we propose an approach for characterizing the server load in IMS networks. The proposed model is based on the quantification of signaling traffic at flow level. This approach allows us to capture the correlation structure of message sequences, while traversing servers across an IMS network. This model is also useful for characterizing the workload of servers for new applications introduced into the telecommunication market. From the simulation results, we verified the model by calculating server utilizations and mean server delays.

Later, we will devise a source model that simulates the IMS user's behaviors in OPNET. Accordingly, the distributions by means of passive measurements are obtained and are applied to the proposed model so that the proposed model can be verified in a different way. Moreover, by utilizing the flow-based approach, an optimization problem that is to minimize node cost, will be considered and formulated in a linear programming problem. Since retransmission scenarios were not considered in this work, future work will extend the proposed model to cover this issue.

VI. Acknowledgement

This work was supported in part by NSERC Grant CRDPJ 354729-07, NSERC Grant RGP 261469-2003, OCE Grant CA-ST-150764-8, and Nortel Networks.

References

[1] P. Agrawal, et.al, "IP Multimedia Subsystems in 3GPP & 3GPP2: Overview and Scalability issues", IEEE Communications Magazine, January, 2008.

[2] V.S. Abhayawardhana, R. Babbage, "A Traffic Model for the IP Multimedia Subsystem (IMS)", IEEE 65th Vehicular Technology Conference, 22-25 April 2007, pp. 783 – 787.

[3] S. Pandey, V. Jain, D. Das, V. Planat, R. Periannan, "Performance Study of IMS Signaling Plane", IP Multimedia Subsystem Architecture and Applications, International Conference, 6-8 Dec. 2007, pp.1 – 5.

[4] V. Koukoulidis, M. Shah, "The IP Multimedia Domain in Wireless Networks: Concepts, Architecture, Protocols and Applications", IEEE 6th International Symposium on Multimedia Software Engineering, 2004.

[5] J. Rosenberg, et. Al. "SIP: Session Initiation Protocol", RFC 3261, June 2002.

[6] T. Borosa, B. Marsic, S. Pocuca, "QoS support in IP multimedia subsystem using DiffServ." Telecommunications, 7th International Conference, Vol: 2, 11-13 Jun 2003, pp: 669 – 672.

[7] P. Calhoun, et al., "Diameter base protocol", RFC 3588, IETF, September 2003.

[8] A. Munir, "Analysis of SIP-Based IMS Session Establishment Signaling for WiMax-3G Networks", Networking and Services, 4th International Conference, 16-21 March 2008, pp. 282 – 287.

[9] A.A Kist, RJ. Hark, "SIP signalling delay in 3GPP", IEEE 6th international Symposium on Communications Interworking of IFIP interworking, Fremantle WA, October 13-16, 2002.

[10] H. Jinho, K. Nakpo, K. Sangyong, K. Jongseog, "A Framework for IMS Interworking Networks with Quality of Service Guarantee", Networking. Seventh International Conference, 13-18 April 2008, pp. 454 – 459.

[11] N. Rajagopal, M. Devetsikiotis, "Modeling and Optimization for the Design of IMS Networks", 39th Annual Simulation Symposium 2006, pp: 34-41.

[12] 3GPP TS 24.228, "Signaling flows for the IP Multimedia call control based on SIP and SDP; Stage 3 (Release 5)", v5.13.0, June 2005.

[13] 3GPP TS 23.228, "IP Multimedia Subsystem (IMS), Stage 2 (Release 7)", v7.6.0, Dec. 2006.

[14]