Fourth International Conference on Networking and Services

# Analysis of SIP-based IMS Session Establishment Signaling for WiMax-3G Networks

Arslan Munir
Mentor Graphics Corporation
Embedded Systems Division
arslan_munir@mentor.com

## Abstract

*The IP multimedia subsystem (IMS) is standardized by the 3rd generation partnership project (3GPP) and 3GPP2 as a new core network domain to support Internet Protocol (IP) based multimedia services over 3G networks. Session Initiation Protocol (SIP) which is an application layer signaling protocol is also standardized by 3GPP and 3GPP2 for session establishment, management, and transformation. In this paper, we study the SIP-based signaling delay for IMS session establishment in 3rd generation (3G) network and worldwide interoperability for microwave access (WiMax) network for different channel rates. In our delay analysis, we take into account transmission, processing and queueing delays at network nodes. The delay analysis of SIP based signaling for IMS provides an insight into the efficiency of SIP signaling for IMS.*

## 1. Introduction

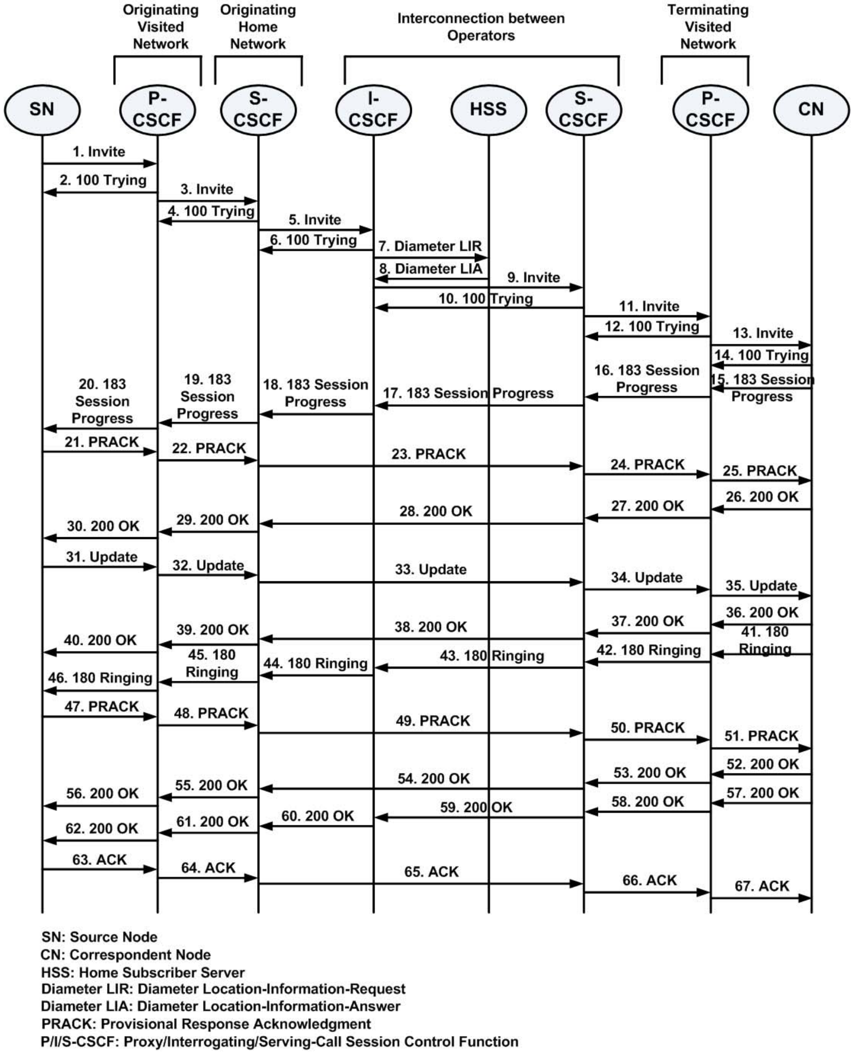
The 3rd generation partnership project (3GPP) and 3GPP2 are standardization bodies responsible for the development of 3G telecommunication system based on the evolution of universal mobile telecommunications system (UMTS) and code division multiple access (CDMA) wireless networks, respectively. The IP multimedia subsystem (IMS) is standardized as a novel 3G core network domain. The IMS permits development of new multimedia and multi-session applications utilizing wireless and wireline transport by providing a service control platform. The IMS also permits service providers to charge according to different policies. Session Initiation Protocol (SIP) is an application layer protocol which is standardized by 3GPP and 3GPP2 for session establishment, management and transformation.

The signaling efficiency for call setup in IMS infrastructure using CDMA2000 is analyzed in [13]. However, it is assumed that both the source node (SN) and the correspondent node (CN) are in a CDMA2000 system. The SIP session setup delay for voice over IP (VoIP) service in 3G wireless networks is studied in [5]. The effect of TCP, UDP, and radio link protocol (RLP) is considered on SIP session setup for VoIP. The performance of SIP-based vertical handoff is analyzed in [2] and analytical expressions for delay of SIP-based handoff to UMTS network from another UMTS network or a wireless local area network (WLAN) and vice versa are given. SIP-based mobility in IPv6 is described in [15] and the delay incurred when a user equipment (UE) moves to a new link and performs the duplicate address detection (DAD) and router selection is examined.

In this paper, we study the SIP-based IMS signaling delay for IMS session establishment procedure shown in Figure 1 for different 3G and WiMax channel rates. The worldwide interoperability for microwave access (WiMax) and 3G are given consideration in our analysis because the two technologies are envisioned to dominate in the future 4G wireless heterogeneous networks. The Signaling Compression (SigComp) is a compression method for general text-based protocols and is developed by Internet Engineering Task Force (IETF) [16]. In our analysis, the SigComp has been considered for SIP signaling to reduce the message size and hence decrease the signaling delays. Our delay analysis of SIP-based IMS signaling is comprehensive taking into account transmission, processing and queuing delays at the network entities. Additionally, provisional responses and authentication procedures involved in the IMS signaling are also considered which, according to the best of our knowledge, were ignored in the previous research in literature. Our delay analysis technique can be applied to any of the signaling flow procedures for different network communication protocols.

The rest of the paper is organized as follows: In Section 2, we present our delay analysis of IMS session setup signaling procedure. Numerical results are presented in Section 3. Conclusions are given in Section 4.

0-7695-3094-X/08 $25.00 © 2008 IEEE
DOI 10.1109/ICNS.2008.7



SN: Source Node
CN: Correspondent Node
HSS: Home Subscriber Server
Diameter LIR: Diameter Location-Information-Request
Diameter LIA: Diameter Location-Information-Answer
PRACK: Provisional Response Acknowledgment
P/I/S-CSCF: Proxy/Interrogating/Serving-Call Session Control Function

**Figure 1. IMS session setup procedure [1], [3].**

## 2. Delay Analysis of IMS Signaling

In this section, we analyze the delay for the IMS signaling procedures. The delay consists of three parts, i.e., transmission delay, processing delay, and queueing delay.

$$D_{total} = D_{trans} + D_{proc} + D_{queue} \tag{1}$$

where $D_{total}$ denotes the total delay for a signaling procedure, $D_{trans}$ denotes the transmission delay, $D_{proc}$ denotes the processing delay and $D_{queue}$ denotes the queueing delay. The transmission delay is the delay incurred due to the transmission of signaling messages which depends on the size of the messages as well as the bandwidth of the channel. The transmission delay considered here also incorporates the delay incurred due to the propagation of signaling messages from one node to another which depends on the distance between the nodes. The processing delay is the delay associated with the encapsulation, decapsulation and routing of packets. The queueing delay is the delay incurred due to the queuing of packets at each node. Our analysis is equally valid for IPv4 as well as IPv6.

### 2.1. Transmission Delay

We only consider wireless link transmission delays as the wired link transmission delays between the core network entities can be considered to be negligible because of the high available bandwidths in the wired links. The wireless link transmission is analytically modeled with and without RLP with TCP (Transmission Control Protocol) as transport layer protocol in [4] and we use this model in our analysis.

The maximum sizes of the IMS signaling messages have

**Table 1. Size of SIP messages involved in IMS session setup**

| SIP Message | Compressed Size |
|---|---|
| INVITE | 810 |
| 183 SESSION PROGRESS | 260 |
| PRACK | 260 |
| 100 TRYING | 260 |
| 180 RINGING | 260 |
| 200 OK | 100 |
| ACK | 60 |
| UPDATE | 260 |

**Table 2. SIP messages $K$ values for different channel rates**

| Message | 19.2 kbps | 128 kbps | 4 Mbps | 24 Mbps |
|---|---|---|---|---|
| INVITE | 17 | 3 | 1 | 1 |
| SIP 183 | 6 | 1 | 1 | 1 |
| SIP 180 | 6 | 1 | 1 | 1 |
| PRACK | 6 | 1 | 1 | 1 |
| SIP 100 | 6 | 1 | 1 | 1 |
| UPDATE | 6 | 1 | 1 | 1 |
| 200 OK | 3 | 1 | 1 | 1 |
| SIP ACK | 2 | 1 | 1 | 1 |

been selected and hence, our results will give an upper bound on the analyzed delays which is of interest to most network analysts. In our analysis, we use SigComp for SIP messages compression. It has been shown in [8] that SIP/SDP (Session Description Protocol) message sizes can be reduced by as much as 88% using SigComp with negligible compression/decompression time. The compression rate for the initial SIP messages such as INVITE has been chosen to be 55% and for the subsequent SIP messages to be 80%. The SIP message sizes have been selected according to [13] and the standard [17] and are shown in Table 1.

We consider 19.2 kbps and 128 kbps channel for 3G network; 4 Mbps and 24 Mbps channel for IEEE 802.16 WiMax. The RLP is used in 3G networks to improve the bit error rate (BER) performance whereas it is not considered for WiMax network due to much higher bandwidth [2]. We need to calculate the values of $K$ i.e. the number of frames in a packet, for the above mentioned channel rates. The RLP frame duration or inter-frame time $\tau$ is assumed to be 20 ms for 3G access network (AN) [4]. We proceed as follows for calculation of value of $K$ for 19.2 kbps channel: Number of bytes in each frame $19.2 \times 10^3 \times 20 \times 10^{-3} \times \frac{1}{8} = 48$ bytes. For SIP INVITE message, and 19.2 kbps channel; the value of K comes out to be $\lceil \frac{810}{48} \rceil = 17$. The WiMax frame duration and inter-frame time is assumed to be 2.5 ms and is independent of the channel bit rate [7]. For 4 Mbps channel: Number of bytes in each frame $4 \times 10^6 \times 2.5 \times 10^{-3} \times \frac{1}{8} = 1250$ bytes. For SIP INVITE message, and 4 Mbps channel; the value of K comes out to be $\lceil \frac{810}{1250} \rceil = 1$. The values of $K$ obtained for different messages following the same methodology are shown in Table 2. We take care of the value of $K$ for a particular signaling message in our analysis.

For IMS session setup, 12 message exchanges are involved between SN and proxy-call session control function (P-CSCF) of the visited IMS network and 12 message exchanges are involved between P-CSCF of the terminating IMS network and CN [1], [3] as shown in Figure 1. When SN is in UMTS network and CN is in WiMax and vice versa, the IMS session setup transmission delay $D_{trans-uw/wu}$ in seconds is given by:

$$D_{trans-uw/wu} = 12 \times D_{RLP} + 12 \times D_{noRLP} \qquad (2)$$

where $D_{RLP}$ and $D_{noRLP}$ denote the average delay for successfully transmitting a TCP segment with no more than $N_{maxTCP}$ retransmission trials with and without RLP operating underneath, respectively and are derived in [4]. When SN as well as CN are in 3G network, the IMS session setup transmission delay $D_{trans-uu}$ in seconds is given as:

$$D_{trans-uu} = 24 \times D_{RLP} \qquad (3)$$

When SN as well as CN reside in WiMax, the IMS session setup transmission delay $D_{trans-ww}$ in seconds is given by:

$$D_{trans-ww} = 24 \times D_{noRLP} \qquad (4)$$

### 2.2. Processing Delay

We calculate the processing delay for different entities in the IMS signaling path. The processing delay for some of the nodes, such as the home subscriber server (HSS), mainly consists of the address lookup table delay. When a query is sent to HSS for a particular IP address, the HSS has to lookup its table for the given IP address. We assume that the HSS table contains the list of all the users $N$ in the network. The IP address lookup is key component in the processing delay for network databases. It has been shown that cache line size can be used to help in multiway search; and binary search can be adapted to perform multiple-column search for long length IP addresses [9]. For rest of the network entities, we assume a fixed processing delay $d_{proc-ed}$ mainly consisting of the delay involved in the encapsulation and decapsulation of packets. The processing delay in nanoseconds at the HSS $d_{proc-hss}$ can be approximated as:

$$d_{proc-hss} = d_{proc-ed} + 100\left(\log_{k+1} N + \frac{L}{S}\right) \text{ ns} \quad (5)$$

where $L$ is the IP address length in bits e.g. $L$ is 32 for IPv4 and 128 for IPv6, $S$ is the machine word size in bits, $k$ is a system-dependent constant, and $d_{proc-ed}$ represents the fixed processing delay due to the encapsulation and decapsulation of packets. We have used the multiplication factor of 100 ns in the above equation because it has been shown in [9] that the lookup time is increased by around 100 ns for each memory access.

The processing delay for the IMS session setup $D_{proc}$ in seconds can be given as:

$$\begin{aligned} D_{proc} &= 7d_{proc-sn} + 24d_{proc-pcscf} + 24d_{proc-scscf} \\ &\quad + 6d_{proc-icscf} + d_{proc-hss} + 5d_{proc-cn} \quad (6) \end{aligned}$$

where $d_{proc-sn}$, $d_{proc-pcscf}$, $d_{proc-scscf}$, $d_{proc-icscf}$, and $d_{proc-cn}$ denote the unit packet processing delay at SN, P-CSCF, serving-call session control function (S-CSCF), interrogating-call session control function (I-CSCF), and CN respectively. The $D_{proc}$ is calculated in seconds after the conversion of each of its constituent delays in seconds from nanoseconds. The coefficients in equation (6) are determined based on the number of messages each network entity has to process and can be verified from Figure 1. It is to be noted that processing is considered for received messages at a node and transmitted messages from a node are taken into processing cost at the immediate receiving node.

### 2.3. Queueing Delay

We calculate the queueing delays for different network entities involved in the IMS signaling. The packet delay to reach from SN to CN depends on the queueing delay at each of the intervening queues which itself depends upon the number of packets at each queue. We have assumed M/M/1 queues for the network entities and Poisson signaling arrival rate process. For a queueing network with M/M/1 queues in tandem, if the input process to the first M/M/1 queue is Poisson, the input process to the next stage M/M/1 queue is also Poisson and independent of the input process and so on [6], [12]. The expected total waiting time or delay in the queueing network consisting of queues in tandem is the sum of the expected waiting times at each queue. For more realistic network model, M/M/1/B queues can be assumed for the network entities which model finite buffer system queues. However, we have not considered M/M/1/B queuing network model for the sake of simplicity.

The queueing delay for the IMS session setup $D_{queue}$ in seconds can be given as:

$$\begin{aligned} D_{queue} &= 7E[w_{sn}] + 24E[w_{pcscf}] + 24E[w_{scscf}] \\ &\quad + 6E[w_{icscf}] + E[w_{hss}] + 5E[w_{cn}] \quad (7) \end{aligned}$$

where $E[w_{sn}]$, $E[w_{pcscf}]$, $E[w_{scscf}]$, $E[w_{icscf}]$, $E[w_{hss}]$, and $E[w_{cn}]$ denotes the expected value of a unit packet queueing delay at SN, P-CSCF, S-CSCF, I-CSCF, HSS, and CN, respectively. Again, we have considered queuing delay at the receive buffer only and assuming that no delay is encountered at transmission buffer at a network node. The coefficients in equation (7) based on our assumption can be verified from Figure 1. The expected waiting time or delay of a packet at SN queue is given by [12]:

$$E[w_{sn}] = \frac{\rho_{sn}}{\mu_{sn}(1-\rho_{sn})} \quad (8)$$

where $\rho_{sn} = \lambda_{e-sn}/\mu_{sn}$ represents the utilization at SN queue, $\mu_{sn}$ denotes the service rate at SN queue and $\lambda_{e-sn}$ represents the *effective arrival rate* (in packets per second) at SN queue. That is, $\lambda_{e-sn} = \sum_{i \in N_{sn}} \lambda_i$, where $N_{sn}$ denotes the number of active sessions including the considered IMS session. The effective arrival rate $\lambda_e$ at a network node can be determined from the utilization at that node. Similarly, the $\lambda_e$ at queues of other network nodes can be calculated and expressions can be determined for the expected waiting time at other network entities.

### 2.4. Total Delay for IMS Session Setup

We calculate the total delay for IMS session establishment procedure. The delay for IMS session setup when SN is in UMTS network and CN is in WiMax and vice versa is given by:

$$D_{total-uw/wu} = D_{trans-uw/wu} + D_{proc} + D_{queue} \quad (9)$$

The delay for IMS session setup when SN as well as CN are in 3G network is given by:

$$D_{total-uu} = D_{trans-uu} + D_{proc} + D_{queue} \quad (10)$$

The delay for IMS session setup when SN as well as CN are in WiMax is given as:

$$D_{total-ww} = D_{trans-ww} + D_{proc} + D_{queue} \quad (11)$$

## 3. Numerical Results

In this section, we present the numerical results for the delay analysis of SIP-based signaling for IMS sessions. The parameter values selected for the analysis are mentioned hereafter. The value of end-to-end frame propagation delay $D$ for 19.2 kbps and 128 kbps channel is taken equal to 100 ms whereas for 4 Mbps and 24 Mbps channel, the valued of $D$ is chosen to be 0.27 ms and 0.049 ms, respectively [2]. Frame duration $T$ as well as inter-frame time $\tau$ is assumed to be 20 ms for 3G AN [4]. WiMax frame duration

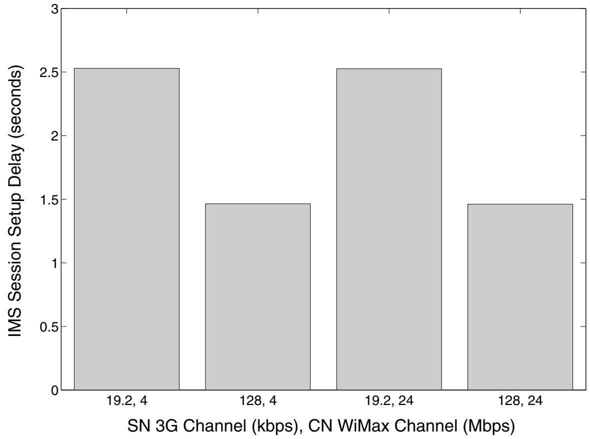


**Figure 2. IMS session setup delay for different channel rates when SN is in 3G network and CN is in WiMax for fixed $\lambda$ and $p$.**

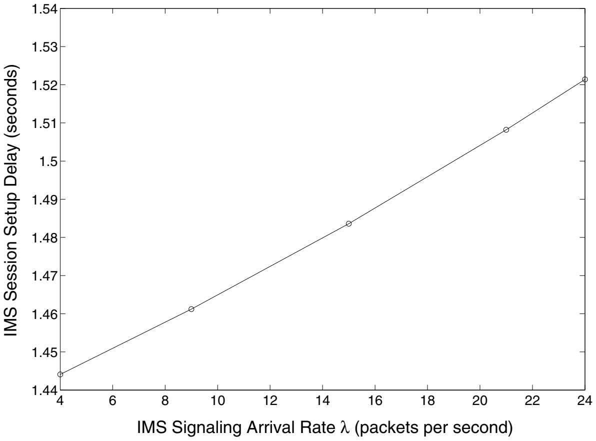


**Figure 3. Effect of changing arrival rate $\lambda$ on IMS session setup delay when SN is in 128 kbps 3G network and CN is in 24 Mbps WiMax for fixed $p$.**

and inter-frame time is assumed to be 2.5 ms and is independent of the channel bit rate [7]. We consider a network containing two 3G base station controllers (BSCs) and three WiMax BSCs. The cell radius for 3G BSC and WiMax BSC is taken to be 1000 m and 700 m, respectively. The user densities in the two networks is taken to be 0.001 per square meter [10], [11], [14]. The number of users resulting from the selection of these cell radii and user densities in 3G cellular and WiMax ANs are $N_{mn1}$ = 5000, and $N_{mn2}$ = 3000, respectively. The IP address length $L$ and processor machine word size $S$ are taken to be 32 bits. The system dependent constant value $k$ is equal to 5 [9]. The maximum number of RLP retransmissions $n$ and maximum number of TCP retransmissions $N_{maxTCP}$ are both taken equal to 3 [4], [5]. The service rate $\mu$ at all the network entities is taken equal to 250 packets/sec. The unit packet processing delay for all the network entities is taken equal to $4 \times 10^{-3}$ seconds [10], [11]. The background utilization due to traffic from other sources is taken to be 0.7 for HSS because it has to handle traffic for inter-system communications from different ANs, and 0.4 for the rest of the entities. The assumption of these values of background utilizations allows us to determine $\lambda_e$ at each of the network nodes. It is worth mentioning that final transmission, processing and queuing delays and hence the total IMS session establishment delay is calculated in seconds after the appropriate conversion of units.

In the first set of experiments, the IMS session setup delay is analyzed for the channel rates of 19.2 kbps and 128 kbps in 3G network; and 4 Mbps and 24 Mbps in WiMax network. For this set of experiments, the probability of a frame being in error $p$ and the IMS signaling arrival rate $\lambda$ in packets per second is kept constant equal to 0.02 and 9, respectively. Figure 2 shows the IMS session setup delay when the SN is in 3G network and CN is in WiMax for different combinations of 3G and WiMax channel rates. It can be observed that the IMS session setup delay is greatly effected by the 3G channel rate i.e. IMS session setup delay decreases considerably as the 3G channel rate increases. Another interesting observation is that the IMS session setup delay is negligibly effected by changing the WiMax channel rate. Similarly, results can be obtained for the cases when both SN as well as CN are in 3G network and in WiMax network based on our analytical model.

In the second set of experiments, the effect of changing IMS signaling arrival rate $\lambda$ in packets per second on IMS session setup delay is analyzed. The frame error probability $p$ is kept constant at 0.02. The results are calculated for $\lambda$ equals to 4, 9, 15, 21, and 24 packets per second. The channel considered for 3G network is 128 kbps and for WiMax is 24 Mbps. Figure 3 shows the effect of increasing arrival rate on IMS session setup delay when SN is in 3G network and CN is in WiMax. It can be seen that the IMS session setup delay increases considerably with the increasing arrival rate.

In the third set of experiments, the effect of changing frame error probability $p$ on IMS session setup delay is analyzed. The arrival rate $\lambda$ is kept constant at 9 packets per second for this set of experiments. The results are calculated for $p$ equals to 0.01, 0.02, 0.05, 0.1, and 0.2. The channel considered for 3G network is 128 kbps and for WiMax is

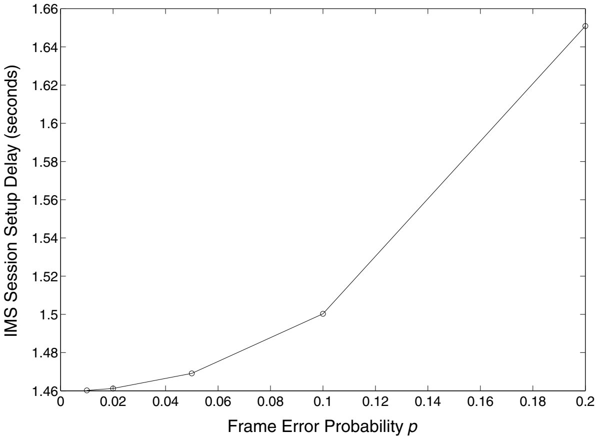


**Figure 4. Effect of changing frame error probability $p$ on IMS session setup delay when SN is in 128 kbps 3G network and CN is in 24 Mbps WiMax for fixed $\lambda$.**

24 Mbps. Figure 4 shows the effect of increasing frame error probability on the IMS session setup delay. It can be observed that the IMS session setup delay increases slowly with the increasing frame error probability.

## 4. Conclusions

In this paper, we analyzed the SIP-based IMS signaling delay for IMS session establishment procedure in 3G and WiMax access networks. The delay analysis is comprehensive since it takes into account transmission, processing, and queueing delays at the network entities. Numerical results indicate that increasing the 3G channel rate can significantly decrease the IMS session setup delay.Also, the IMS session setup delay increases with the increase in the signaling arrival rate and frame error probability.

## Acknowledgment

The initial part of the work was supported by Bell Canada and Natural Sciences and Engineering Research Council of Canada (NSERC) under the supervision of Dr. Vincent Wong. The later part is supported by Mentor Graphics Embedded Systems Division.

## References

[1] 3GPP. Signalling flows for the IP multimedia call control based on Session Initiation Protocol (SIP) and Session Description Protocol (SDP); Stage 3 (Release 5). TS 24.228 (v5.15.0), September 2006.

[2] N. Banerjee, W. Wu, K. Basu, and S. Das. Analysis of SIP-based Mobility Management in 4G Wireless Networks. *Elsevier Computer Communications*, 27(8):697–707, May 2004.

[3] G. Camarillo and M.-A. Garcia-Martin. *The 3G IP Multimedia Subsystem (IMS): Merging the Internet and the Cellular Worlds*. John Wiley and Sons, 2004.

[4] S. Das, E.Lee, K. Basu, and S. Sen. Performance Optimization of VoIP Calls over Wireless Links using H.323 Protocol. *IEEE Transactions on Computers*, 52(6):742–752, June 2003.

[5] H. Fathi, S. Chakraborty, and R. Prasad. Optimization of SIP Session Setup Delay for VoIP in 3G Wireless Networks. *IEEE Transactions on Mobile Computing*, 5(9):1121–1132, September 2006.

[6] E. Gelenbe and G. Pujolle. *Introduction to Queueing Networks*. John Wiley and Sons, 1998.

[7] C. Hoymann, K. Klagges, and M. Schinnenburg. Multihop Communication in Relay Enhanced IEEE 802.16 Networks. In *Proc. of 17th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Helsinki, Finland, September 2006.

[8] H. Jin and A. Mahendran. Using SigComp to Compress SIP/SDP Messages. In *Proc. of IEEE International Conference on Communications (ICC'05)*, Seoul, Korea, May 2005.

[9] B. Lampson, V. Srinivasan, and G. Varghese. IP Lookups Using Multiway and Multicolumn Search. *IEEE/ACM Transactions on Networking*, 7(3):324–334, June 1999.

[10] C. Liu and C. Zhou. An Improved Interworking Architecture for UMTS-WLAN Tight Coupling. In *Proc. of IEEE Wireless Communications and Networking Conference (WCNC'05)*, Atlanta, Georgia, March 2005.

[11] C. Liu and C. Zhou. HCRAS: A Novel Hybrid Internetworking Architecture between WLAN and UMTS Cellular Networks. In *Proc. of IEEE Consumer Communications and Networking Conference (CCNC'05)*, Las Vegas, Nevada, January 2005.

[12] J. Medhi. *Stochastic Models in Queueing Theory*. Academic Press, An imprint of Elsevier Science, 2003.

[13] M. Melnyk and A. Jukan. On Signaling Efficiency for Call Setup in all-IP Wireless Networks. In *Proc. of IEEE International Conference on Communications (ICC'06)*, Istanbul, Turkey, June 2006.

[14] S. Mohanty and J. Xie. Performance Analysis of a Novel Architecture to Integrate Heterogeneous Wireless Systems. *Elsevier Computer Networks*, 51(4):1095–1105, March 2007.

[15] N. Nakajima, A. Dutta, S. Das, and H. Schulzrinne. Handoff Delay Analysis and Measurement for SIP based mobility in IPv6. In *Proc. of IEEE International Conference on Communications (ICC'03)*, Anchorage, Alaska, May 2003.

[16] R. Price, C. Bormann, J. Christoffersson, H. Hannu, Z. Liu, and J. Rosenberg. Signaling Compression (SigComp). RFC 3320, January 2003.

[17] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler. SIP: Session Initiation Protocol. RFC 3261, June 2002.