Analysis of Signaling Cost for a Roaming User in a Heterogeneous Mobile Data Network

Kumudu S. Munasinghe and Abbas Jamalipour
School of Electrical and Information Engineering
University of Sydney, NSW 2006, Australia
kumudu@ee.usyd.edu.au, a.jamalipour@ieee.org

Abstract—**In an all-IP environment of internetworked heterogeneous mobile data networks, ongoing data sessions from roaming users are subjected to frequent vertical handoffs. Under such circumstances, careful consideration must be given for the selection of appropriate vertical handoff mechanisms to ensure seamless service continuity and desired Quality of Service (QoS) levels. Therefore, efficient methods for evaluating and comparing such techniques are essential. One such evaluation technique is signaling cost analysis. This paper presents an analytical model for evaluating signaling cost of vertical handoffs in a heterogeneous mobile networking environment at the core network level for a roaming user. The numerical analysis and evaluation is based on a framework designed for interworking between Universal Mobile Telecommunications System (UMTS), CDMA2000 technology, and mobile WiMAX (Worldwide interoperability for Microwave Access) Networks. Results and analysis illustrate the behavior of the signaling cost metric against session arrival rate, network mobility rate, and the call-to-mobility rate.**

Keywords- **IMS; UMTS; WiMAX; SIP; Mobile IP; Mobility Management; Signaling Cost, Queuing Networks**

I. INTRODUCTION

It is a well known fact that ubiquitous data services and relatively high data rates across heterogeneous data networks could be achieved by interworking 3G cellular networks with Broadband Wireless Access (BWA) technologies (say, WiMAX) [1]. This will enable a user to access 3G cellular services via a WiMAX network, when roaming within the coverage of a WiMAX network and vice-versa. Thus BWA networks can be considered as a complementary technology for 3G cellular data networks and will eventually become a compulsory element of the future all-IP Next Generation Mobile Network (NGMN) [2].

When such an interworked heterogeneous networking environment is considered, ongoing data sessions are frequently subjected to vertical handoffs. Therefore, it is important for appropriate mechanisms capable of seamless session handoff with guaranteed QoS to be in place. This is an ongoing area of research with a wealth of resources available [3], [4]. On the other hand, in the process of selecting an appropriate vertical handoff mechanism for a roaming user, efficient techniques for evaluating cost effective vertical handoff techniques are essential. One such evaluation method is the analysis of the signaling cost. The concept of using the signaling cost for evaluating mobility management during handoffs for Mobile IP (MIP) [5] and Session Initiation Protocol (SIP) [6] sessions have initially been explored in [7]. This works has further been extended in [8] by considering location update, and session setup procedures in the analysis. Much recently, contributions to the signaling cost estimation mechanism were made by additionally considering the probability of a handoff scenario taking place [9]. However, according to our knowledge, the signaling cost evaluation method for a vertical handoff scenario for a roaming user in a heterogeneous networking environment is unexplored and yet to be resolved, which motivates the following contribution.

Therefore, the key contribution of this paper lies in presenting a novel analytical model for evaluating the signaling cost of vertical handoffs for a roaming user at the core network level in a heterogeneous mobile networking environment. The analysis investigates the behavior of the signaling cost metric against data session arrival rate, network mobility rate, and the call-to-mobility rate. The vertical handoff mechanism and heterogeneous networking platform used for the analysis is based on an authors' previous contribution [10]. The significance of our proposed heterogeneous networking platform is that it uses a novel approach, that is, the use of the 3GPP's IP Multimedia Subsystem (IMS) for supporting real-time session negotiation and management [11]. The remainder of this paper is organized as follows. The next section briefly introduces the heterogeneous networking platform and the vertical handoff mechanism used for signaling cost analysis. Followed by are the sections on signaling cost analysis and numerical results prior to the concluding remarks.

II. MOBILITY AND VERTICAL HANDOFF MANAGEMENT IN HETEROGENEOUS MOBILE NETWORKS

The interworking platform used in this analysis is illustrated in Fig. 1. Interested readers may refer to [10] for a more specific and detailed information on its architectural design. One of the primary design considerations of this architecture worth noting is that all networks are loosely coupled for data routing and tightly (or centrally) coupled at the IMS for control signaling. Therefore, session mobility management is facilitated via the IMS at the application layer. In order to guarantee terminal mobility, MIPv4 has also been implemented at the IP layer.

An UMTS core network is connected to the all-IP network through the Gateway GPRS Support Node (GGSN), which also acts as its MIP Foreign Agent (FA). Once the system

978-1-4244-2324-8/08/$25.00 © 2008 IEEE.

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE "GLOBECOM" 2008 proceedings.

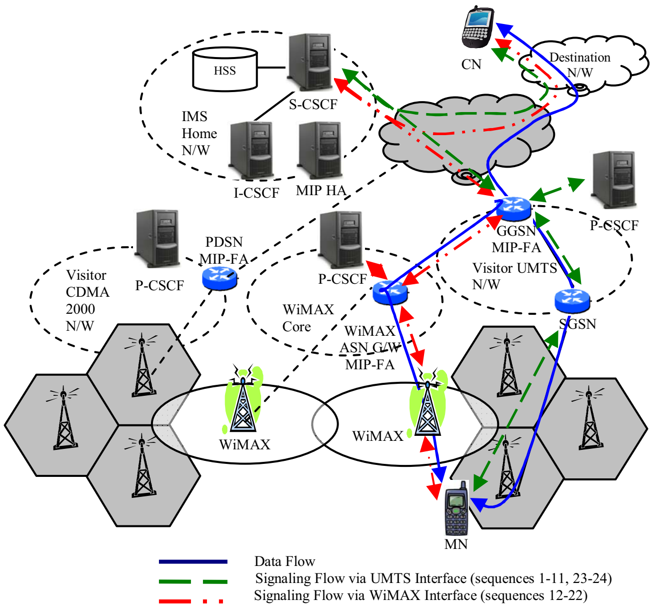

Fig. 1. The Proposed Interworking Architecture.

acquisition is done by a Mobile Node (MN) connected to the UMTS network, the next step is to set up a data pipeline. The actual IP address allocation for the MN is initiated by sending the MIP registration request to its Home Agent (HA) via the GGSN (i.e., the MIP-FA). This mechanism is based on the specifications given under [12]. The MN acts as an IMS-SIP client and sends a SIP registration message to its home system through the Proxy-Call Session Control Function (P-CSCF). Once authorized, a suitable Serving-Call Session Control Function (S-CSCF) for the MN is assigned and its subscriber profile is sent to the designated S-CSCF.

After the activation of the PDP context and the service registration, the MN is ready to establish a media/data/call session. As illustrated in Fig. 2, the sequence of the SIP session origination procedure can be described as follows. The mobile origination procedure is initiated by a SIP INVITE message sent from the UMTS interface of the source MN (step 4). This initial message is forwarded from the P-CSCF in the UMTS core network to the S-CSCF of the originating (or Home) network, via the IMS elements of the terminating network, and finally to the destination. This SIP INVITE carries a request to follow the precondition call flow model. This is important because some clients require certain preconditions (that is, QoS levels) to be met before establishing a session. Next, this model requires that the destination responds with a 183 Session Progress containing a SDP answer (step 5).

The acknowledgement for the reception of this provisional response by a PRecondition ACKnowledgment (PRACK) request follows afterwards (step 6). When the PRACK request successfully reaches the destination a 200 OK response is generated by the destination with an SDP answer (step 7). Next an UPDATE request is sent by the source containing another SDP offer, in which the source indicates that the resources are reserved at his local segment (step 8). Once the destination receives the UPDATE request, it generates a 200 OK response (step 9). Once this is done, the MN can start the media/data flow and the session will be in progress (via the UMTS interface).

When this MN roams between WiMAX and UMTS systems (say), inter-network roaming takes place. The message flow for an inter-network roaming (i.e., for a vertical handoff) from UMTS to WiMAX can be described as follows. Firstly the standard WiMAX link layer access registration procedures are performed. Next the WiMAX interface performs the MIP registration procedures with the ASN Gateway (MIP FA) as explained previously (steps 13-14). This is when the ASN Gateway (MIP-FA) forwards this request (via the CSN) to the MIP-HA and the HA assigns the home IP address to the new WiMAX interface. Lastly the exchanging of a MIP Binding Update message between the MN and the CN for avoiding triangular routing (step 15) [5].

The next stage is the taking place of the IMS-SIP session handoff procedures. This requires sending a SIP Re-INVITE

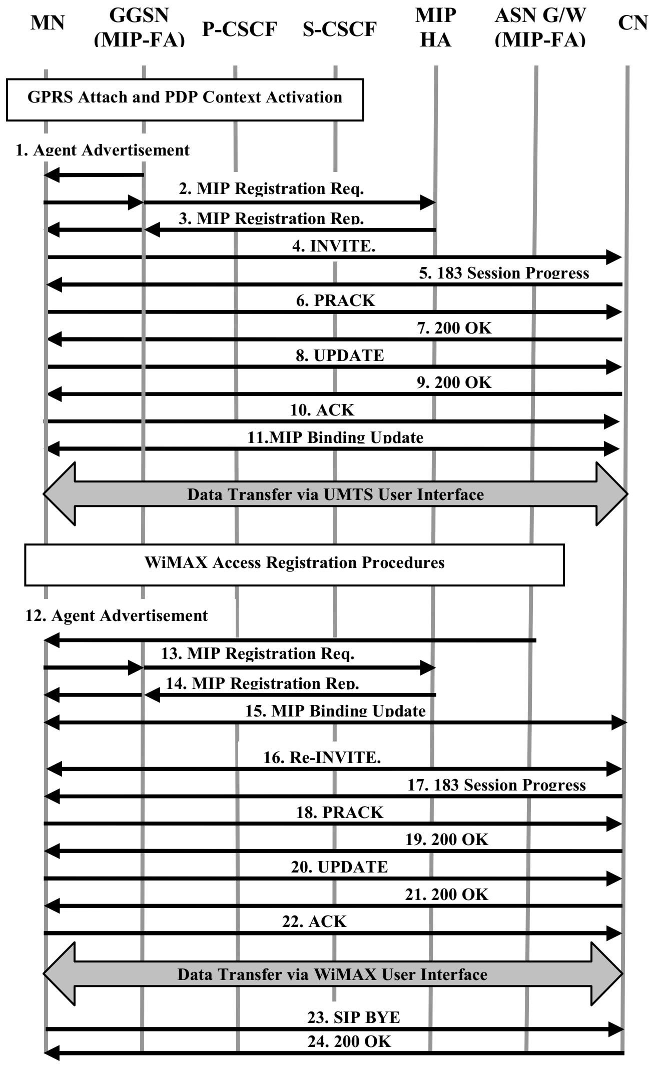

Fig. 2. Vertical Handoff Signaling.

(with same Call-ID and other identifiers corresponding to the ongoing session) to the destination (step 16). Followed by this is a resource/preconditions reservation for the WiMAX interface. Once this is successfully done the new session flow can be initiated. It is important to note that until such time that the new data flow is initiated via the WiMAX interface, the data flow via the UMTS interface remains active. Thus the model follows the make-before-break handoff mechanism as proposed in our previous works [13]. Inter-system roaming from WiMAX to UMTS can also take place in a similar manner. Furthermore, since this design is an extension to our WLAN-UMTS-CDMA200 interworking platform, WiMAX-CDMA2000 roaming can also be accommodated within this architecture in a similar manner.

III. SIGNALING COST ANALYSIS

The resultant signaling cost of mobility management during vertical handoff can be analyzed as follows. The primary assumption made in this analysis is that the session arrivals follow a Poisson process. Thus an essential defining characteristic for all user initiated sessions (e.g., VoIP) is that, sessions are mutually independent. Therefore, as sighted in [14], for large populations, where each user is independently contributing a small portion of the overall traffic, human initiated call/data sessions can be assumed to follow a Poisson arrival process [15]. Based on the traces of wide-area traffic, there is further evidence that Poisson arrivals appear to be suitable for traffic at the session level when sessions are human initiated [16]. The signaling cost or overhead is the accumulative traffic load on exchanging signaling messages during the MN's communication session. Therefore the signaling cost incurred by a message can be defined as:

$$Cost = P \times S_{message} \times H_{a-b} \tag{1}$$

where, P is the probability that each handoff will occur, $S_{message}$ is the av. size of a signaling message, H_{a-b} is the av. number of hops between a and b.

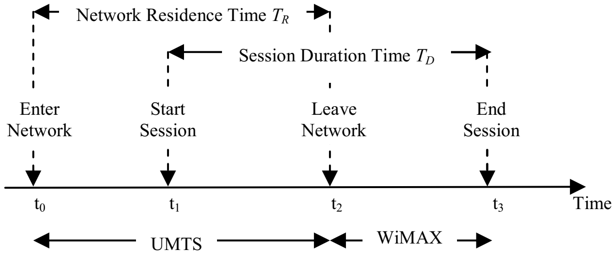

Fig.3. Timing Diagram.

When a MN moves from one network (say, from UMTS) to another (say, to WiMAX), the following condition must be satisfied for a successful vertical session handoff to take place [9]. That is, a data session that starts from the MN's current network must remain active until it has moved out of this network, as shown in Fig. 3. Based on this condition, the arrival probability of a session that is likely to be subjected to a vertical handoff (P_1) for an inter-network roaming MN can be derived. Let's assume that session arrivals follow Poisson process with the av. arrival rate λ, thus the probability that there is one session arrival in a time period t becomes $\lambda t e^{-\lambda t}$ [17]. Hence, P_1, the session arrival probability for an inter-network roaming MN between t_0 and t_2, can be expressed as:

$$P_1(t_0 \le t_1 \le t_0 + T_R) = \int_0^\infty \lambda t e^{-\lambda t} f_{T_R}(t)\,dt \tag{2}$$

where, $f_{T_R}(t)$ is the probability density function (*pdf*) of the network residence time T_R. It is assumed that the residence time of the MN in a given network, T_R, is exponentially distributed with a mean $1/\eta$, where η is the inter-network mobility rate. Hence $f_{T_R}(t)$ can be expressed as:

$$f_{T_R}(t) = \eta e^{-\eta t} \tag{3}$$

By substituting $f_{T_R}(t)$ in (2) and solving the equation P_1 can be derived as:

$$P_1(t_0 \le t_1 \le t_0 + T_R) = \int_0^\infty \lambda t e^{-\lambda t} \eta e^{-\eta t}\,dt$$

$$P_1 = \frac{\lambda \eta}{(\lambda + \eta)^2} \tag{4}$$

The second condition is that, for a vertical handoff to take place, the session duration time T_D must be greater than network residence time T_R. It is assumed that the session duration time, T_D, is exponentially distributed with a mean $1/\mu$, where μ is the mean message service rate of the (session initiating) wireless link (i.e., UMTS in this case). Hence the *pdf* of T_D, $f_{T_D}(t)$ can be expressed as:

$$f_{T_D}(t) = \mu e^{-\mu t} \tag{5}$$

Therefore the probability P_2 for this condition, which is vertical session handoff probability, can be expressed as:

$$P_2(T_D \rangle T_R) = \int_0^\infty \int_t^\infty \mu e^{-\mu y} f_{T_R}(t)\,dy\,dt$$

$$P_2(T_D \rangle T_R) = \int_0^\infty \int_t^\infty \mu e^{-\mu y} \eta e^{-\eta t}\,dy\,dt$$

$$P_2 = \frac{\eta}{(\mu + \eta)} \tag{6}$$

For clarity and convenience sake, the units for μ (i.e., the mean message processing/service rate of the session initiating wireless link) are changed from packets/sec to bits/sec. If the *pdf* of packet size, x, in bits be $\mu_{wl} t e^{-\mu_{wl} x}$ with a mean packet length of $1/\mu_{wl}$ bits/packet, and the capacity of the communication channel i be C_i bits/sec. The product $\mu_{wl}C_i$ is then the service rate in packets/sec. Therefore, $\mu_{wl}C_i$ is substituted for μ in (6). Thus P_2 can be finally expressed as:

$$P_2 = \frac{\eta}{(\mu_{wl} C_i + \eta)} \tag{7}$$

Therefore, the total signaling cost for the scenario in Fig. 3 can be expressed as the sum of the two individual signaling costs associated with the above two conditions. Thus, P_1 and P_2 are substituted for P in the generalized expression given in (1) for calculating these individual signaling costs. Additionally, since the signaling cost is calculated for a roaming MN, av. network mobility rate *(η)* and av. session arrival rate *(λ)* components also contribute to the final equation. The SIP INVITE message sequence (steps 2-11 from Fig. 2) is associated with P_1, the session arrival probability and session arrival rate. Similarly, the SIP Re-INVITE message sequence (steps 13-22 from Fig. 2) is associated with P_2, the vertical handoff probability and inter-network mobility rate. Hence, the total signaling cost incurred by the vertical handoff can be expressed as:

$$Cost = P_1\lambda\sum_{i=1}^{n_1}\left(S_{Invite-i}\times H_{(a-b)-i}\right)+P_2\eta\sum_{i-1}^{n_2}\left(S_{\mathrm{Re}\,Invite-i}\times H_{(a-b)-i}\right) \quad (8)$$

where n_1 and n_2 represent the number of messages involved in each handoff/message sequence. If η is the av. network mobility rate of a MN and λ is the av. session arrival rate, λ/η may be defined as the Call-to-Mobility Rate (*CMR*). Thus (8) can be re-arranged as:

$$Cost = \left[P_1\eta\sum_{i=1}^{n_1}\left(S_{Invite-i}\times H_{(a-b)-i}\right)\right]\frac{\lambda}{\eta}+P_2\eta\sum_{i=1}^{n_2}\left(S_{\mathrm{Re}\,Invite-i}\times H_{(a-b)-i}\right) \quad (9)$$

$$Cost = \left[P_1\eta\sum_{i=1}^{n_1}\left(S_{Invite-i}\times H_{(a-b)-i}\right)\right]CMR+P_2\eta\sum_{i=1}^{n_2}\left(S_{\mathrm{Re}\,Invite-i}\times H_{(a-b)-i}\right) \quad (10)$$

IV. NUMERICAL RESULTS

This section presents numerical results relating to the behavior of signaling cost analysis against av. session arrival rate (*λ*) *calls/min* and av. network mobility rate (*η*) *min*$^{-1}$. Table I provides the typical MIPv4 and SIP message sizes and Fig. 4 provides the relative distances in hops used for the numerical evaluation. Fig. 5 illustrates the behavior of normalized signaling cost versus av. session arrival rate (*λ*) for different av. session mobility rate (*η*) values and a constant service rate. According to the graphs in Fig. 5, it is clear that the signaling cost increases against *λ* for increasing values of *η*.

TABLE I
MESSAGE SIZES AND PARAMETER VALUES

Message	Size (Bytes)	Message	Size (Bytes)
INVITE	736	BYE	550
Re-INVITE	731	MIP Agent Ad.	28
183 Ses. Pro.	847	MIP Reg. Req.	60
PRACK	571	MIP Reg. Rep.	56
200 OK	558	MIP BU	66
UPDATE	546	MIP BACK	66
ACK	314	C_i	2–70 Mbps

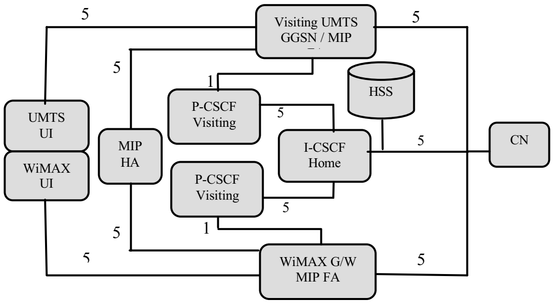

Fig. 4. Relative distances in hops.

In general, this is due to the increase of session arrival probability (P_1) with the increase of *λ*. Nevertheless, it is interesting to observe the behavior of P_1 against *λ* for different *η* values (Fig. 6) to exactly find out the reasons for different gradient levels of the three graphs in Fig. 5. Therefore, according to Fig. 6, for *λ* ranging from 0.01-0.1 *calls/min*, when $\eta = 0.01$ *min*$^{-1}$, a negative slope is observed. This negative gradient has contributed towards slowing down the increase of signaling cost against increasing *λ* within the considered range. Similarly, in Fig. 6, for the same *λ* range, when $\eta = 0.1$ *min*$^{-1}$ a positive gradient is observed.

This contributes to a rather rapidly increasing signaling cost for the graph corresponding to $\eta = 0.1$ *min*$^{-1}$ in Fig. 5. Since P_2 remains constant for a given *η*, it does not impose a dramatic effect in this case. It is also important to note that the range of *λ* is kept below $\mu_{wl}C_i$, the service rate. This is because, in the event that *λ* reaches $\mu_{wl}C_i$, the utilization of the system will rapidly reach 100%. Thus the graphs in Fig. 5 are only plotted up to $\lambda = 0.1$ *calls/min*.

Fig. 7 illustrates the behavior of normalized signaling cost against av. network mobility rate (*η*) for different av. session arrival rate (*λ*) values and a constant service rate. In this case, normalized signaling cost generally increases as (*η*) increases. Also this increase becomes rapid for values of (*η*) ranging from 1-10 *min*$^{-1}$. The reason for such behavior is that as the network mobility rate (*η*) increases, more sessions can be subjected to vertical handoff, which eventually increases the session handoff probability (P_2) giving rise to the signaling cost.

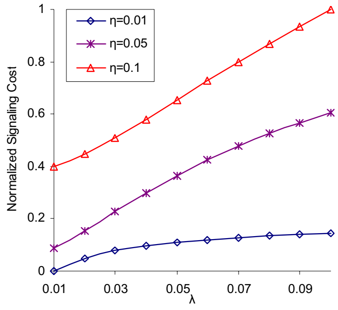

Fig. 5. Signaling Cost vs. Av. Session Arr. Rate *(λ calls/min)*.

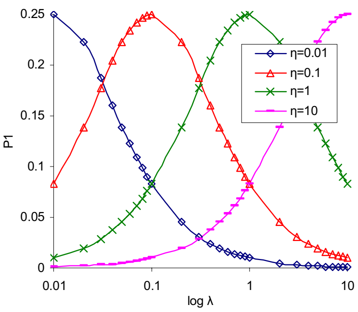

Fig. 6. P_1 vs. Av. Session Arrival Rate (*λ calls/min*).

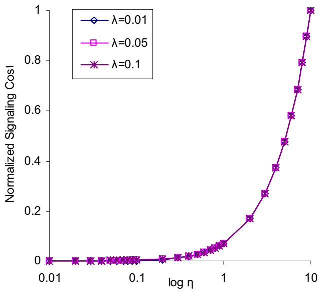

Fig. 7. Signaling Cost vs. Av. Net. Mob. Rate (η min^{-1}).

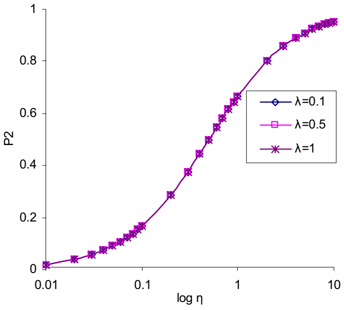

Fig. 8. P_2 vs. Av. Net. Mobility Rate (η min^{-1}).

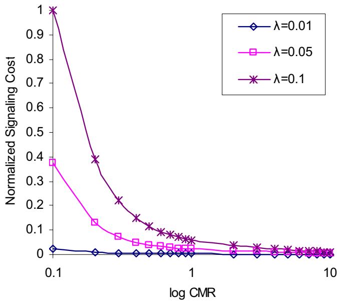

Fig. 9. Normalized Signaling Cost vs. CMR (λ *constant*).

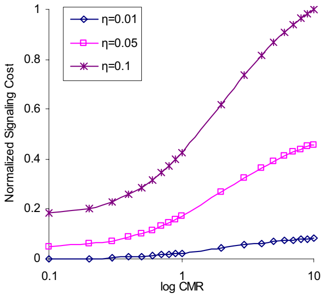

Fig. 10. Normalized Signaling Cost vs. CMR (η *constant*).

For comparison purposes, the P_2 curve against increasing network mobility rate is illustrated in Fig. 8. According to Fig. 8, it is clear how P_2 approaches 1 with the increase of η. The effect of P_1 in this case is relatively minimal since P_1 does not exceed 0.25, much similarly to the pattern indicated in Fig. 6 (i.e., since P_1 curves behaves the same despite λ and η being interchanged).

The next investigation is the behavior of signaling cost against the CMR. Fig. 9 illustrates normalized signaling cost against the CMR by having λ as a constant. As per the illustration of the graphs in Fig. 9, the normalized signaling cost reduces exponentially as the CMR increases. As the CMR increases by keeping λ as a constant, η tends to decrease rapidly, which has a direct impact on P_2 that reduces it exponentially against increasing the CMR. However, the impact of decreasing η does not tend to have a drastic effect on P_1. P_1 shows a closely similar pattern as in Fig. 6 with a maximum peak of 0.25 for CMR=1. Hence, in this case, the signaling cost curve is shaped according to the behavior pattern of P2 as CMR increases. Also note that the signaling cost increases as λ increases.

Last but not least, Fig. 10 illustrates normalized signaling cost against the CMR by having η as a constant. As per the illustration of the graphs in Fig. 10, the normalized signaling cost increases as the CMR increases and eventually reaches a saturation point. As the CMR increases by keeping η as a constant, λ tends to increase rapidly, which eventually results in increasing signaling cost. As in the above case, P_1 behaves in a similar manner with a maximum peak of 0.25 for CMR=1. However, the notable point is that P_2 remains constant for a chosen η value, which eventually contributes towards shaping the signaling cost curves by reaching a saturation level. Furthermore, signaling cost also increases as η increases.

V. CONCLUSIONS

This paper presented a novel analytical model for evaluating the signaling cost of vertical handoffs for a roaming user in a heterogeneous mobile network environment at the core network level. Using the proposed analytical model, an in-depth analysis was performed for investigating the behavior of the signaling cost metric. Numerical results revealed interesting behavioral patterns for the signaling cost metric as data session arrival rates, network mobility rates, and call-to-mobility rates are varied accordingly. This analytical model could be successfully used for designing, evaluating and optimizing cost effective handoff mechanisms in heterogeneous mobile networking environments.

REFERENCES

[1] L. Ma, "The Competition and Cooperation of WiMAX, WLAN, and 3G," in *Proc. of IEEE International Conference on Mobile Technology, Applications and Systems*, Guangzhou, China, Nov. 2005.
[2] M. Buddhikot, G. Chandranmenon, S. Han, Y. W. Lee, S. Miller, and L. Salgarelli, "Integration of 802.11 and third-generation wireless data networks," in *Proc. IEEE INFOCOM*, San Franciso, California, Apr. 2003.
[3] K. D. Wong, M. Barton, B. Kim, V. K. Varma, S. Ramesh, G. Hayward, and J. A. Friedhoffer, "UMTS signaling over 802.11 wireless LAN," in *Proc. of 58th IEEE Vehicular Technology Conference*, 2003.
[4] V. K. Varma, S. Ramesh, K. D. Wong, M. Barton, G. Hayward, and J. A. Friedhoffer, "Mobility management in integrated UMTS/WLAN networks," in *Proc. of the IEEE ICC*, Anchorage, Alaska, 2003.
[5] C. Perkins, "IP Mobility Support for IPv4," IETF RFC 3344, 2002.
[6] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler, "SIP: Session Initiation Protocol," RFC 3261, 2002.
[7] Q. Wang and M. A. Abu-Rgheff, "Interacting mobile IP and SIP for efficient mobility support in all IP wireless networks," in *Proc. of IEE Int. Conf. on 3G Mobile Comm. Tech.*, London, UK, Oct. 2004.
[8] Q. Wang and M. A. Abu-Rgheff, "Mobility management architectures based on joint mobile IP and SIP protocols," *IEEE Wireless Communications*, vol. 13, pp. 68-76, Dec. 2006.
[9] U. Narayan and X. Jiang, "Signaling Cost Analysis of Handoffs in Mixed IPv4/IPv6 Mobile Environment," in *Proc. of IEEE Globecom*, Washington, D.C., Nov. 2007.
[10] K. S. Munasinghe and A. Jamalipour, "A Unified Mobility and Session Management Platform for Next Generation Mobile Networks," in *Proc. of IEEE Globecom*, Washington DC, Nov. 2007.
[11] 3GPP, "IP Multimedia Subsystem (IMS)," 3GPP TS 23.228 Version 6.10.0 Release 6, 2005.
[12] 3GPP, "Combined GSM and Mobile IP Mobility Handling in UMTS IP CN," 3GPP TR 23.923 v 3.0.0, 2000.
[13] K. Munasinghe and A. Jamalipour, "A 3GPP-IMS based Approach for Converging Next Generation Mobile Data Networks," in *Proc of the IEEE ICC*, Glasgow, UK, Jun. 2007.
[14] T. M. Chen, "Network Traffic Modeling," in *The handbook of computer networks*, H. Bidgoli, Ed.: Wiley, 2007.
[15] J. Roberts, "Internet Traffic, QoS and Pricing," *Proceedings of the IEEE*, vol. 92, pp. 1389-1399, 2004.
[16] V. Paxon and S. Floyd, "Wide area traffic: The failure of Poisson Modeling," *IEEE/ACM Transactions on Networking*, vol. 3, pp. 226-244, 1995.
[17] L. Kleinrock, *Queuing Systems Volume 1: Theory*: John Wiley & Sons, 1975.