Analysis of Vertical Session Handoff for Self-Similar Traffic 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 an appropriate vertical handoff mechanism to ensure seamless service continuity and desired Quality of Service (QoS) levels. Therefore, efficient methods for analyzing vertical handoffs for IP based data sessions are essential. This paper presents an analytical model for evaluating vertical session handoffs in a heterogeneous mobile networking environment by considering a *G/M/1* queuing model, where *G* is a Pareto distribution. This is since Internet traffic data indicate that long-tailed distributions serve as better models for packet inter-arrival times and service lengths. 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 of the analysis are compared against the results obtained from a classical *M/M/1* queuing model assuming Poisson arrivals and exponential service times.**

***Keywords*- Queuing Networks; Pareto Distribution, Poisson Distribution, *G/M/1* queue, *M/M/1* queue, IMS; UMTS; CDMA2000; WiMAX; SIP; Mobile IP; Mobility Management;**

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 (i.e., UMTS and CDMA2000 Systems) with Broadband Wireless Access networking technologies (BWA) (e.g., 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 may 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 Internet data traffic to/from roaming users are frequently subjected to vertical handoff. Therefore, it is important for appropriate mechanisms capable of seamless session handoff with guaranteed QoS to be in place. Hence, efficient methods for analyzing vertical handoffs for such Internet data traffic sessions also become highly desirable.

Recent studies on the Internet traffic have strongly indicated that the Poisson packet arrivals and exponential packet lengths studied in the classical queuing theory are basically inappropriate for modeling Internet traffic [3] [4] [5]. There is a wealth of resources in the area of analyzing and characterizing Internet traffic. According to these, it is a well accepted fact that long-tailed distributions serve as better models for packet inter-arrival times and service lengths for Internet traffic [6] [7] [8] [9]. However, according to our knowledge, a method for analyzing an Internet data session being subject to a vertical handoff by 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 an analytical model for analyzing vertical session handoffs in a heterogeneous mobile networking environment by considering a long-tailed distribution. The analysis studies a *Pareto/M/1* queue (i.e., Pareto rather than Poisson arrival rates) to model data sessions that are subjected to vertical handoffs in a packet switched network. 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]. A clear advantage of involving the IMS is its ability for real-time session negotiation and management using SIP for achieving seamless session mobility during vertical handoffs. Additionally, MIP also plays an important role in supporting terminal mobility management in this approach.

The reminder 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 is the section on *Pareto/M/1* queuing analysis. Subsequently the vertical handoff analysis and numerical results follow prior to the concluding remarks.

II. MOBILITY AND VERTICAL HANDOFF MANAGEMENT IN HETEROGENEOUS MOBILE NETWORKS

The interworked networking platform used in this analysis is illustrated in Fig. 1. Interested readers may refer to [10] for more specific and detailed information on the 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

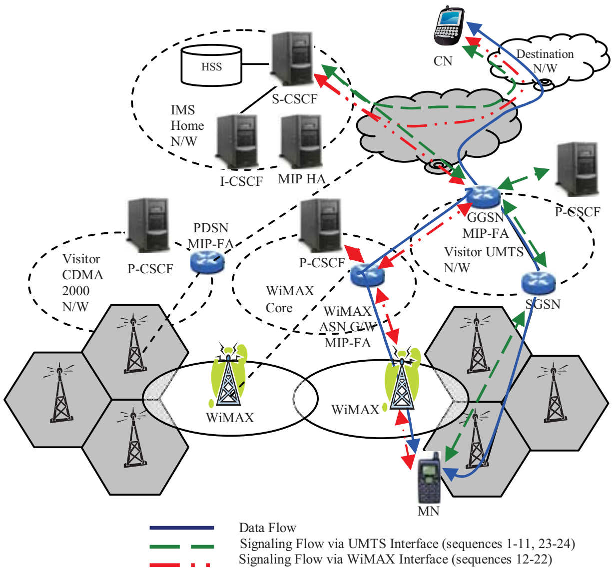

Fig. 1. The Proposed Interworking Architecture.

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.

The 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 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

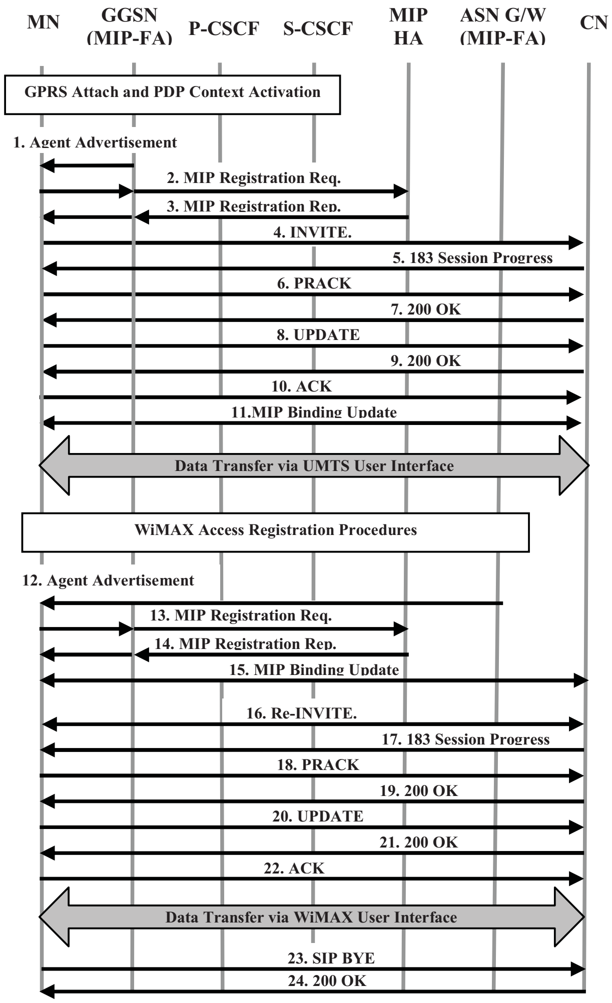

Fig. 2. Vertical Handoff Signaling.

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) [13].

The next stage is the taking place of the IMS-SIP session handoff procedures. This requires sending a SIP Re-INVITE (with same Call-ID and other identifiers corresponding to the ongoing session) to the destination SIP UAC (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 [14]. 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. PARETO/M/1 QUEUING MODEL

Internet traffic has been described as having one or more of the following related characteristics: self-similar (or fractal) traffic traces, long-range dependence, burstiness on multiple scales, and long or heavy-tailed packet inter-arrival times or service requirements [15].

Self-similarity implies that the traffic looks the same over any time scale. Furthermore, as first shown in [3], Poisson traffic does not have the same characteristic. Long-range dependence is defined with respect to the autocorrelation function of a stationary discrete-time stochastic process, *R(k)*. It measures the level of correlation of the process with itself and measured *k* periods away. The process is said to be long-range dependent if $\Sigma k\ R(k) = \infty$, thus implying that there is at best a slow and non-exponential decline in the autocorrelation function with increasing lags *k*.

Furthermore, it may be argued that a self-similar process is also long-range dependent. The Hurst parameter is often used to describe the degree of self-similarity in long-range processes [15]. The concept of burstiness means that packets arrive in several short inter-arrival times followed by a much longer time. Examples of long-tailed distributions are the Pareto, the log-normal, the folded Cauchy, and the DFR form of the Weibull. In this analysis, a method for studying Pareto queues is presented.

The standard form for the two-parameter Pareto distribution function defined over the nonnegative real numbers can be written as:

$$F(x) = 1 - 1/\left[(\alpha + x)^{\beta}\right] \forall(\alpha, \beta \rangle 0) \tag{1}$$

As a critical motivation for the subsequent procedure, such a distribution function can be directly derived as a gamma *(α,β)* mixture of ordinary exponential densities. With no loss in generality, the one-parameter version of the Pareto can be given as [16]:

$$F(x) = 1 - 1/\left[(1 + x)^{\beta}\right] \tag{2}$$

The corresponding density function is:

$$f(x) = \beta/\left[(1 + x)^{\beta+1}\right] \tag{3}$$

and it is shown that the Pareto is indeed a long-tailed distribution, where β measures the initial rate of decline of the density function curve [16]. In the following scenario, a Pareto arrival distribution into the queuing system is considered. From the standard analysis of a *G/M1* queue, the steady-state probability for the number of customers Q in system just before an arrival is given for all nonnegative n as [17]:

$$\Pr\{Q = n\} = q_n = (1 - r)r^n \tag{4}$$

For *Pareto/M/1*, the usual approach for obtaining the stationary delay time distributions and system size probabilities requires solving a root finding problem involving the Laplace-Stieltjes Transform (LST), A*(s), of the inter-arrival time distribution function [17], [18]. Thus r is the root of the fundamental branching process equation obtained by solving for z is:

$$z = A^{*}[\mu(1 - z)] \tag{5}$$

where $1/\mu$ is the expected service time [17]. The system utilization, ρ, which is λ/μ, where λ is the customer arrival rate, and for the problem to have a non-trivial solution, one must have $\rho < 1$. The unique root of the fundamental equation of the branching process, say r, then becomes the parameter of a geometric distribution for steady-state system sizes at the embedded arrival points [16]. These geometric probabilities are then combined with convolutions of the exponential service distribution to derive the stationary line-delay distribution.

Unfortunately, for the case of Pareto arrivals, a closed form for $A^*(s)$ does not exist. This paper uses a method proposed by Harris and Marchal for finding Coxian distribution fits for arbitrary distribution functions using Laplace transform approximations [19]. It turns out that their technique, which is called as the Transform Matching Method (TMM)) works especially well for distributions defined over the full real line but without all moments. Thus use TMM for $A^*(s)$ and then use Newton's method to solve for the root r.

Once the root is found, the queue and system waiting-time distribution functions can easily be derived for $t \geq 0$ as [17]:

$$W_q(t) = 1 - re^{-\mu(1-r)t} \tag{6}$$

$$W(t) = 1 - e^{-\mu(1-r)t} \tag{7}$$

A close observation of the above queue and system waiting-time distribution functions indicates that they have the same functional form as the *M/M/1* queue except with r replacing ρ. Thus the expected queue waiting time, *Wq*, and system waiting time, *W*, can be expressed as [17] [16]:

$$W_q = \frac{r}{\mu(1-r)} \tag{8}$$

$$W = \frac{1}{\mu(1-r)} \tag{9}$$

IV. VERTICAL HANDOFF ANALYSIS

An analytical model is derived for evaluating the proposed scheme for analyzing vertical session handoff management for Internet data traffic.

A. *Handoff Delay*

A standard vertical handoff delay during mid-session mobility consists of the following sub-procedures (or delays); D_1 = link layer handoff delay, D_2= movement detection delay, D_3 = address allocation delay, D_4 = session re-configuration delay, and D_5 = packet re-transmission delay [20]. The vertical handoff delay at the network layer (and above) are calculated independent of the link layer delay D_1 and mainly consist of D_3 and D_4. According to our proposed architecture for IMS based vertical handoff, there is no Dynamic Host Configuration Protocol (DHCP) related address allocation; hence it can be argued that is D_4 the main contributor for network layer based vertical handoff delay, D. The session re-configuration delay, D_4 mainly consists of the previously mentioned IMS based session negotiation and handoff and Home Subscriber Server (HSS) related message exchange delays.

In order to derive an expression for D, we must first derive an expression for analyzing the end-to-end transmission delay. Hence, let us assume that the end-to-end transmission delay for a packet size S sent from network A to network B over a number of hops via a wireless and wired links to be expressed as:

$$D(S, H_{a-b}) = D_{wl} + D_w + L_{wl} + L_w \tag{10}$$

where, D_{wl} is the total delay at the wireless interface (say, Base Station - BS), D_w is the total delay at the wired link, L_{wl} is the latency of the wireless link, and Lw is the latency of the wired link. In order to derive D_{wl} and D_w a *Pareto/M/1* queuing model has been applied to the packet flow of the data session at the wireless BS and other networking elements of the IMS on the path of signaling and data routing of Fig. 2.

It is important to note that to apply the results of *Pareto/M/1* analysis, it is assumed that the service times that a packet experiences at different nodes are independent of each other. However, this assumption is untrue, since the service time is proportional to the packet length, and a packet has the same length as it traverses the network. Nevertheless, it has been found that this independence assumption can be used in large networks [18]. Using the results from the *Pareto/M/1* model, expressions for D_{wl} and D_w can be expressed as:

$$D_{wl} = \frac{1}{\mu_{wl}(1-r_{wl})} \tag{11}$$

where, μ_{wl} is the service rate and r_{wl} is the root of the fundamental branching process equation obtained by solving for z at the wireless interface. For clarity and convenience sake, the units for μ_{wl} are changed from packets/sec to bits/sec. If the probability density function of for packet size, x, in bits be $\mu e^{-\mu x}$ with a mean packet length of $1/\mu$ bits/packet, and the capacity of communication channel i be C_i bits/sec. The product μC_i is then the service rate in packets/sec. Therefore, for channel i, we have

$$D_{wl} = \frac{1}{\mu C_i(1-r_{wl})} \tag{12}$$

where, D_{wl} includes both queuing and transmission delays. Also note that the mean packet size does not depend on the channel as the capacity and the input rates do. However, when D_w is considered, it can be expressed as a collection of delays of multiple *Pareto/M/1* queues. It is also assumed that if the output of several *Pareto/M/1* servers feed into the input queue of another server, the resulting input process is also a Pareto process [21]. This assumption has also been derived for *G/M/1* queues in [18]. Therefore, by using the derived result from [18], the total wired network delay experienced by a packet can be expressed as:

$$D_w = \frac{1}{\lambda_w}\sum_j \lambda_j \left(\frac{1}{\mu C_j(1-r_w)}\right) \tag{13}$$

where, λ_w is the total packet arrival rate to the network, λ_j is the packet arrival rate at j^{th} node, and μC_j is the service rate in packets/sec at the j^{th} node. Thus by combining equations (10), (12) and (13) we get:

$$D(S, H_{a-b}) = \frac{1}{\mu C_i(1-r_{wl})} + \left\{\frac{1}{\lambda_w}\sum_j \lambda_j \left(\frac{1}{\mu C_j(1-r_w)}\right)\right\} + L_{wl} + L_w \tag{14}$$

Now, an expression for the vertical handoff delay D can be expressed by applying (14) to the entire IMS signaling flow involved in the vertical handoff mechanism as illustrated in Fig. 2. Thus the final expression for D is a combination of the following end-to-end delay components as indicated in equation (15).

$$\begin{aligned} D_{IMS} = {} & D(S_{MIPReq}, H_{UMTS\text{-}MIP\text{-}HA}) + D(S_{MIPRep}, H_{UMTS\text{-}MIP\text{-}HA}) \\ & + D(S_{MIP\text{-}BU}, H_{UMTS\text{-}CN}) + D(S_{ReINVITE}, H_{WiMAX\text{-}CN}) \\ & + D(S_{183\text{-}SP}, H_{WiMAX\text{-}CN}) + D(S_{PRACK}, H_{WiMAX\text{-}CN}) \\ & + D(S_{OK}, H_{WiMAX\text{-}CN}) + D(S_{UPDATE}, H_{WiMAX\text{-}CN}) \\ & + D(S_{OK}, H_{WiMAX\text{-}CN}) + D(S_{ACK}, H_{WiMAX\text{-}CN}) \\ & + D(S_{BYE}, H_{UMTS\text{-}CN}) + D(S_{OK}, H_{CN\text{-}UMTS}) + \Delta \end{aligned} \tag{15}$$

where, Δ is the additional IMS (application layer) related latency due to HSS lookup process. The important point to note here is that the derivation of equation (15) has not taken into account the errors that may cause various messages to be damaged or lost. This is since for successful session establishment, the entire message flow must take place and if any message is damaged or lost the vertical handoff process will fail. Hence it has been assumed that the channel is error free during the process of the vertical handoff taking place. It is also worth reminding that make-before-break handoff is applied in the proposed handoff scenarios, which helps compensate for large handoff delays. For the purpose of a complete analysis of vertical handoff delay, the standard straight forward case of break-before-make handoff scenario is used.

B. *Packet Loss*

The total packet loss (*Pkt_loss*) during a vertical session handoff can be defined as the sum of all lost packets during the vertical handoff while the MN is receiving the downlink data packets. It is assumed that the packet loss begins when the layer 2 handoff is detected and all in-flight packets are lost during the vertical handoff time. Thus, it can be expressed as:

$$Pkt\_loss = \left[\frac{1}{2T_{ad}} + D \right] \times \lambda_{wl} \times N_m \qquad (16)$$

where, T_{ad} is the time interval between P-CSCF discovery times, λ_{wl} is the downlink packet arrival rate at the wireless interface, and N_m is the average number of vertical handoffs during a single session [20]. N_m plays a major role in the calculation of packet loss since the packet loss due to vertical handoff is directly proportionate to the number of handoffs it is subjected within a given session.

V. NUMERICAL RESULTS

This section presents numerical results relating to the behavior of vertical handoff delay and transient packet loss against system utilization for the case where the shape parameter $\beta = 1.5$. In order to better understand the behavior of the *Pareto/M/1* queue, its performance has been compared against the known closed form values for an *M/M/1* queue. The results used for the performance comparison for an *M/M/1* queue is obtained from one of the authors' previous works on vertical session handoff analysis [22]. Table I provides the typical MIPv4 and SIP message sizes and Fig. 4 provides the relative distances in hops used in the numerical evaluation.

Fig. 5 illustrates the graphs for WiMAX-to-UMTS vertical handoff delay against the system utilization for a *Pareto/M/1* queuing analysis and a *M/M/1* classical queuing analysis. Both analytical methods show approximately close behavioral patterns up to the point of 60% of the system utilization. However, beyond this point, the vertical handoff delay increases according to the nature of packet arrival patterns (i.e., Poisson or Pareto). For example, since Poisson arrivals are relatively smoother and not as bursty as Pareto arrivals, the

TABLE I
MESSAGE SIZES AND PARAMETER VALUES

Message	Size (Bytes)	Message	Size (Bytes)
INVITE	736	MIP Reg. Req.	60
Re-INVITE	731	MIP Reg. Rep.	56
183 Ses. Pro.	847	MIP BU	66
PRACK	571	MIP BACK	66
200 OK	558	C_i	2-70 Mbps
UPDATE	546	L_{wl}	2ms
ACK	314	L_w	0.5ms
BYE	550	λ_d	33kbps
MIP Agent Ad.	28	T_{ad}	1sec

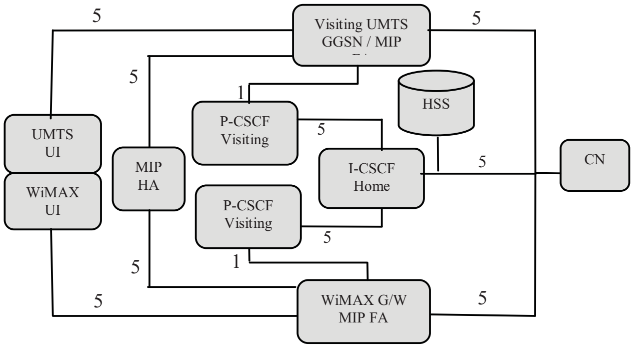

Fig. 4. Relative distances in hops.

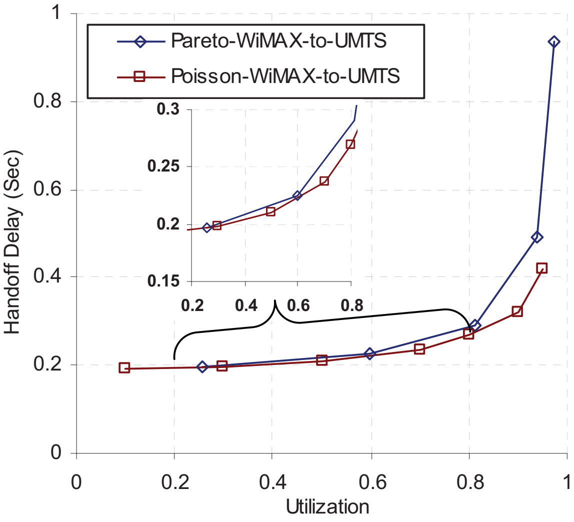

Fig. 5. WiMAX-to-UMTS Handoff Delay vs. System Utilization.

results clearly illustrates an exponentially increasing delay. On the other hand, as the system utilization grows beyond 80%, the *Pareto/M/1* queuing model tends to demonstrate its characteristic heavy tailed behavior.

On the other hand, Fig. 6 illustrates relatively lower handoff delays for the graphs corresponding to UMTS-to-WiMAX vertical handoff delay against the system utilization for a Pareto/M/1 queuing analysis and a *M/M/1* classical queuing analysis. This indicates that when a session is transferred to a network with relatively lower link bandwidth, a relatively higher vertical handoff delay may be expected.

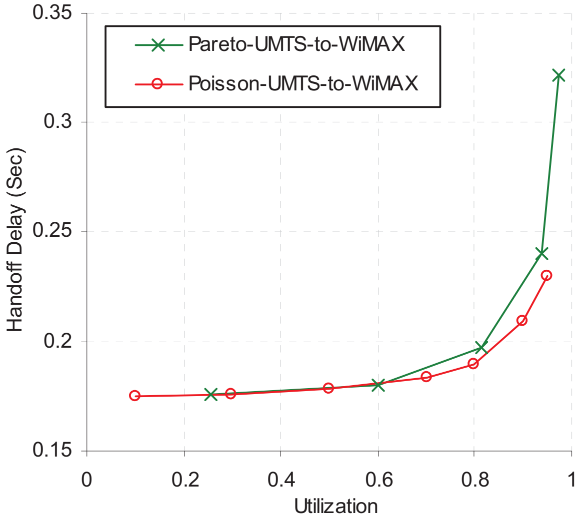

Fig. 6. UMTS-to-WiMAX Handoff Delay vs. System Utilization.

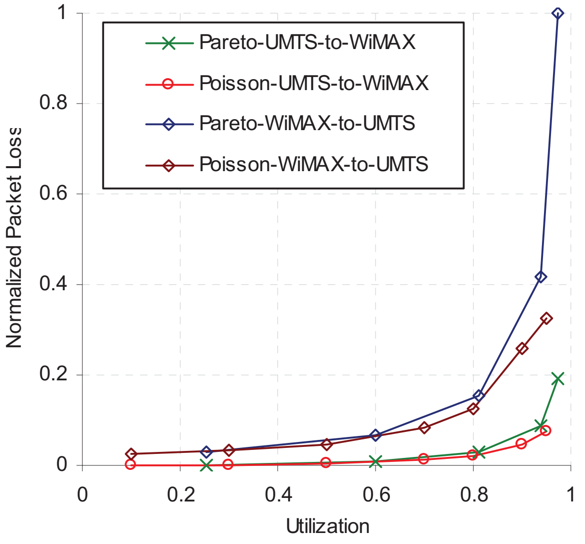

Fig. 7. Transient Packet Loss vs. System Utilization.

Fig. 7 illustrates the normalized transient packet loss during vertical handoffs as the system utilization increases (in the case of a break-before-make handoff scenario). The voice codec considered for the downlink packet transmission in this case is a GSM codec. According to equation (16), the packet loss during a vertical handoff is directly proportional to the vertical handoff delay. Therefore, relatively high vertical handoff delays indicated by the WiMAX-to-UMTS graphs in Fig. 5 directly relate to the two high packet loss curves in Fig. 7. Similarly, the packet loss is relatively low in Fig. 7 for a UMTS-to-WiMAX handoff, which is in line with the two relatively low handoff delay graphs shown in Fig. 6. Further, the exponential and heavy-tailed behaviors can also be observed in Fig. 7 for Poisson and Pareto based models respectively.

VI. CONCLUSIONS

This paper presents a novel *Pareto/M/1* queuing based analytical model for analyzing vertical handoffs for a roaming user in a heterogeneous mobile networking environment. Using the proposed analytical model, an in-depth analysis is performed for investigating the vertical session handoff delay from WiMAX-to-UMTS (and vice-versa) against the system utilization. The results for the *Pareto/M/1* queuing analysis are then compared against the *M/M/1* classical queuing analysis. According to the comparison, both queuing methods show approximately close behavioral patterns up to the point of 60% of the system utilization. As the system utilization grows beyond 80%, the *Pareto/M/1* queuing model shows its characteristic heavy tailed behavior.

REFERENCES

[1] L. Ma, "The Competition and Cooperation of WiMAX, WLAN, and 3G," in *Proceedings of IEE 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 *Proceedings of the IEEE INFOCOM 2003*, San Franciso, California, Apr. 2003.

[3] V. Paxon and S. Floyd, "Wide area traffic: The faliure of Poisson Modeling," *IEEE/ACM Transactions on Networking*, vol. 3, pp. 226-244, 1995.

[4] A. T. Andersen, A. Jensen, and B. F. Nielsen, "Modeling And Performance Study Of Packet-Traffic With Self–Similar Characteristics Over Several Time-Scales With Markovian Arrival Process (MAP)," in *Proceedings of the Nordic Teletraffic Seminar NTS*, 1995.

[5] S. Basu, A. Mukherjee, and S. Klivansky, "Time Series Models for Internet Traffic," in *Proceedings of the IEEE INFOCOM*, Mar. 1996.

[6] A. Feldmann, A. C. Gilbert, and W. Willinger, "Data Networks As Cascades: Investigating the Multifractal

Nature Of The Internet," *ACM SIGCOMM Computer Communication Review*, vol. 28, pp. 42-55, 1998.

[7] A. Feldmann, A. C. Gilbert, W. Willinger, and T. Kurtz, "The Changing Nature of Network Traffic: Scaling Phenomena," *Computer Communication Review*, vol. 28, pp. 5-29, 1998.

[8] H. Fowler and W. Leland, "Local Area Network Traffic Characteristics, With Implications for Broadband Network Congestion Management," *IEEE Journal in Selected Areas in Communications*, vol. 9, pp. 1139-1149, 1991.

[9] V. S. Frost and B. Melamed, "Traffic Modeling For Telecommunications Networks," *IEEE Communications Magazine*, vol. 32, pp. 70-81, 1994.

[10] K. S. Munasinghe and A. Jamalipour, "A Unified Mobility and Session Management Platform for Next Generation Mobile Networks," in *Proceedings of the 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] C. Perkins, "IP Mobility Support for IPv4," IETF RFC 3344, 2002.

[14] K. S. Munasinghe and A. Jamalipour, "A 3GPP-IMS based Approach for Converging Next Generation Mobile Data Networks," in *Proceedings of the IEEE International Conference on Communications*, Glasgow, UK, Jun. 2007.

[15] T. Fowler, "Fractals, Long Range Dependence and Packet Traffic," in *Proceedings of the Advanced Simulation Technologies Conference*, San Diego, CA, 1999.

[16] M. J. Fischer, D. M. B. Masi, D. Gross, and J. F. Shortle, "One-Parameter Pareto, Two-Parameter Pareto, Three-Parameter Pareto: Is there a modeling difference" *The Telecommunications Review*, 2005.

[17] G. Gross and C. M. Harris, *Fundamentals of Queuing Theory*, 3 ed. New York: John Wiley, 1998.

[18] L. Kleinrock, *Queuing Systems Volume 1: Theory*: John Wiley & Sons, 1975.

[19] C. M. Harris and W. G. Marchal, "Distribution Estimation Using Laplace Transforms," *INFORMS Journal of Computing*, vol. 10, pp. 448-458, 1998.

[20] S. C. Lo, G. Lee, W. T. Chen, and J. C. Liu, "Architecture for mobility and QoS support in all-IP wireless networks," *IEEE JSAC*, vol. 22, pp. 691-705, 2004.

[21] B. Bashforth and C. Williamson, "Statistical Multiplexing of Self-Similar Video Streams: Simulation Study and Performance Results," in *Proceedings of the International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems*, Montreal, Canada, Jul. 1998.

[22] K. S. Munasinghe and A. Jamalipour, "Interworked WiMAX-3G Cellular Networks: An Architecture for Mobility Management and Performance Evaluation," *IEEE Transactions on Wireless Communications*, 2009 (in press).