2011 International Conference on Network and Electronics Engineering
IPCSIT vol.11 (2011) © (2011) IACSIT Press, Singapore

On the Provisioning of Guaranteed QoS in Wireless Internet with Self-Similar Traffic Input based on G/M/1 Queueing System and Limited Service Polling Model

Mohsin Iftikhar, Abdul Malik Rahhal and Mansour Zuair

Computer Science Department
College of Computer and Information Sciences
King Saud University, Riyadh, Saudi Arabia
miftikhar@ksu.edu.sa; amr@ksu.edu.sa; Zuair@ksu.edu.sa

Abstract. Over the past decade, we have witnessed a growing popularity of new wireless architectures such as 3G/4G, Wi-Fi and WiMAX due to the increase in demand for wireless Internet access. The all-IP based future mobile and wireless network model is expected to be the most dominant architecture for QoS provisioning in next-generation wireless networks, mainly due to its scalability and capability of inter-working heterogeneous wireless access networks. Recently, the rapid growth of various wireless infrastructures and the interesting mixture of wireless traffic generated by large number of devices (PDAs, Laptops and cell-phones) have diverted the attention of wireless research community towards understanding the nature of traffic carried by different wireless architectures. A series of recent studies on GPRS aggregated traffic, WAP and Web traffic has proven that just like fixed IP traffic, wireless traffic also exhibits strong long-range dependency. However, much of the current understanding of wireless traffic modeling builds on classical Poisson distributed traffic, which can yield misleading results and hence poor wireless network planning. In this paper, we contribute to the accurate modeling of wireless IP traffic by considering three different classes of traffic that exhibit long-range dependency and self-similarity. We consider a model of three queues based on G/M/1 queueing system and analyze it on the basis of novel scheduling logic and derive exact bounds on packet delay for the corresponding traffic classes. We also develop a comprehensive discrete event simulator to understand the QoS behavior of the corresponding traffic classes under this proposed scheduling scheme. The novel scheduling logic outperforms the traditional schemes such as priority and round robin and serves as the basis to offer guaranteed QoS relevant to the diversified requirement of different applications in this heterogeneous mixture of wireless networks.

Keywords: QoS, 3G/4G, Wi-Fi, WiMAX, GPRS

1. Introduction

With the increasing demand of Internet connectivity and because of the flexibility and wide deployment of IP technologies, there has been a paradigm shift towards IP-based solutions [1]. Several Wireless IP architectures have been proposed [2-8] based on three main IP QoS models, IntServ [9], DiffServ [10] and MPLS [11]. To provide differential treatment to multiple traffic classes within these different kinds of network domains, several queueing tools have been developed that can be implemented in the routers. Examples include; Priority Queueing (PQ), Custom Queueing (CQ), Weighted Fair Queueing (WFQ), Class Based Weighted Fair Queueing (CBWFQ) and Low Latency Queueing (LLQ) [12]. Regardless of the queueing discipline implemented in a router, the scheduler at the output port of the router serves multiple queues simultaneously. This kind of single server/multiple queues system is generally called a polling model.

Recent research has shown that wireless data traffic exhibits self-similarity and long-range dependency [13-16]. The properties and behavior of self-similar traffic is very different from traditional Poisson or

Markovian traffic. To guarantee tight bound QoS parameters to heterogeneous end-users of Mobile Internet, it is essential to model the traffic behavior accurately through wireless IP domains. There is a need to determine end-to-end QoS parameters such as delay, jitter, throughput, packet loss, availability and per-flow sequence preservation.

In this paper, we contribute to the accurate modeling of wireless traffic behavior by analyzing G/M/1 queueing system, taking into account three different classes of wireless IP traffic that exhibit long-range dependence and self-similarity. We consider three queues served by the server according to a limited service polling model and derive exact bounds on packet delay for corresponding traffic classes. The study of polling models is important since it gives very good insight into the qualitative behavior of many proposed and implemented queueing disciplines and forms the basis to derive exact expressions of different QoS parameters such as delay, throughput and jitter, thus leading towards offering guaranteed service to the end-user.

The paper is organized as follows. Section II reviews related work. Section III is devoted to explain self-similar traffic with multiple classes and the calculation of interarrival times respectively. Section IV explains the procedure of formulating the imbedded Markov chain along with finding out the packet delay. In Section V, we provide the simulation results. The applications of the current modeling are discussed in Section VI. Finally, conclusion and future work is given in Section VII.

2. RELATED WORK

Generally, polling models can be classified as Exhaustive, Gated and Limited Service. The exact details of the systems are beyond the scope of this paper. Instead, the readers are referred to [17] for a detailed discussion of polling systems. In this section, we constrain the discussion to explaining the limited service polling model on which this paper is based.

In the limited service discipline, a station (queue) is served until either 1) the buffer is emptied or 2) a specified number of packets have been served, whichever occurs first. If at most k packets are served in one cycle from a queue, we refer to the model as a k-limited polling model [17]. A large amount of research has been undertaken regarding limited service polling models. A detailed approximate analysis of limited service polling model has been given in [18-19]. The major weakness of the existing work is that the assumption of only Poisson arrivals has been considered as traffic input and second even in the case of Poisson arrivals only approximate results are available regarding the limited service discipline.

Few studies have focused on wireless traffic modeling and here we discuss the most relevant work. The influence of self-similar input on GGSN performance in the UMTS Release 5 IM-subsystem has been analyzed on the basis of a FBM/D/1/W queueing system (FBM-Fractional Brownian Motion) in [20]. In this work, different probabilistic parameters of GGSN such as average queue length and average service rate were also found. The work in [21] presents modeling and a simulation study of the Telus Mobility (a commercial service provider) Cellular Digital Packet Data (CDPD) network. The collected results on average queueing delay and buffer overflow probability indicate that genuine traffic traces produce longer queues as compared to traditional Poisson based traffic models. To get an overview of the analysis done in wireless IP traffic modeling with self-similar input, we refer the readers to [22-25]. These studies are merely based on characterization of wireless traffic. Consequently, the issue of providing guaranteed QoS to the end-user of Mobile Internet has not been addressed properly. To provide differential treatment to multiple traffic classes with different QoS demands, there is a need to accurately determine end-to-end QoS parameters such as delay, jitter, throughput, packet loss, availability and per-flow sequence preservation.

To overcome the limitations of the previous work and in the light of achieving the objective of end-to-end QoS in Wireless Internet, we consider a model of three queues based on G/M/1, which takes into account three classes of self-similar input traffic denoted by *SS/M/1*, and we analyze it on the basis of a limited service polling model with zero switch over time. Our traffic model [26] is parsimonious with few parameters. It is similar to on/off processes, in particular to its variation N-Burst model studied in [27] where packets are incorporated. However, only a single type of traffic is considered in [27]. We present a novel analytical approach and derive expressions for the steady state queue length distribution at the time of arrivals, as well as

for the bounds on delay. The current paper is the extension of our prior work on 1-limited polling model [35], which is also known as an alternating service model.

3. SELF-SIMILAR TRAFFIC MODEL

The traffic model considered in this paper has been studied in [26]. It belongs to a particular class of self-similar traffic models, more recently referred to as the telecom process in [28]. The model captures the dynamics of packet generation while accounting for the scaling properties of the traffic in telecommunication networks.

The traffic is found by aggregating the number of packets generated by several sources. Each source initiates a session with a heavy-tailed distribution, in particular a Pareto distribution whose density is given by $g(r) = \delta b^{\delta} r^{-\delta-1}$, $r > b$, where δ is related to the Hurst parameter by $H = (3-\delta)/2$. The packets arrive according to a Poisson process with rate α, locally, over each session. The sessions also arrive according to a Poisson process with rate λ. In the framework of a Poisson point process, the model represents an infinite number of potential sources. With our model, we had been able to find the interarrival time distributions for different classes of traffic.

For each class, the traffic $Y(t)$ measured as the total number of packets injected in [0, t] is found by

$$Y(t) = \sum_{S_i \le t} U_i (R_i \wedge (t - S_i))$$

where U_i, R_i, S_i denote the local Poisson process, the duration and the arrival time of session i, respectively. Hence, $Y(t)$ corresponds to the sum of packets generated by all sessions initiated in [0,t] until the session expires if that happens before t, and until t if is does not. The stationary version of this model based on an infinite past is considered in calculations below. The packet sizes are assumed to be fixed because each queue corresponds to a certain type of application where the packets have fixed size or at least fixed service distribution. The traffic model Y is long-range dependent and almost second-order self-similar; the auto covariance function of its increments is that of fractional Gaussian noise. Three different heavy traffic limits are possible depending on the rate of increase in the traffic parameters [26]. Two of these limits are well known self-similar processes, fractional Brownian motion and Levy process, which do not account for packet dynamics in particular.

$$\overline{F}_T(t) = P\{T > t\} = \frac{1}{\rho}\Big\{ e^{-\nu} e^{-\eta} \cdot \exp[-\lambda t(1-e^{-\alpha t})] \quad (\exp[\nu e^{-\alpha t}] - 1)(\exp[\eta(1-e^{-\alpha t})/\alpha t] - 1)$$

$$+ e^{-\nu} e^{-\eta} \exp[-\lambda t(1-e^{-\alpha t})](\exp[\nu e^{-\alpha t}] - 1) + e^{-\nu} e^{-\eta} \exp[-\lambda t(1-e^{-\alpha t})](\exp[\eta(1-e^{-\alpha t})/\alpha t] - 1)\Big\}$$

where $\nu = \lambda \int_t^{\infty} (y-t) g(y)\, dy$, $\eta = \lambda \left[\int_0^t y g(y)\, dy + t\, \overline{G}(t) \right]$ $\rho = 1 - \exp[-\lambda \delta b/(\delta - 1)]$

and the complementary distribution function $\overline{G}$ is that of the Pareto density g defined earlier for session duration.

The complementary distribution function $\overline{F}_T$ above is for a single class of packets. Let T_i denote the interarrival time of a type i packet, and f_{T_i} denote its density function; that is, $f_{T_i}(t) = -d\overline{F}_{T_i}/dt$ using $(\lambda_i, \alpha_i, \delta_i, b_i)$. For the queueing analysis of the present paper, we also need cross interarrivals $T_{ij}, i, j = 1,2$, occurring between a type i packet followed by a type j packet. The density functions of T_{ij} are found as:

$$f_{T_{11}}(t) = f_{T_1}(t)\, \overline{F}_2^{0}(t) \qquad f_{T_{12}}(t) = f_2^{0}(t)\, \overline{F}_{T_1}(t)$$

$$f_{T_{22}}(t) = f_{T_2}(t)\, \overline{F}_1^{0}(t) \qquad f_{T_{21}}(t) = f_1^{0}(t)\, \overline{F}_{T_2}(t)$$

where

$$\overline{F}_i^{0}(t) = e^{-\nu_i} e^{-\eta_i} \exp[-\lambda_i t(1-e^{-\alpha_i t})] \exp[\nu_i e^{-\alpha_i t}] \exp[\eta_i (1-e^{\alpha_i t})/\alpha_i t]$$

and f_i^0 is the corresponding density function. The detailed derivation of interarrival time calculations has been given in [29]. In this paper, we consider three classes of traffic streams arriving at router. The generalization to three classes is trivial and its detailed derivation has been given in [30]. Here we just present the final results as follows:

$$f_{T_{ii}}(t) = f_{T_i}(t)\,\overline{F}_j^{0}(t)\,\overline{F}_k^{0}(t)$$

$$f_{T_{ij}}(t) = f_{T_j}^{0}(t)\,\overline{F}_{T_i}(t)\,\overline{F}_{T_k}^{0}(t)$$

where we multiply the corresponding density with complementary probabilities to make sure that the desired transition occurs from type *i* arrival to type *i* or *j*, and $i, j, k \in \{1,2,3\}$.

4. ss/m/1 WITH MULTIPLE CLASSES AND LIMITED SERVICE POLLING SCHEME

We consider a model of three queues based on self-similar input traffic denoted by SS/M/1, and analyze it on the basis of limited service polling model with zero switch over time. The scheduling logic of this limited service polling model is as follows: It visits the first queue and serves 2 packets, and then it goes to second queue and serves 1 packet. After that it again goes back to first queue and serves 2 packets and then it goes to third queue and serves 1 packet. Hence during each cycle, it serves 4 packets from queue 1 and 1 packet from queue 2 and queue 3 each. Let the service time distribution have rate μ_1, μ_2 and μ_3 for class 1, class 2 and class 3 packets, respectively. Since the scheduler serves four packets from queue one and one packet from queue no. 2 and one packet from queue no. 3 during each cycle, hence, we need to differentiate between the first, second, third and fourth packet of queue 1 of the same cycle and then we need to classify between class 1, class 2 and class 3 packets as well. Therefore S_1^1 is the first packet of queue 1 of the same cycle, S_1^2 is the second packet of queue 1 of the same cycle, S_1^3 is the third packet of queue 1 of the same cycle, S_1^4 is the fourth packet of queue 1 of the same cycle and S_2 is the packet of queue 2 and S_3 is the packet of queue 3. That's why the notation S_n^m can be used to differentiate between these four different kinds of packets, where $m = 1,2,3,4$ and $n = 1,2,3$. The subscript m will be used only when $n = 1$. Where S_n^m is the service time required by class n packets.

The usual imbedded Markov chain [31] formulation of G/M/1 is based on the observation of the queueing system at the time of arrival instants, right before an arrival. At such instants, the state of the system is the number of packets that arriving packet sees in the queue plus packets in service, if any, excluding the arriving packet itself. We specify the states and the transition probability matrix *P* of the Markov chain with the self-similar model for three types of traffic.

Let $\{X_n : n \geq 0\}$ denote the imbedded Markov chain at the time of arrival instants. As the service is alternating, the type of packet in service is important at each arrival instant of a given type of packet to determine the queueing time. Therefore, we define the state space as:

$$S = \{(i_1, i_2, i_3, a, s) : a \in \{a_1, a_2, a_3\},$$
$$s \in \{s_1^1, s_1^2, s_1^3, s_1^4, s_2, s_3, I\}, i_1, i_2, i_3 \in Z_+\}$$

We generate the transition probability matrix *P* of the Markov chain by specifying the transition probabilities from all the states in the states space i.e. non-idle states, states with empty queues and arrival at full queue. We only write down one transition in detail:

Transition from $(i_1,i_2,i_3,a_1,s_1^1) \rightarrow (j_1,j_2,j_3,a_2,s_1^2)$

Here a transition occurs from an arrival of class 1 traffic to an arrival of class 2 traffic, such that the class 1 arrival has seen the first packet of class 1 traffic in service of some cycle, with i_1 packets of queue 1 in the system (equivalently, total of queue 1 and the packet in service) and i_2 packets of class 2 and i_3 packets of class 3 in the system. The transition occurs to a state, where the new arrival (class 2 packet) sees j_1 packets of class 1, j_2 packets of class 2 and j_3 packets of class 3 in the system, with a second packet of class 1 in service of some cycle. Recall that there are two kind of classification, one is between first, second, third and fourth packet of same cycle of class 1 (queue 1) and then between class 1, class 2 and class 3 packets (queue 1, queue 2 and queue 3). Since in the previous state an arrival of class 1 has seen the first packet of class 1 in service, and in the next state an arrival of class 2 sees the second packet of class 1 in service, it is implied definitely that the first packet (S_1^1) of class 1 has completed its service. Now as the new class 2 arrival finds second packet of class 1 (S_1^2) in service, but because of memory-less property of exponential service time,

we do not know that how many cycles have been completed. To make this idea more clearly, we assume that in the previous state the first packet of class 1 (S_1^1) which was in service belongs to a cycle *A*. Now as the new arrival finds the second packet of class 1(S_1^2) in service, there are many possibilities, the first possibility is that it belongs to the same cycle *A*, in this case, only one packet of class 1 has been served and no packet has been served from queue 2 and 3. If S_1^2 belongs to the next cycle, for example cycle *B*, then definitely 5 packets have been served from queue 1 and 1 packet has been served from queue 2 and one packet has been served from queue 3. Hence, the maximum number of class 1 packets that can be served are i_1 (if the total number of packets in queue 1 i.e. i_1 is odd), otherwise the maximum number of packets that can be served are $i_1 - 1$ (if the total number of packets in queue 1 i.e. i_1 are even), if the arriving packet is still to find a class 1 packet in service. However, j_1 includes the type 1 packet that arrived in the previous state. Hence, we have:

$$j_1 = i_1 + 1 - k, \quad j_n = i_n - (\frac{k-1}{4}), n = 2,3$$

until either both queue 2 and queue 3 are exhausted or only one packet (in case of odd number of packets) or two packets (in case of even number of packets) from class 1 remain in the system, the second packet (S_1^2)of class 1, being in service, whichever occurs first. We consider queue 2 and queue 3 as a single queue and denote it as QUEUE I_2 So there are two possibilities:

(1) If $i_1 / 2 \geq I_2$ and $k = 1,5,...4i_n - 2$, *n=2, 3* or when $i_1 / 2 < I_2$ & $k = 1,5,.....i_1(odd)$ or $i_1 - 1(even)$: The transition probability is:

$$P\{X_{n+1} = (i_1 - k + 1, i_2 - (\frac{k-1}{4}), i_3 - (\frac{k-1}{4}), a_2, s_1^2) \mid X_n = (i_1, i_2, i_3, a_1, s_1^1)\} = \int_0^\infty \int_0^t \int_{t-x}^\infty f_{S_1^2}(s) f_{S_1^k + S_2^{\frac{k-1}{4}} + S_3^{\frac{k-1}{4}}}(x) f_{T_{12}}(t)\, ds\, dx\, dt$$

Recall that $f_{T_{12}}(t)$ denotes the probability density function for the Interarrival of two packets where a type 1 packet is followed by a type 2 packet and we used the fact, that the remaining service time of a type 1 packet in service has the same exponential distribution Exp (μ_1), due to memory-less property of the Markovian service time.

(2) On the other hand, merely class 1 packets are served if both queue 2 and queue 3 are exhausted. Therefore, If $i_1 / 2 \geq I_2$ and $k = 4i_n + 2.....i_1(odd)$ or $i_1 - 1(even)$ then we have:

$$P\{X_{n+1} = (i_1 + 1 - k, 0, 0, a_2, s_1^2) \mid X_n = (i_1, i_2, i_3, a_1, s_1^1)\} = \int_0^\infty \int_0^t \int_{t-x}^\infty f_{S_1^2}(s) f_{S_1^k + S_2^{I2} + S_3^{I3}}(x) f_{T_{12}}(t)\, ds\, dx\, dt$$

Similarly we can write down all possible states.

Steady state distribution π as seen by an arrival can be found by solving $\pi P = \pi$ using the transition matrix *P* of the Markov chain analyzed above. In routers, the buffer size is limited hence the numerical calculation of π is straightforward.

To the best of our knowledge, no previous analytical expressions are available for the waiting time of a G/M/1 queue with this polling scheduler. As in our model, during each cycle, the scheduler serves 4 packets from queue 1 and only one packet from queue 2 queue 3 each. We study queue 1 in detail. Consider the steady state distribution at the time of packet arrivals to queue 1. An arriving packet of class 1 will wait for the service completion of the one already in service plus the service times of packets in queue 1, 2 and 3 according to the limited service logic. Since we have considered queue 2 and queue 3 as a single queue and denote it as QUEUE I_2 So there are two possibilities: $i_1 / 2 < I_2$ and $i_1 / 2 \geq I_2$. By considering these two possibilities and the scheduling logic we can find the exact bounds for class 1 packet as follow:

$$C_1 \leq E[W_1] \leq C_1'$$

$$C_1 = \sum_{j_1=1}^{J_1-1} \sum_{j_2=0}^{\lfloor J_1/4 \rfloor} \sum_{j_3=0}^{\lfloor J_1/4 \rfloor} \pi(j_1, j_2, j_3, a_1, s_1^1)\{1/\mu_1 + (j_1 - 1)/\mu_1 + j_2/\mu_2 + j_3/\mu_3\} + \sum_{j_1=1}^{J_1-1} \sum_{j_2=\lfloor J_1/4 \rfloor}^{J_2} \sum_{j_3=\lfloor J_1/4 \rfloor}^{J_3} \pi(j_1, j_2, j_3, a_1, s_1^1)\{1/\mu_1 + (j_1 - 1)/\mu_1 + \lfloor j_1/4 \rfloor/\mu_2 + \lfloor j_1/4 \rfloor/\mu_3\} +$$

$$\sum_{j_1=1}^{J_1-1} \sum_{j_2=0}^{\lfloor J_1/4 \rfloor} \sum_{j_3=0}^{\lfloor J_1/4 \rfloor} \pi(j_1, j_2, j_3, a_1, s_1^2)\{1/\mu_1 + (j_1 - 1)/\mu_1 + j_2/\mu_2 + j_3/\mu_3\} + \sum_{j_1=1}^{J_1-1} \sum_{j_2=\lceil J_1/4 \rceil}^{J_2} \sum_{j_3=\lceil J_1/4 \rceil}^{J_3} \pi(j_1, j_2, j_3, a_1, s_1^2)\{1/\mu_1 + (j_1 - 1)/\mu_1 + \lceil j_1/4 \rceil/\mu_2 + \lceil j_1/4 \rceil/\mu_3\} +$$

$$\sum_{j_1=1}^{J_1-1}\sum_{j_2=0}^{\lfloor J_1/4\rfloor}\sum_{j_3=0}^{\lfloor J_1/4\rfloor}\pi(j_1,j_2,j_3,a_1,s_1^3)\{1/\mu_1+(j_1-1)/\mu_1+j_2/\mu_2+j_3/\mu_3\}+\sum_{j_1=1}^{J_1-1}\sum_{j_2=\lceil J_1/4\rceil}^{J_2}\sum_{j_3=\lceil J_1/4\rceil}^{J_3}\pi(j_1,j_2,j_3,a_1,s_1^3)\{1/\mu_1+(j_1-1)/\mu_1+\lfloor j_1/4\rfloor/\mu_2+\lfloor j_1/4\rfloor/\mu_3\}+$$
$$\sum_{j_1=1}^{J_1-1}\sum_{j_2=0}^{\lfloor J_1/4\rfloor}\sum_{j_3=0}^{\lfloor J_1/4\rfloor}\pi(j_1,j_2,j_3,a_1,s_1^4)\{1/\mu_1+(j_1-1)/\mu_1+j_2/\mu_2+j_3/\mu_3\}+\sum_{j_1=1}^{J_1-1}\sum_{j_2=\lceil J_1/4\rceil}^{J_2}\sum_{j_3=\lceil J_1/4\rceil}^{J_3}\pi(j_1,j_2,j_3,a_1,s_1^4)\{1/\mu_1+(j_1-1)/\mu_1+\lceil j_1/4\rceil/\mu_2+\lceil j_1/4\rceil/\mu_3\}+$$
$$\sum_{j_1=0}^{J_1-1}\sum_{j_2=1}^{\lfloor J_1/4\rfloor}\sum_{j_3=0}^{\lfloor J_1/4\rfloor}\pi(j_1,j_2,j_3,a_1,s_2)\{1/\mu_2+j_1/\mu_1+(j_2-1)/\mu_2+j_3/\mu_3\}+\sum_{j_1=0}^{J_1-1}\sum_{j_2=\lceil J_1/4\rceil}^{J_2}\sum_{j_3=\lceil J_1/4\rceil}^{J_3}\pi(j_1,j_2,j_3,a_1,s_2)\{1/\mu_2+j_1/\mu_1+\lfloor j_1/4\rfloor/\mu_2+\lfloor j_1/4\rfloor/\mu_3\}+$$
$$\sum_{j_1=0}^{J_1-1}\sum_{j_2=0}^{\lfloor J_1/4\rfloor}\sum_{j_3=1}^{\lfloor J_1/4\rfloor}\pi(j_1,j_2,j_3,a_1,s_3)\{1/\mu_3+j_1/\mu_1+j_2/\mu_2+(j_3-1)/\mu_3\}+\sum_{j_1=0}^{J_1-1}\sum_{j_2=\lceil J_1/4\rceil}^{J_2}\sum_{j_3=\lceil J_1/4\rceil}^{J_3}\pi(j_1,j_2,j_3,a_1,s_3)\{1/\mu_3+j_1/\mu_1+\lfloor j_1/4\rfloor/\mu_2+\lfloor j_1/4\rfloor/\mu_3\}$$

$$C_1^{/}=\sum_{j_1=1}^{J_1-1}\sum_{j_2=0}^{J_2}\sum_{j_3=0}^{J_3}\pi(j_1,j_2,j_3,a_1,s_1^1)\{1/\mu_1+(j_1-1)/\mu_1+\lfloor j_1/4\rfloor/\mu_2+\lfloor j_1/4\rfloor/\mu_3\}+\sum_{j_1=1}^{J_1-1}\sum_{j_2=0}^{J_2}\sum_{j_3=0}^{J_3}\pi(j_1,j_2,j_3,a_1,s_1^2)\{1/\mu_1+(j_1-1)/\mu_1+\lceil j_1/4\rceil/\mu_2+\lceil j_1/4\rceil/\mu_3\}+$$
$$\sum_{j_1=1}^{J_1-1}\sum_{j_2=0}^{J_2}\sum_{j_3=0}^{J_3}\pi(j_1,j_2,j_3,a_1,s_1^3)\{1/\mu_1+(j_1-1)/\mu_1+\lfloor j_1/4\rfloor/\mu_2+\lfloor j_1/4\rfloor/\mu_3\}+\sum_{j_1=1}^{J_1-1}\sum_{j_2=0}^{J_2}\sum_{j_3=0}^{J_3}\pi(j_1,j_2,j_3,a_1,s_1^4)\{1/\mu_1+(j_1-1)/\mu_1+\lceil j_1/4\rceil/\mu_2+\lceil j_1/4\rceil/\mu_3\}+$$
$$\sum_{j_1=0}^{J_1-1}\sum_{j_2=1}^{J_2}\sum_{j_3=0}^{J_3}\pi(j_1,j_2,j_3,a_1,s_2)\{1/\mu_2+j_1/\mu_1+\lfloor j_1/4\rfloor/\mu_2+\lfloor j_1/4\rfloor/\mu_3\}+\sum_{j_1=0}^{J_1-1}\sum_{j_2=0}^{J_2}\sum_{j_3=1}^{J_3}\pi(j_1,j_2,j_3,a_1,s_3)\{1/\mu_3+j_1/\mu_1+\lfloor j_1/4\rfloor/\mu_2+\lfloor j_1/4\rfloor/\mu_3\}$$

Since the scheduler serves one packet from queue 2 and one packet from queue 3 during each cycle, hence the expected time for a randomly arriving class 2/class 3 packet will be same. Therefore, we can write down the bounds on expected waiting time of class 2/3 packet as: $C_2 < E[W_2] = C_2^{/}$,

$$C_2=\sum_{j_2=0}^{J_2-1}\sum_{j_3=0}^{J_3}\sum_{j_1=1}^{4j_2+4}\pi(j_1,j_2,j_3,a_2,s_1^1)\{1/\mu_1+(j_1-1)/\mu_1+j_2/\mu_2+j_2/\mu_3\}+\sum_{j_2=0}^{J_2-1}\sum_{j_3=0}^{J_3}\sum_{j_1=4j_2+5}^{J_1}\pi(j_1,j_2,j_3,a_2,s_1^1)(1/\mu_1+(4j_2+3)/\mu_1+j_2/\mu_2+j_2/\mu_3\}+$$
$$\sum_{j_2=0}^{J_2-1}\sum_{j_3=0}^{J_3}\sum_{j_1=1}^{4j_2+3}\pi(j_1,j_2,j_3,a_2,s_1^2)\{1/\mu_1+(j_1-1)/\mu_1+j_2/\mu_2+j_2/\mu_3\}+\sum_{j_2=0}^{J_2-1}\sum_{j_3=0}^{J_3}\sum_{j_1=4j_2+4}^{J_1}\pi(j_1,j_2,j_3,a_2,s_1^2)\{1/\mu_1+(4j_2+2)/\mu_1+j_2/\mu_2+j_2/\mu_3\}+$$
$$\sum_{j_2=0}^{J_2-1}\sum_{j_3=0}^{J_3}\sum_{j_1=1}^{4j_2+2}\pi(j_1,j_2,j_3,a_2,s_1^3)\{1/\mu_1+(j_1-1)/\mu_1+j_2/\mu_2+j_2/\mu_3\}+\sum_{j_2=0}^{J_2-1}\sum_{j_3=0}^{J_3}\sum_{j_1=4j_2+3}^{J_1}\pi(j_1,j_2,j_3,a_2,s_1^3)(1/\mu_1+(4j_2+1)/\mu_1+j_2/\mu_2+j_2/\mu_3\}+$$
$$\sum_{j_2=0}^{J_2-1}\sum_{j_3=0}^{J_3}\sum_{j_1=1}^{4j_2+1}\pi(j_1,j_2,j_3,a_2,s_1^4)\{1/\mu_1+(j_1-1)/\mu_1+j_2/\mu_2+j_2/\mu_3\}+\sum_{j_2=0}^{J_2-1}\sum_{j_3=0}^{J_3}\sum_{j_1=4j_2+2}^{J_1}\pi(j_1,j_2,j_3,a_2,s_1^4)\{1/\mu_1+4j_2/\mu_1+j_2/\mu_2+j_2/\mu_3\}+$$
$$\sum_{j_2=1}^{J_2-1}\sum_{j_3=0}^{J_3}\sum_{j_1=0}^{4j_2}\pi(j_1,j_2,j_3,a_2,s_2)\{1/\mu_2+j_1/\mu_1+(j_2-1)/\mu_2+j_2/\mu_3\}+\sum_{j_2=1}^{J_2-1}\sum_{j_3=0}^{J_3}\sum_{j_1=4j_2+1}^{J_1}\pi(j_1,j_2,j_3,a_2,s_2)\{1/\mu_2+(j_2-1)/\mu_2+4j_2/\mu_1+j_2/\mu_3\}+$$
$$\sum_{j_2=0}^{J_2-1}\sum_{j_3=1}^{J_3}\sum_{j_1=0}^{4j_2}\pi(j_1,j_2,j_3,a_2,s_3)\{1/\mu_3+j_2/\mu_2+j_1/\mu_1+j_2/\mu_3\}+\sum_{j_2=0}^{J_2-1}\sum_{j_3=1}^{J_3}\sum_{j_1=4j_2+1}^{J_1}\pi(j_1,j_2,j_3,a_2,s_3)\{1/\mu_3+j_2/\mu_2+(4j_2+4)/\mu_1+j_2/\mu_3\}$$

$$C_2^{/}=\sum_{j_2=0}^{J_2-1}\sum_{j_3=0}^{J_3}\sum_{j_1=1}^{J_1}\pi(j_1,j_2,j_3,a_2,s_1^1)(1/\mu_1+(4j_2+3)/\mu_1+j_2/\mu_2+j_2/\mu_3\}+\sum_{j_2=0}^{J_2-1}\sum_{j_3=0}^{J_3}\sum_{j_1=1}^{J_1}\pi(j_1,j_2,j_3,a_2,s_1^2)\{1/\mu_1+(4j_2+2)/\mu_1+j_2/\mu_2+j_2/\mu_3\}+$$
$$\sum_{j_2=0}^{J_2-1}\sum_{j_3=0}^{J_3}\sum_{j_1=1}^{J_1}\pi(j_1,j_2,j_3,a_2,s_1^3)(1/\mu_1+(4j_2+1)/\mu_1+j_2/\mu_2+j_2/\mu_3\}+\sum_{j_2=0}^{J_2-1}\sum_{j_3=0}^{J_3}\sum_{j_1=1}^{J_1}\pi(j_1,j_2,j_3,a_2,s_1^4)\{1/\mu_1+4j_2/\mu_1+j_2/\mu_2+j_2/\mu_3\}+$$
$$\sum_{j_2=1}^{J_2-1}\sum_{j_3=0}^{J_3}\sum_{j_1=0}^{J_1}\pi(j_1,j_2,j_3,a_2,s_2)\{1/\mu_2+(j_2-1)/\mu_2+4j_2/\mu_1+j_2/\mu_3\}+\sum_{j_2=0}^{J_2-1}\sum_{j_3=1}^{J_3}\sum_{j_1=0}^{J_1}\pi(j_1,j_2,j_3,a_2,s_3)\{1/\mu_3+j_2/\mu_2+(4j_2+4)/\mu_1+j_2/\mu_3\}$$

5. SIMULATION RESULTS

A comprehensive discrete-event simulator was built to understand and evaluate the QoS behaviour of self-similar traffic under newly proposed scheduling logic. The simulation engine is highly modular by design allowing free customization of the traffic generator and the scheduling logic. This allows for the ready evaluation of any scheduling discipline under any specific kind of input traffic. The key element for the scheduler logic is the `Scheduler` class. Here we used the template method design pattern [32]. This allows any scheduling algorithm to be loosely coupled but easily integrated, overriding the existing program skeleton. `LimitedServicePollingScheduler` was actually implemented to analyze the corresponding QoS behaviour. For the higher priority class (class 1 packets), we set the session arrival rate to $\lambda_1 = 6s^{-1}$, the in-session packet arrival rate to $\alpha_1 = 50s^{-1}$ (the characteristic of VoIP traffic) and the service rate to $\mu_1 = 2500s^{-1}$. For queue 2 and queue 3, we set the session arrival rate $\lambda_2 = \lambda_3 = 50s^{-1}$, the in-session packet arrival rate to $\alpha_2 = \alpha_3 = 6s^{-1}$ and the service rate to $\mu_2 = \mu_3 = \mu_1$. We investigated the effects of varying the Hurst parameter (0.5<H<1) on various QoS parameters. The QoS results from the simulation studies with 95% confidence interval are presented. Gross et al. study a related issue in detail in [32] and conclude that care must be taken in simulations involving Pareto distributions as they can lead to large errors

due to the heavy tail. It should also be noted though, that the bulk of empirical evidence [33] suggests that H ~ [0.7, 0.85] is the region of interest in network traffic. Fig. 1 shows waiting time (in ms), Packet Loss Rate (PLR) and Queue length vs. Hurst Parameter for limited service polling model. We can see the significant detrimental impact of increasing the Hurst Parameter (the degree of self-similarity) on the QoS offered. We can also notice the effect of this novel proposed polling scheduler, as the burstiness of the traffic increases, we see a significant increase in the QoS parameters of low priority queues. But this scheduler easily outperforms other traditional schedulers such as priority. In our prior work, we have implemented priority scheduler [30] where we noticed that at Hurst parameter 0.9, the queueing delay for lowest priority i.e. queue no. 3 was round about 30 *ms*. But here the queueing delay for lowest priority queue (queue no. 3) at H=0.9 is almost 12 *ms*, which shows a major improvement. Our scheduler on one side not only provides a priority service to real time traffic (queue 1) whilst at the same time, it provides a fair service to low priority queues as well (queue 2 and queue 3) so that low priority queues must not be starved for bandwidth.

6. APPLICATIONS OF THE MODEL

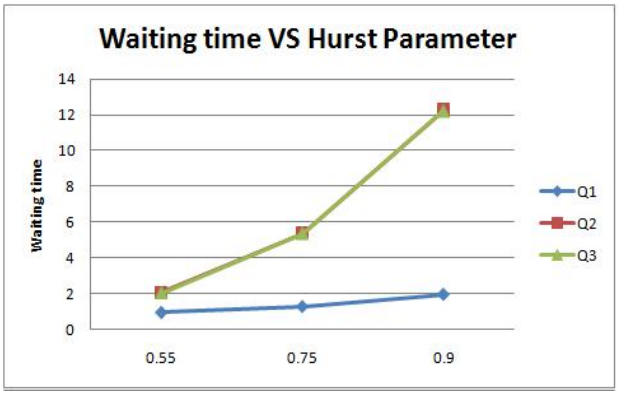
Here we give an overview of the prime application of the model. 4G systems are leaning towards all-IP network architecture for transporting IP multimedia services. To transport 4G services through IP networks without loosing end-to-end QoS provisioning, a consistent and efficient QoS mapping between 4G traffic classes and IP QoS classes is required. According to 3GPP, UMTS-to-IP QoS mapping is performed by a translation function in the GGSN router that classifies each UMTS packet flow and maps it to a suitable IP QoS class [34]. In order to make accurate mappings and to ensure guaranteed QoS parameters to the end user of mobile Internet, it is essential to being able to accurately model the end-to-end behavior of different classes of wireless IP traffic (conversational, interactive, streaming and background) passing through IP QoS domain. Our model is directly applicable to the problem of determining the end-to-end queueing behavior of IP traffic through both Wired and wireless IP domains, but modeling accuracy is more crucial in resource constrained environments such as wireless networks. For example, our model is directly able to analyze the behavior of different QoS classes of 4G traffic passing through a IP QoS domain, in which routers are implemented with the proposed scheduling logic. Thus, the model enables tighter bounds on actual behavior so that over-provisioning can be minimized. It also enables translations of traffic behavior between different kinds of QoS domains so that it is possible to map reservations made in different domains to provide session continuity.

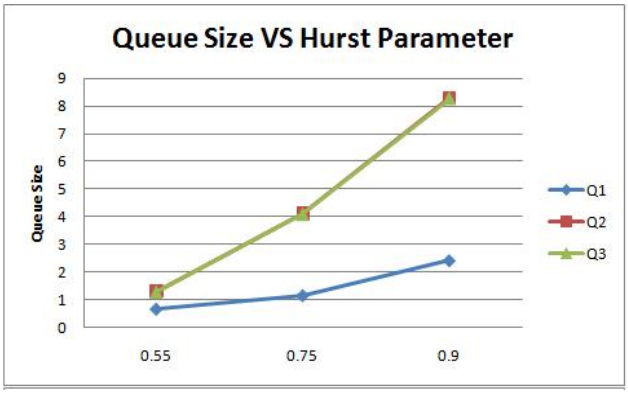
7. CONCLUSION AND FUTURE WORK

In this paper, we have presented a novel analytical model based on G/M/1 queueing system for accurate modeling of wireless IP traffic behavior under the assumption of three different classes of self-similar traffic. We have analyzed it on the basis of limited service polling service and explicit expressions of expected waiting time for the corresponding classes have been derived. We have also implemented a discrete event simulator to simulate the QoS behavior of multiple classes of self-similar traffic. The simulation results quantify the affect of implementation of new proposed scheduler in IP network. The model represents an important step towards the overall aim of finding realistic (under self-similar traffic assumptions) end-to-end QoS behavior (in terms of QoS parameters such as delay, jitter and throughput) of multiple traffic classes passing through heterogeneous wireless IP domains (IntServ, DiffServ and MPLS). Our future work will focus to analyze the performance of different QoS domains implemented with different queueing disciplines. Further we intend to implement a test bed to validate our proposed QoS models.

8. Acknowledgment

The current work is supported by NPST project grant (10-INF1112-02) at King Saud University, Riyadh, Saudi Arabia.





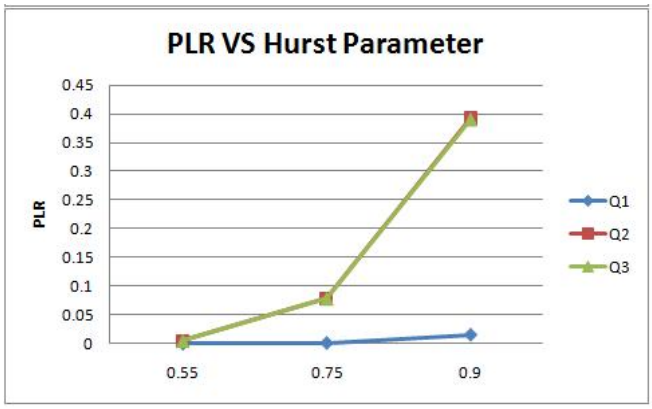


Fig. 1: Expected Waiting time (in *ms*), queue length and PLR vs. Hurst Parameter for 3 classes of traffic in limited service polling scheme

9. References

[1] J. Yang and I. Kriaras, "Migration to all-IP based UMTS networks, " *IEEE 1st International Conference on 3G Mobile Communication Technologies* , 27-29 March, 2000, pp. 19-23

[2] K. Venken, J. De Vriendt and D. De Vleeschauwer, "Designing a DiffServ-capable IP-backbone for the UTRAN", *IEEE 2nd International Conference on 3G Mobile Communication Technologies*, 26-28 March 2001, pp. 47-52

[3] S. Maniatis, C. Grecas and I. Venieris, "End-to-End QoS Issues Over Next Generation Mobile Internet", *IEEE Symposium on Communication and Vehicular Technology*, 2000, SVCT-2000, 19 Oct, 2000, pp. 150-154

[4] P. Newman, Netillion Inc. "In Search of the All-IP Mobile Network", *IEEE Communication Magazine,* vol. 42, issue 12, Dec. 2004, pp. S3-S8

[5] G. Araniti, F. Calabro, A. Iera, A. Molinaro and S. Pulitano, "Differentiated Services QoS Issues in Next Generation Radio Access Network: a New Management Policy for Expedited Forwarding Per-Hop Behavior", *IEEE Vehicular Technology Conference*, VTC 2004-Fall, vol. 4, 26-29 Sept. 2004, pp. 2693-2697

[6] S. Uskela, "All IP Architectures for Cellular Networks", *2nd International Conference on 3G Mobile Communication Technologies*, 26-28 March 2001, pp. 180-185

[7] Jeong-Hyun Park, "Wireless Internet Access for Mobile Subscribers Based on GPRS/UMTS Network" *IEEE Communication Magazine*, vol. 40, issue 4, April 2002, pp. 38-39

[8] K. Daniel Wong and Vijay K. Varma, "Supporting Real-Time IP Multimedia Services in UMTS", *IEEE Communication Magazine,* vol. 41, issue 11, Nov. 2003, pp. 148-155

[9] W. Stallings, "Integrated Services Architecture: The Next-Generation Internet", *International Journal of Network Management*, 9, 1999, pp. 38-43

[10] S. Blake et al., "An Architecture for Differentiated Services", *IETF RFC 2475*, Dec. 1998

[11] Rosen E. et al., "Multiprotocol Label Switching (MPLS) Architecture", *RFC 3031*, Jan. 2001

[12] W. Odom and M. J. Cavanaugh, "IP Telephony Self-Study Cisco DQoS Exam Certification Guide", Cisco Press, 2004, pp. 3-314

[13] R. Chakravorty, J. Cartwright and I. Pratt, "Practical Experience with TCP over GPRS", *in IEEE GlobeCom*, Nov. 2002

[14] D. Schwab and R. Bunt, "Characterizing the use of a Campus Wireless Network", *in IEEE INFOCOM,* March 2004

[15] X. Meng, S. Wong, Y. Yuan and S.Lu, "Characterizing Flows in Large Wireless Data Networks", *in ACM Mobicom*, Sep 2004

[16] A. Balachandran, G. M. Voelker, P. Bahl and P. Venkat Rangan, "Characterizing user behavior and network performance in a public Wireless LAN", *Sigmetrics Performance Evaluation. Review*, vol. 30. no. 1, 2002, pp. 195-205

[17] H. Takagi, "Analysis of Polling Systems", Research Reports and Notes, Computer System Series, The MIT press, 1986

[18] O. J. Boxma and B. W. Meister, "Waiting-Time Approximations in Multiple Queue Systems with Cyclic Service", Performance Evaluation 7 (1987), pp. 59-70

[19] O. J. Boxma and B. W. Meister, "Waiting-Time Approximations for Cyclic-Service Systems with Switchover Times", Performance Evaluation 7 (1987), pp. 299-308

[20] A. Krendzel, Y. Koucheryavy, J. Harju and S. Lopatin, "Traffic and QoS management in Wireless Multimedia Networks" COST 290:: *Wi-QoST*, Working group N3 http://www.cost290.org

[21] M. Jiang, M. Nikolic, S. Hardy and L. Trajkovic, "Impact of Self-Similarity on Wireless Data Network Performance", *IEEE ICC,* 2001, vol. 2, pp. 477-481

[22] J. Ridoux, A. Nucci and D. Veitch, "Characterization of Wireless Traffic based on Semi-Experiments", *Technical Report-LIP6*, December 2005

[23] Z. Sahinoglu and S. Tekinay, "On Multimedia Networks: Self-Similar Traffic and Network Performance", *IEEE Communication Magazine*, vol. 37, issue 1, Jan. 1999, pp. 48-52

[24] I. Norros, "On the use of Fractional Brownian Motion in theory of connectionless networks", *IEEE Journal on Selected Areas in Communications*, vol. 13. no. 6, August 1995, pp. 953-962

[25] P. Benko, G. Malicsko and A. Veres, "A Large-scale, passive analysis of end-to-end TCP Performances over GPRS", *in IEEE INFOCOM*, March 2004

[26] M. Caglar, "A Long-Range Dependant Workload Model for Packet Data Traffic", *Mathematics of Operations Research*, 29, 2004, pp. 92-105

[27] H. P. Schwefel, L. Lipsky, "Impact of aggregated self-similar ON/OFF traffic on delay in stationary queueing models (extended version)", *Performance Evaluation*, 43, 2001, pp. 203-221

[28] I. Kaj, "Limiting fractal random processes in heavy-tailed systems", In Fractals in Engineering, New Trends in Theory and Applications, Eds.J. Levy-Lehel, E. Lutton, *Springer-Verlag London,* 2005, pp. 199-218

[29] M. Iftikhar et al, "Multiclass G/M/1 queueing system with self-similar input and non-preemptive priority" Journal of Computer Communications, Elsevier, vol. 31, issue 5, pp. 1012-1027

[30] M. Iftikhar et. al, "Towards the formation of comprehensive SLAs between heterogeneous wireless DiffServ domains" Springer Journal of Telecommunication Systems, (2009) 42: 179-199

[31] E. Cinlar, Introduction to Stochastic Processes, 1975

[32] D. Gross et al, 'Difficulties in simulating queues with Pareto service, " in proc. of the 2002 winter simulation conference

[33] Gi. Kihong Park et al, "On the relationship between file size, transport protocols and self-similar network traffic," in proc. of International conference on network protocols, Oct. 1996. pp. 171-180

[34] R. Ben Ali, Y Lemieux and S. Pierre, "UMTS-to-IP QoS Mapping for Voice and Video Telephony Services, *IEEE Network* , vol. 19, issue 2, March/April 2005, pp. 26-32

[35] M. Iftikhar et al. , "An Alternating Service Model with Self-Similar Traffic Input to Provide Guaranteed QoS in Wireless Internet", in proc. of IEEE ICI, Uzbekistan, 19-21 Sept. 2006