TRAFFIC DELAY ESTIMATION FOR 3G MOBILE IP SERVICES

ISSN: 2141 – 3290
www.wojast.com

[1]EKPENYONG, M. E. AND [2]ISABONA J.
[1]*Department of Computer Science*
University of Uyo, P.M.B. 1017, Uyo, Nigeria,
mosesekpenyong@gmail.com
[2]*Benson Idahosa University, P.M.B. 1100, Benin City, Benin*
josabone@yahoo.com

ABSTRACT: An important Quality of Service (QoS) factor for wireless network planning is traffic delay. In any telecommunication system, traffic predicts the network capacity, and demands efficient resource (e.g., bandwidth) management as a requirement to convey the expected traffic load. In this paper, we propose a queuing model and resource allocation strategy that evaluates the network throughput (for users) performance. The proposed model is capable of giving a true picture of packet losses and unexpected delays in the event of climate change. One of the effects of climate change considered in this paper is fading. Fading, can notoriously hamper the network, and could cause noticeable delays during communication, thus presenting annoying side effects such as cross talking, voice echoing, not reachable terminals and dropped calls. We evaluate the performance of the proposed model by studying the blocking and delay properties of an existing third generation (3G) cellular system and simulate the model in practice. Simulation results show that the obtained delay is below the third generation partnership project (3GPP) standard specification, and can be used by network operators to minimize traffic delays in data and Voice over Internet Protocol (VoIP) services.

INTRODUCTION

The transition of mobile cellular systems from second generation (2G) to third generation (3G) technology is driven by the popularity in the mobile voice services and demand for other services such as data, video and multimedia applications, and international roaming. New data services in 3G systems can have dramatic impact on the usage of the cellular network, particularly as emerging applications perversely grow. Also, the rapid expansion of users understanding of the current network traffic characteristic is vital for capacity planning of future wireless/cellular data networks (Adya, *et. al* 2001; Crovella and Taqqu, 1999). Given this dramatic change, the path of next generation wireless network is definitely leading us towards an all IP network, where voice will be transformed into packets. The primary requirements of 3G systems (Moyo and Chitamu, 2000; Kasera and Narang, 2004) include:

- Support of bit rates up to 2Mbps
- Mechanisms to efficiently support "bandwidth on demand" and variable bit rate services.
- Spectrum for services with different quality of service requirements (e.g. speech, video, packet data, etc.), which differ in terms of:
  - delay requirements: support for both real-time and non-real time services.
  - error requirements: support for bit error rate as low as 10^{-6}.
  - bandwidth requirements: support for services requiring variable bandwidth, and also includes support for asymmetric uplink and downlink bandwidths for application such as web-browsing.

- Simultaneous co-existence with 2G systems and support for inter-system handover.
- Higher spectrum efficiency.

These requirements in addition to the limited spectrum availability, present an increasing demand for bandwidth efficient multiple access schemes. At the moment, the Wideband Code Division Multiple Access (WCDMA) is highly preferred as the multiple access technology for 3G systems. When compared with other commercial mobile technologies, CDMA provides better capacity for uninterrupted voice and data communications, which allows for seamless connectivity and represents the most common platform for building 3G technologies (Mohan and Ravichandran, 2009). In 3G WCDMA systems, data applications are expected to finally dominate the overall traffic volume (Grewal and Dedourek, 2004), because the traffic generated by current data applications is inherently “bursty” and asymmetric, with higher data rate.

Traffic in the upcoming wireless networks is expected to be purely non-stationary. The traffic of real-time classes, observed at the Base Stations (BS) is anticipated to vary as a result of the small cell size and increased handoff rates. It is natural in a non-stationary environment, that the performance of the scheduler gets degraded due to the presence of arrival rate estimation errors. Also, inefficient radio resource distribution and unequal delay variation from the target mainly results from estimation errors (Mohan and Ravichandran, 2009). To address this problem, we concentrate on the traffic load and compute the queuing delay based on a traffic model, with a proposed radio resource management algorithm that controls the queuing delay.

EFFECT OF FADING ON CELLULAR TRAFFIC

The presence of reflectors in the environment surrounding a transmitter and receiver could create multiple paths that a transmitted signal can traverse. As a result, the receiver sees the superposition of multiple copies of the transmitted signal, each traversing a different path. Each signal copy experiences differences in attenuation, delay and phase shift while travelling from the source to the receiver. This can result in either constructive or destructive interference, amplifying or attenuating the signal power seen at the receiver. Strong destructive interference is frequently referred to as deep fade and may result in temporary failure of communication, due to a severe drop in the channel signal-to-noise ratio. Fading channel models are often used to model the effects of electromagnetic transmission of information over the air in cellular networks and broadcast communication. Fading channel models are also used in underwater acoustic communications to model the distortion caused by the water. Mathematically, fading is usually modelled as a time-varying random change in the amplitude and phase of the transmitted signal. In this paper, we report on the performance of the cellular network with and without fading. The essence is to provide an insight into the cause of network traffic problems and the need for an optimised approach to solving these problems.

The idea of traffic modelling lies in constructing models that captures the important statistical properties of the underlying (measured) data. Traditional traffic models which predict the aggregate traffic processed by the switches have been developed for wire-line networks. Unlike fixed networks, cellular telephone traffic must support mobile customers. Traffic is implemented as a queue or network of queues where various performance measures of the system are calculated or estimated. These measures include the queue length, server utilization, waiting time and traffic loss. In most estimation techniques, it is assumed that the mobile network is homogeneous, with independent cells, and studied using stochastic models (Hong and Rappaport, 1986). In such systems, every cell is considered to be statistically identical and independent to each other. The model presented in (Nguyen, 2005), which conforms with the model in the third generation partnership project (3GPP) of 1998, uses a multi-layer model to describe sources, and at the lowest layer, implements the Weibull and Pareto distributions for ON and OFF packet durations, respectively, of internet traffic.

Case Study

The essence of using experimental data is to enhance the accuracy of predictions. In this paper, we obtained two classes of data: the traffic and Erlang capacity data at busy hours. These data were measured from the Globacomm (GLO) Nigeria main control-switch room at the Lagos Headquarters in Nigeria. GLO has a total of 6,742 base stations servicing the Nigerian airspace, with a sectorised cell capacity of 2000 mobile users. The data were obtained over a period of two weeks for the purpose of simulating the performance of the system under study.

Figures 1 and 3 show the traffic occurrence and capacity spread at the various base stations within the study period. These figures reveal that some base stations were not fully utilized (i.e., either experienced no traffic or had less capacity to manage). This is explained by the long tails observed in both figures.

In Figures 2 and 4, we extract and plot the average traffic and Erlang capacity of the system under study. We observed that on the average, the traffic and network capacity fluctuate over the period and are not significantly influenced by the number of base stations, as evident from the respective R^2 values (i.e., 0.0213 and 0.0255, respectively). The imperfect nature of the traffic could therefore be attributed to propagation related factors such as users' mobility, fading due to weather change, congestion, blocking, etc.; whereas capacity fluctuation could be as a result of variations in number of users accessing the network during peak and non-peak periods. Trend line equations are fit in the plots to enable the prediction of new empirical results.

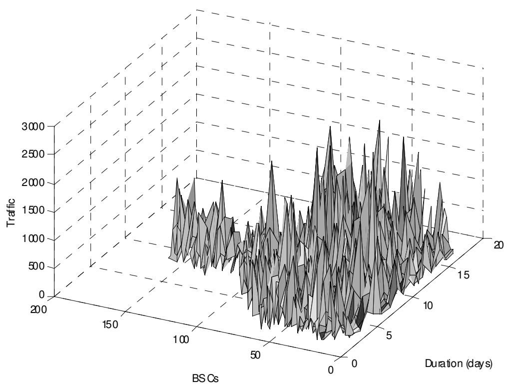

Figure 1. Traffic analysis at busy hours

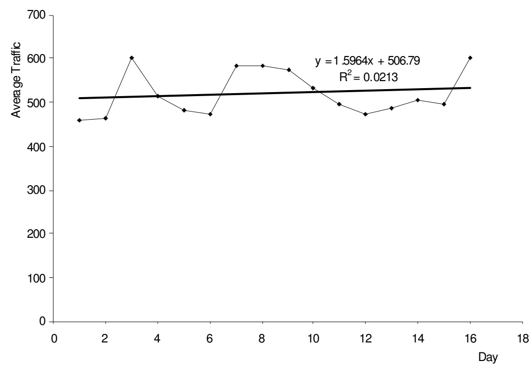

Figure 2. Average traffic vs. duration

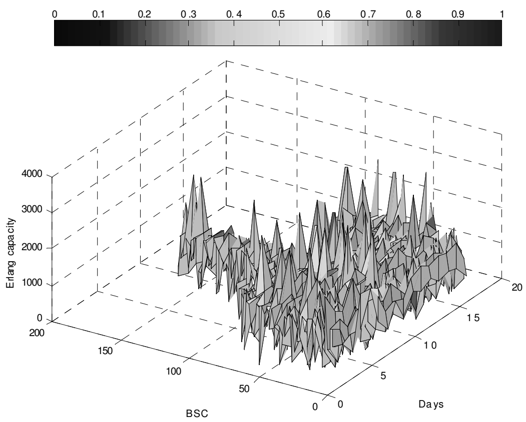

Figure 3. Erlang capacity analysis

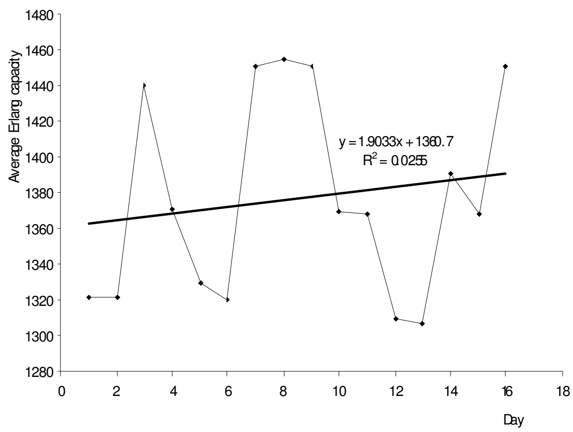

Figure 4. Average Erlang capacity vs. duration

SYSTEM MODEL

Blocking Probability Model

Data traffic can be conveyed either through a circuit or packet-switched system. Circuit-switched systems usually have constant bit rates, while packet-switched systems may have variable bit rates. To handle data traffic, we examined an m-server queuing system with *k* customers (k > m). Besides, we assume that the population is infinite (M/M/m/k). This model is suitable for data services that are delay tolerant, because when all *m* channels are busy and upon acceptance of a new call attempt, the call can immediately join the queue. The birth and death coefficients as shown in Figure 5 are:

$$\lambda_n = \begin{cases} \lambda, & n < k-1 \\ 0, & n \geq k \end{cases} \tag{1}$$

and

$$\mu_n = \begin{cases} n\mu, & n \leq m \\ m\mu, & n \geq n \leq k \\ 0, & k < n \end{cases} \tag{2}$$

where *n* represents the number of subscribers in the queue whose call attempts have been accepted.

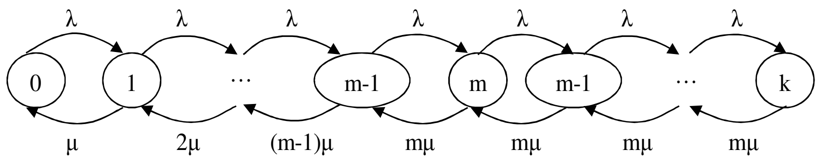

Figure 5. State transition-rate diagram for m-servers, finite storage k and infinite population (M/M/m/k).

Assuming P_n is the probability of being in the state *n* (or there exist *n* subscribers in the system), P_n can be obtained as:

$$P_n = C_n P_o \tag{3}$$

where

$$C_n = \begin{cases} \left(\frac{\lambda}{\mu}\right)^n \frac{1}{n!}, & n < m \\ \left(\frac{\lambda}{\mu}\right)^n \frac{1}{m!}\frac{1}{m^{n-m}}, & m \leq n \leq k \\ 0, & k < n \end{cases} \tag{4}$$

and

$$P_0 = \begin{cases} \dfrac{1}{1+\sum_{n=1}^{m-1}\left(\frac{\lambda}{\mu}\right)^n \frac{1}{n!}}, & n<m \\ \dfrac{1}{1+\sum_{n=1}^{m-1}\left(\frac{\lambda}{\mu}\right)^n \frac{1}{n!}+\frac{1}{m!}\sum_{n=m}^{k}\left(\frac{\lambda}{\mu}\right)^n \frac{1}{m^{n-m}}}, & m\leq n\leq k \end{cases} \tag{5}$$

replacing γ with $\lambda/_{m\mu}$ in equation (5), we have,

$$P_0 = \begin{cases} \dfrac{1}{1+\sum_{n=1}^{m-1}\left(\frac{\lambda}{\mu}\right)^n \frac{1}{n!}}, & n<m \\ \dfrac{1}{\sum_{n=0}^{m}\left(\frac{\lambda}{\mu}\right)^n \frac{1}{n!} + \frac{1}{m!}\left(\frac{\lambda}{\mu}\right)^m \gamma \frac{(1-\gamma^{k-m})}{1-\gamma}} & m \le n \le k \end{cases} \tag{6}$$

Now, substituting c_n and P_0 in equations (4) and (5) respectively into equation (3), we arrive at the data blocking probability model:

$$P_{B(data)} = \begin{cases} \left[\left(\frac{\lambda}{\mu}\right)^n \frac{1}{n!}\right]\left[\dfrac{1}{1+\sum_{n=1}^{m-1}\left(\frac{\lambda}{\mu}\right)^n \frac{1}{n!}}\right], & for \quad n<m \\ \left[\left(\frac{\lambda}{\mu}\right)^n \left(\frac{1}{m!}\frac{1}{m^{n-m}}\right)\right]\left[\dfrac{1}{1+\sum_{n=1}^{m-1}\left(\frac{\lambda}{\mu}\right)^n \frac{1}{n!} + \frac{1}{m!}\sum_{n=m}^{k}\left(\frac{\lambda}{\mu}\right)^n \frac{1}{m^{n-m}}}\right] & for \quad m \le n \le k \end{cases} \tag{7}$$

where $P_{B(data)} \cong P_n$, represents the probability of data blocking. This model is suitable for computing blocking probabilities in Telnet, worldwide web and email traffics (Soldani, 2005).

Now, assuming that all servers are busy (i.e., no queue exists), such that some calls are lost, equation (6) reduces to the regular Erlang-B formula (Koo, *et. al*, 2007):

$$P_m = \frac{\dfrac{\left(\frac{\lambda}{\mu}\right)^m}{m!}}{\sum_{k=0}^{m}\dfrac{\left(\lambda/\mu\right)^k}{k!}} \tag{8}$$

where P_m describes the fraction of time all servers are busy. Calls follow an exponential distribution (i.e., has a memory-less property), with λ arrival rate. From equation (8), we deduce the blocking probability for VoIP connection, thus:

$$P_{B(VoIP)} = \frac{\dfrac{\left(\lambda/\mu\right)^{Nc}}{N_c!}}{\sum_{k=0}^{Nc}\dfrac{\left(\lambda/\mu\right)^k}{k!}} \tag{9}$$

where $P_{B(VoIP)}$ is the blocking probability of VoIP (at peak or busy periods)

N_c is the number of trunks or channels.

Queue Delay Model

If T_q is the average long-term waiting time of a subscriber in the queue, then, in accordance with the Little's rule (Leon-Garcia, 2008), we can write

$$L_q = \lambda T_q \tag{10}$$

where L_q is the average number of subscribers in the queue (in the long-term) and can be represented as:

$$L_q = \sum_{n=0}^{\infty} nP_n = \frac{P_0}{m!}\left(\frac{\lambda}{\mu}\right)^m \frac{\gamma}{(1-\gamma)^2}\left[1-\gamma^{k-m+1}-(1-\gamma)(k-m+1)\gamma^{k-m}\right] \qquad (11)$$

where

$$\gamma = \lambda / m\mu \quad \text{and} \quad T_q = \frac{L_q}{\lambda} = \frac{P_o}{m!}\left(\frac{\lambda}{\mu}\right)^m \frac{\gamma}{(1-\gamma)^2}\left[1-\gamma^{k-m+1}-(1-\gamma)(k-m+1)\gamma^{k-m}\right] \qquad (12)$$

with P_0 already defined in equation (6).

Dealing with Fading

In wireless communications, signal fading is caused by multi-path effect. Multi-path effect means that a signal transmitted from a transmitter may have multiple copies traversing different paths to reach a receiver. Thus, at the receiver, the received signal should be the sum of all these multi-path signals. Because the paths traversed by these signals are different; some are longer, and some are shorter. The one at the direction of Light of Signal (LOS) should be the shortest. These signals interact with each other. If signals are in phase, they would intensify the resultant signal; otherwise, the resultant signal is weakened due to out of phase. This phenomenon is called channel fading. In general, there are two criteria to measure channel fading: the doppler spread, and the delay spread.

In this paper, we use Gaussian probabilities to manage fading. The Gaussian probability function is well known and is employed in communication networks to minimise the effect of fading. Thus, for a Gaussian random variable X with mean, $\overline{m}$, and standard deviation, s, the cumulative distribution function is given by

$$F_X(x) = P\{X \le x\} = \frac{1}{2} + \frac{1}{2} erf\left(\frac{x-\overline{m}}{s\sqrt{2}}\right) \qquad (13)$$

were erf is the error function. For the generation of this function, we invoke the MATLAB built-in error function, *erf*, defined by

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \qquad (14)$$

and simulate the reduction of this phenomena (fading) by apply equation (13) to equations (7), (9) and (12), respectively, and observe the extent of fading reduction in both data and voice traffic, as well as the delay profile of the network.

SIMULATION AND DISCUSSION OF RESULTS

Table 1 defines the input parameters used for the simulation. The simulation data are ideal environmental parameters, some of which were captured from the field. Shown in Figures 6 and 7 are plots of blocking probability and traffic intensity for data and VoIP users, respectively, taking into consideration, the number of active users (as modelling parameter). As can be seen in Figure 6, there is a sharp increase in data blocking when k=10, compared to k=5, before the system regains stability. This can be likened to a typical hotspot environment, where the number of users accessing the network is more than the available resources to cater for them at a time. Therefore, proper considerations should be given to hotspot services, especially in urban or densely populated areas during network planning. In Figure 7, there is a linear increase in both parameters (blocking probability and traffic intensity). This is typical of bursty traffics and possible contribution of weather changes, resulting in fading.

Table 1. Simulation model parameters

Parameter	**Value**
Number of subscribers in the queue whose calls have been accepted(n)	n<m = 5, n>m = 8
Total number of servers (m)	$n \le m \le k$: n<m = 3, n>m = 5
Average Traffic intensity(ρ)	0 – 20
Number of trunks or channels (n_c)	5
Total number of customers	5, 10
Average arrival rate(λ)	0.5
Average service rate(μ)	0.5
Probability of no customers in the queue(ρ_0)	0.5
Time in seconds (s)	0.1 - 1

The graphs also reveal the importance of parameter prediction in a network system, as it enables network operators to measure the contribution of certain factors affecting the system's performance. As shown in Figures 6 and 7, other factors contributed to deteriorating the performance of the network (i.e., fading only, gave lower blocking probabilities) and could misinform network operators when estimating the network parameters.

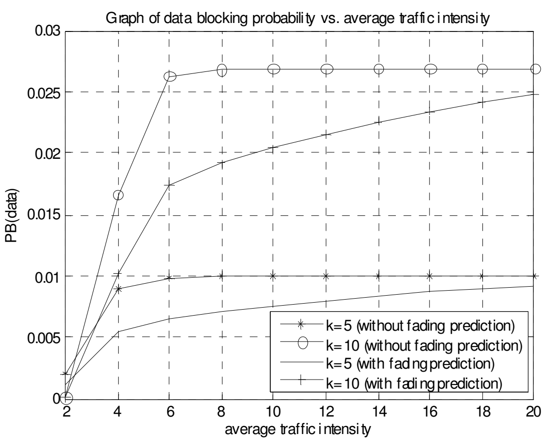

Figure 6. Data blocking probability vs. average traffic intensity

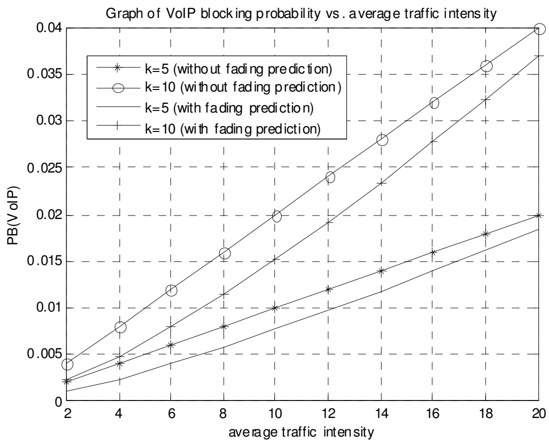

Figure 7. VoIP blocking probability vs. average traffic intensity

Figure 8 shows a graph of the delay performance as a function of traffic intensity in VoIP systems. The graph indicates that when the number of channel increases, the system delay also increases, but as k increases, (for instance, from k = 5 to k = 10), the delay decreases. In practice, if the VoIP call manager (sometimes called the VoIP server) is overwhelmed with requests, or its connection to the network is compromised, call setup delays can reach a point where users abandon their calls before they are able to connect to the other party. This scenario has been observed in recent times in poorly planned networks, as mobile subscribers experience long delays before their calls get connected to the receiver, and would rather prefer to retransmit the request. If IP phones are misconfigured for instance, or the IP connection to the server is impaired, calls remain open in the call queue long after the parties have disengaged or disconnected. Thus to achieve high quality voice, the maximum desired one-way latency should be 150ms (Gonia, 2004). If round trip delays exceed 250ms, voice users will certainly notice delays, and callers will start talking over each other, and sometimes a caller will echo her/himself while transmitting. It is however problematic to determine appropriate QoS classes and performance targets. Delay itself is specified differently (as a mean according to ITU-T, but as a maximum in 3GPP). The ITU-T delay bound for streaming applications and VoIP has a mean value not greater than 1s and more stringent than 3GPP that proposes a maximum value of 280ms (Stafford, 2004). A quick comparison shows that our proposed delay profile is effective in controlling the system, and is well below the recommended standard. The model can also cope with packet losses and unexpected latency within the network. This can be achieved in practice by retransmitting the failed messages (i.e., messages that could not be transmitted successfully) after some predefined condition.

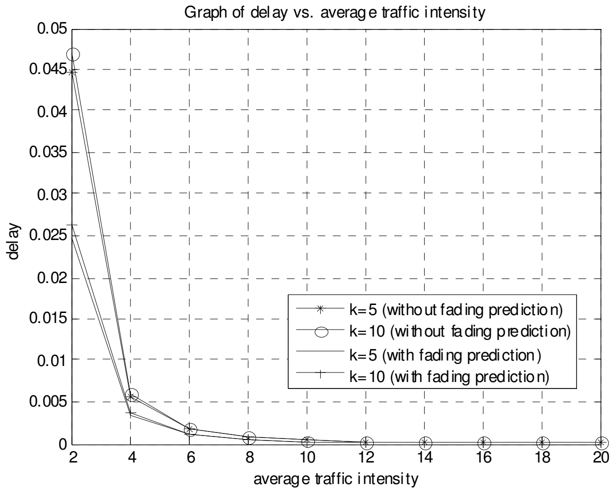

Figure 8. Delay vs. average traffic intensity

Again, with fading prediction, there were reductions in the system's delay for k = 5 and k = 10, respectively. But as the traffic intensity rises, there was no significance difference in network latency.

CONCLUSION

The tremendous growth of the wireless network increases the demand and support of different multimedia (e.g., voice, audio, video, data, etc.) applications available over the network. This however could lead to indeterminate states such as congestions, should the network has to maintain high resources level for the quality of service requirements of such application(s). We have defined performance metrics from steady state distributions, using the respective system parameters of a realistic cellular network and proposed models for computing the blocking probability and queuing delay of the system to eliminate lost calls in a communication network and implemented the traffic delay system using a robust simulation toolkit. Results obtained revealed that control should be placed on the number of users accessing a wireless communication system, with the implementation of an appropriate resource management technique (Dadkhah, *et.al.* 2007) and the prediction of important parameters.

ACKNOWLEDGEMENTS

We appreciate the field engineers of GLOBACOMM Nigeria for their assistance during the data collection phase of this research.

REFERENCES

Adya, A., Bahl, P. and Qiu, L. (2001). Understanding the Browse Patterns of Mobile Clients. In *Proceedings of SIGCOMM Internet Management Workshop (IMW),* San Francisco, California.

Crovella, M. and Taqqu, M. (1999). Estimating the Heavy Tail Index from Scaling Properties. Methodology and Computing. *Applied Probability,* 1(1): 55-79.

Dadkhah, C., Abdoli, M. and Hakkak, M. (2007). A New Radio Resource Management Algorithm for Mixed Traffic Transmission in Mobile System. In: *Proceedings of the 18th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications.*

Gonia, K. (2004). *Latency and QoS for Voice over IP*. SANS Institute, 2004.

Grewal, J. and Dedourek, J. (2004). Provision of QoS in Wireless Networks. In: *Proceedings of 2nd Annual Conference on Communication Networks and Service Research (CNSR'04),* Fredericton, pp. 337-340.

Hong, D. and Rappaport, S. (1986). Traffic Model and Performance Analysis for Cellular Mobile Radio Telephone Systems with Prioritized and Non-prioritized Handoff Procedures". *IEEE Transactions on Vehicular Technology,* 35(3): 77-92.

Kasera, S. and Narang, N. (2004). *3G Networks: Architecture, Protocols and Procedures Based on 3GPP Specifications for UMTs Networks*. Tata-McGraw-Hill Professional, Networking Series.

Koo, I., Yang, J. R. and Kim, K. (2007). Erlang capacity analysis of CDMA systems supporting voice and delay – tolerant data services under delay constraint". *IEEE Transactions on Vehicular Technology*, 56(4): 2375-2385.

Leon-Garcia, A. (2008). *Probability, Statistics and Random Processes for Electrical Engineering*, 3rd Edition, Prentice Hall.

Mohan, N. and Ravichandran, T. (2009). An efficient call admission control and adaptive scheduling for wcdma wireless networks, *European Journal of Scientific Research,* 4: 718-727.

Moyo, R. and Chitamu, P. (2000). Performance and Capacity Issues for 3G WCDMA in Micro and Macro Cellular Environments. In *Proceedings of the South African Telecommunication Networks and Application Conference*.

Nguyen, T. V. (2005). *Capacity Improvement using Adaptive Sectorization in WCDMA Cellular Systems with Non-Uniform and Packet Mode Traffic"*, PhD. Thesis, Victoria University, Melbourne.

Soldani, D. (2005). QoS Management in UMTS Terrestrial Radio Access FDD Networks. PhD. Dissertation, Helsinki University of Technology, Espoo, Finland.

Stafford, M. (2004). *Signaling and Switching for Packet Telephony.* Artech House, USA.