TELETRAFFIC ANALYSIS OF A CALL ADMISSION CONTROL SCHEME WITH TCP PROTOCOL INDUCED FACTORS

T. Walingo*

** Centre of Radio Access and Rural Technologies, University of KwaZulu-Natal, Private Bag X54001, Durban, South Africa E-mail: Walingo@ukzn.ac.za*

Abstract: Future networks face tremendous challenges towards providing guaranteed Quality of Service (QoS) for the multiple traffic types employing numerous protocols. This work presents an expanded parameter Call Admission Control (CAC) scheme and intelligent scheduling to provide QoS on modern networks. The expanded parameter CAC scheme features the lower layers Signal to Interference Ratio (SIR) and network delay as admission parameters. The SIR is due to the transmission on the code division multiple access network and delay is due to the operation of the Transmission Control Protocol (TCP). TCP delay is a good representative of the aggregated delays of the entire network; both access and core network as it is at a higher layer. TCP was ideally designed for wired networks. TCP performance is degraded and leads to substantial delay while operating on wireless lossy links as it reduces the sending rate and resends lost packets. This work presents an analytical framework for evaluating the performance of a wireless TCP based CAC model which features Batch Markovian Arrival Process (BMAP) traffic, a better representative of the future traffic characteristics than the traditional Poisson traffic. A teletraffic analysis of a network with TCP is done and the impact of TCP induced delays on the network is investigated for a future generation CAC scheme.

Keywords: BMAP, Call admission control, CDMA, Multimedia traffic, TCP.

1. INTRODUCTION

Future networks design needs to overcome the numerous network and traffic challenges. The network challenges include those associated with increasing the network capacity, accommodating diverse heterogeneous networks, mobility management and dealing with diverse network protocols e.g. Code Division Multiple Access (CDMA) and TCP with their own inherent problems [1][2][3]. They also handle diverse ever growing applications over cellular networks with their own challenges. The traffic challenges include; diversity in QoS requirements, multimedia traffic types, real time or non-real time with different QoS metrics (BER and delay). Furthermore, the traffic may exhibit other properties such as burstiness, correlation and self-similarity. An efficient CAC scheme is one of the methods employed to guarantee QoS on a CDMA based network. This heterogeneous mix of services with varying QoS metrics and characteristics must be supported and provided with their guaranteed QoS. Each traffic type has got key QoS defining metric(s). These key metric(s) for a traffic type must be identified and made the principal admission parameter(s). The most common QoS parameters for most traffic on a CDMA network are the SIR and delay. SIR has traditionally been the de facto parameter for CAC on CDMA based networks [4][5]. A combination of several admission parameters has been encountered in seldom [6][7]. In this work, the employed CAC considers both SIR and delay. The SIR is on the CDMA wireless link while the delay can be due to scheduling on the wireless link or in the core network. In a realistic network, delay is a result of the effects of the numerous factors; scheduling, processing, transmission and routing. Access network delays and core network delays that arise due to transmission, scheduling and routing protocols have been widely explored as compared to other protocol induced delays [8][9]. Some network protocols like TCP, the transport layer protocol, Automatic Repeat request (ARQ), the link layer protocol, several Media Access Control (MAC) protocols like CDMA introduce delays in networks due to their operation. CDMA and ARQ protocol induced delays have received sufficient attention unlike the TCP induced delays [7][8][10]. TCP is one of the de facto transport protocols on the internet today. There are many TCP variants [11][12]; Tahoe, Reno, New Reno, SACK, Vegas, Fast TCP etc. However, their basic core functionality is the same. TCP was ideally designed for wired networks where the losses are not so great [3]. However, for an *anything anytime anywhere* network, TCP will operate on wireless networks with higher packet losses than expected. This leads to a degradation of the TCP's performance as it was not meant to perform on very lossy links. TCP affects the whole network in the following ways; firstly, during congestion it reduces the sending rate on the network and secondly, when a packet is lost it resends the packet. These factors introduce latency in the whole network. They cause delays and greatly impact on the teletraffic performance of the network. TCP delay is an aggregation of delays at various layers of the protocol stack. It is a comprehensive summary of the user's end to end delay including the access and core network delays. It is therefore imperative that a CAC scheme should feature the delay introduced by TCP. An analytical framework for investigating a CAC for a future Next Generation Networks (NGN) with TCP protocol induced delay is presented in this work.

Modern network models need to be used to effectively evaluate the performance of current and future CAC schemes. The analytical traffic models need to evolve with the evolving traffic characteristics. Traditionally, simple traffic models like the Poisson model have been used. However, they have been shown to be unsuitable for current and future IP based traffic [13]. BMAP traffic models have been favored to evaluate the performance of the CAC scheme. This is because BMAP is a generalization of a wide variety of traffic types and can effectively represent modern traffic than the Poisson model. The BMAP model suffers from drawbacks of complexity and fuels the matrix state space explosion of Markovian analytical models. It can easily render the analysis intractable. An approximate Markovian model to evaluate the BMAP based traffic models without too much error margins is developed and used in the teletraffic analysis.

Closely related works [14][15][16], have addressed in parts the various sections of this work and not as a complete unit. A TCP-aware CAC scheme to regulate the packet-level dynamics of TCP flows is proposed and analyzed in [14]. A TCP call is admitted to the system and TCP call adjustment is done by reducing the transmission rate to fit in the available bandwidth. The work does not employ modern traffic models and the wireless link is not modeled. The work in [15] models the TCP window and addresses the inefficiencies that arise due to the peculiar window evolution. It introduces TCP feedback into the CAC procedures in different non terrestrial wireless architectures. The work is simulation based and no analytical model is developed. A more recent work in [16] develops an analytical model to investigate the performance experienced by TCP sessions sharing a wireless channel by a fixed point approach; two models for different packet loss scenarios, imperfect error correction and buffer overflow. Different analytical queuing models are investigated and compared. The paper justifies the use of simple queuing models, M/M/1/K or Geo/Geo/1/K as compared to the BMAP/G/1/K. Though debatable, there is no call admission control employed and the wireless link is not explicitly modeled. The main contribution of this work is to provide a unified analytical model of a multiclass heterogeneous TCP based wireless network featuring; modern BMAP traffic and a CAC based on the wireless SIR and TCP feedback delay. Performance evaluation for the various TCP protocols over different wireless channel models is done.

This paper is organized as follows; Section 2 presents the system model; network model, CAC model and the wireless channel model. The analytical evaluation of SIR capacity and TCP based delay capacity is done in the section. The teletraffic queuing based analytical model is developed in Section 3 where the following is presented; non-homogeneous BMAP arrival process and the Markovian model. The performance measures for evaluating the developed models are presented in Section 4. The obtained results are presented and discussed in Section 5 and the conclusions drawn in Section 6.

2. SYSTEM MODEL

2.1 Network Model

The network consists of mobile stations of different classes; high priority traffic sensitive to both delay and packet loss, medium priority traffic that can tolerate some delay and packet loss violations and low priority best effort traffic. The traffic classes are grouped into the three groups depending on their desired BER (SIR) and delay thresholds. The CDMA based access network is employed. A call admitted in the network starts a TCP session for transmission of its traffic. TCP operates with independent sessions for each source and class. The different sessions affect each other by increasing interference on the wireless link and congestion in the core network.

A particular call i of class k arrives according to a $BMAP_k$ distribution. The call requires SIR threshold SIR_{Tk} and delay threshold d_{Tk}. If admitted, the call generates traffic that is scheduled for transmission on the wireless queues as shown in the complete network system model of Fig. 1.

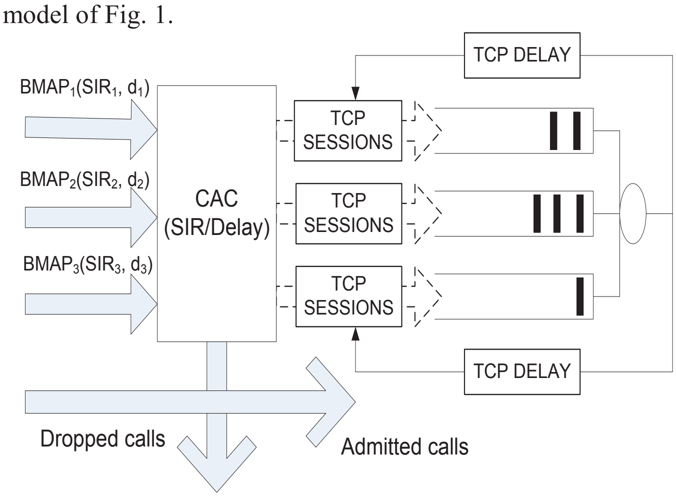

Fig. 1: System model

2.2 Call Admission Control Algorithm

The call admission control employed is as follows. An arriving call requests admission from the base station. To admit a new call i of class k, the guaranteed QoS (SIR and delay bounds), for the particular call should be provided and the QoS of the existing calls should not be severely affected by admitting the call. The SIR capacity is determined, if there is no capacity, the call is dropped; the delay capacity is determined, if available, the call is admitted. The analytical capacities represented by (2) and (16) are calculated in the sections that follow.

2.2.1 Capacity Based on interference

The received signal to noise ratio $(E_b/N_o)_{ik}$ of user i $\{i=1,2,..M_k\}$ of class k $\{k=1,..N\}$ is given by:

$$(E_b/N_o)_{ik} = \frac{G_{ik}h_{ik}p_{ik}}{(1+f)\sum_k \sum_{j\neq i} h_{jk}\alpha_{jk}p_{ik} + \eta_o W}, \quad (1)$$

where G_{ik} is the processing gain, h_{ik} the path gain, p_{ik} the transmitted power, α_{jk} the source activity factor, η_o the noise factor, W the bandwidth, f the ratio of the external to internal interference. The probability of accepting a call based on SIR under the minimum power requirement [17][18], p_A^S, is the given by [19]:

$$P_A^S = P(Z \leq \varphi), 0 < \varphi \leq 1. \quad (2)$$

where

$$Z = (C(t) + \Delta(t) + f(t)C(t)) \leq \varphi \quad (3)$$

and g_{ik} is the power index, Δ_{ik} the power constraint and $C(t)$ are respectively given by:

$$g_{ik} = \frac{(E_b/N_o)_{T,ik}}{((E_b/N_o)_{T,ik} + G_{ik})}, \quad (4)$$

$$\Delta_{ik} = \frac{\eta_o W}{\min_i (p_{ik}h_{ik}/g_{ik})}, \quad (5)$$

$$C(t) = \sum_k \sum_i \alpha_{ik}(t) g_{ik}, \quad (6)$$

The mean and variances of Z can easily be derived [19]. If (2) is satisfied, then there is capacity based on SIR.

2.2.2 CAC capacity based on TCP Protocol Delay

To determine TCP delay capacity, the TCP window evolution needs to be modeled. The behavior of various TCP protocols; window evolution, slow start, congestion avoidance, fast recovery and timeout has extensively been presented in literature [3][20][21]. Though there are many variants [11][12], the basic TCP protocols are used. TCP Window evolves in cycles which can be divided into rounds. After packet loss detection, a TCP cycle begins with either slow start or congestion avoidance and ends with the successful conclusion of fast recovery mechanism or on the basis of a timeout. A round starts with the transmission of w packets where w is the current size of the congestion window. To model the evolution, an analytical loss window model [22] is employed. Let W_i denote the maximum window size reached in the i th cycle.

$$W_i = \begin{cases} \min\{W_{rec}, W_m, W_d\} \\ W_d, if \ W_d < W_{rec} < W_m \end{cases} \quad (7)$$

where W_m is the maximum window size allowed on the wireless link due to the constraint of the sum of wireless link and the buffers in the system, W_{rec} is the maximum receiver buffer size that the receiver advertises at the beginning of TCP flow establishment. Finally, W_d is the window where a packet drops due to congestion or the wireless channel losses. The sequences of window sizes at which packets are dropped in successive cycles $\{W_i\}$ form a Markov chain with transition matrix $P(W_{i+1} = w/W_i = w_d)$. The steady state loss window probability $P(w_d)$, w_d = 1, 2… W_m can be found from the transition matrix. The transition probability matrix is characterized by [22]

$$P(W_{i+1} = w/W_i = w_d) = \begin{cases} P(D_{sw_d})P(S_s^s(w))P(L_s(w)) & \text{for } 1 \leq w \leq w_s - 1 \\ P(D_{sw_d})P(S_s^c(w))P(L_c(w)) & \text{for } w \geq w_s \\ P(D_{cw_d})P(S_c^c(w))P(L_c(w)) & \text{for } w \geq w_s \end{cases} \quad (8)$$

The terms of the equation are defined by the following events with w_s as the slow start threshold:

- The event that the next cycle starts in congestion avoidance given a packet loss at congestion window of w_d is D_{cw_d}. Its probability is $P(D_{cw_d})$.
- The event that the next cycle starts in slow start given a packet loss at congestion window of w_d is D_{sw_d}. Its probability is $P(D_{sw_d}) = 1 - P(D_{cw_d})$.
- The event that packets successfully reach the receiver to attain a window size w belonging in slow start given that it started from slow start is $S_s^s(w)$. Its probability is $P(S_s^s(w))$.
- The event that packets successfully reach the receiver to attain a window size w belonging in congestion avoidance given that it started from slow start is $S_s^c(w)$. Its probability is $P(S_s^c(w))$.
- The event that packets successfully reach the receiver to attain a window size w belonging in congestion avoidance given that it started from congestion avoidance is $S_c^c(w)$. Its probability is $P(S_c^c(w))$
- The event that a packet loss in a cycle results in a loss window w belonging in slow start is $L_s(w)$. Its probability is $P(L_s(w))$.
- The event that a packet loss in a cycle results in a loss window w belonging in congestion avoidance is $L_c(w)$. Its probability is $P(L_c(w))$.

The probabilities are calculated in the sections that follow.

a) Calculation of $S_s^s(w)$, $S_s^c(w)$ and $S_c^c(w)$

Let w_d, the drop window size in round d, be in slow start. To obtain a loss window size of w_d the following events must occur; $x_s = w_d - 1$ packets have to be successful and the w_d'th packet has to be dropped. The probability that enough packets successfully reach the receiver to attain a window size w belonging in slow start given that it started from slow start $P(S_s^s(w))$ is given by:

$$P(S_s^s(w)) = T_{w_d}(w_d - 1), \tag{9}$$

where $T_{w_d}(w_d - 1)$ is defined in (23) of the wireless section and noting that a packet is transmitted in a slot.

Let w_d, the drop window size in round d, be in congestion avoidance. To obtain a loss window size of w_d in congestion avoidance starting from congestion avoidance the following events must occur; x_{ca} packets have to succeed in the congestion avoidance

$$x_{ca} = w_s r_{ca} + \frac{(r_{ca} - 1) r_{ca}}{2} \tag{10}$$

where $r_{ca} = w_d - w_s$ is the number of successful congestion avoidance rounds and w_s is the slow start window threshold at which TCP enters congestion avoidance. The probability that enough packets successfully reach the receiver to attain a window size w belonging in congestion avoidance given that it started from congestion avoidance $P(S_c^c(w))$ is given by

$$P(S_c^c(w)) = T_{w_d}(x_{ca}), \tag{11}$$

where $T_{w_d}(x_{ca})$ is defined in (23).

Let w_d, the drop window size in round d, be in congestion avoidance. To obtain a loss window size of w_d in congestion avoidance starting from slow start the following events must occur; $x_t = x_s + x_{ca}$ packets must succeed. x_s is the number of the successful slow start packets while x_{ca} are the successful congestion avoidance packets. The probability that enough packets successfully reach the receiver to attain a window size w belonging in congestion avoidance given that it started from slow start $P(S_s^c(w))$ is given by:

$$P(S_s^c(w)) = T_{w_d}(x_t), \tag{12}$$

where $T_{w_d}(x_t)$ is defined in (23).

b) Calculation of D_{sw_d} and D_{cw_d}

The next cycle starts in slow start if the previous cycle ended in a timeout. Therefore, $P(D_{sw_d})$ is equal to the timeout probability P_{to} at the loss window of w_d. $P(D_{cw_d})$ is found from $P(D_{sw_d}) = 1 - P(D_{cw_d})$, since the next cycle either begins in slow start or congestion avoidance. For TCP Tahoe and Old Tahoe, $P(D_{sw_d}) = 1$, therefore $P(D_{cw_d}) = 0$. For the other TCP protocols the probabilities are computed from the timeout probability P_{to} discussed in the section below.

c) Timeout probability Calculation

When all the means of recovering from a packet by the TCP protocol fails, timeout occurs where the window reinitializes itself and restarts. Timeouts can be classified as direct or indirect. A direct timeout occurs if the number of duplicate Ack's that arrive at the sender is less than a certain threshold Ω (normally 3). For a small window $w_d \leq 3$, the number of duplicate Ack's will not arrive at the sender and therefore, $P_{to} = 1$. For a larger loss window, the number of packets successfully delivered in the loss window w_d, x_d, should be less than Ω. The probability of a direct timeout P_{to} is given by:

$$P_{to} = \sum_{x_d=1}^{\Omega} T_{w_d}(x_d) \tag{13}$$

An indirect timeout occurs if the TCP algorithm goes to fast retransmit and then the first recovery fails and a timeout occurs. This is given by the probability that the algorithm goes to fast retransmit and less packets go through for a first recovery to succeed and a timeout occurs. The probability of indirect timeout P_{to} is given by

$$P_{to} = \left(1 - \sum_{x_d=1}^{\Omega} T_{w_d}(x_d)\right) \sum_{x_d=1}^{\Omega} T_{w_d}(x_d) \quad . \tag{14}$$

d) Calculation of $L_s(w)$ and $L_c(w)$

The probability of a packet loss in a cycle resulting in a loss window w belonging in slow start, $P(L_s(w))$, and the probability of a packet loss in a cycle resulting in a loss window w belonging in congestion avoidance, $P(L_c(w))$, are identical and reduce to the probability of a packet loss after successful delivery of several packets. This is the probability of failure to deliver a packet in one slot. It is given by:

$$P(L_s(w)) = T_1(1) . \tag{15}$$

The values of $P(D_{sw_d})$ and $P(D_{cw_d})$ distinguish between various TCP protocols. The probabilities depend on the wireless channels packet loss probabilities.

e) The TCP based delay capacity

The TCP window evolves according to the network dynamics. Its growth is dependent on packet loss due to congestion or the wireless channel and the time of sensing the losses. The TCP window is therefore a good representative of the whole network delay due to the Round Trip Time (RTT) of sending a packet and receiving the acknowledgement. The TCP induced delay d_{Tk} can be translated into minimum packets to be transmitted per RTT n_{kRtt}. This directly translates to the TCP's window size w_{kRtt} noting that each window/round is transmitted in one RTT. A call will satisfy admission criteria if the number of packets transmitted per RTT is above n_{kRtt} (window above w_{kRtt}). Therefore the delay bound of a particular traffic can be estimated from the average window size. The delay based probability of admitting a user of class k is given by:

$$P_A^D = \sum_{w=w_{kRtt}}^{W_m} \pi_w \tag{16}$$

where π_w is the probability of a TCP window size of w, $0 \le w \le W_m$. For the TCP model the loss window characterizes the maximum throughput of a TCP session. Together with the round trip time, this determines the least delay that can be guaranteed for a particular TCP session.

2.3 *Wireless Model and packet error probability*

The wireless channel is modelled as a Finite State Markov Chain (FSMC). The FSMC Model represents the channel with multiple states with each state corresponding to a transmission mode [23]. It can be reduced to a Two State Markov Chain (TSMC) if the number of states reduces to two. In the FSMC the range of the received SNR is partitioned into a finite number of intervals. Let $(\Psi_0 = 0) < \Psi_0 < \Psi_1 \cdots < (\Psi_L = \infty)$ be the thresholds of the received SNR. Then the channel is in state $S_l, l = 0,1,\cdots,L-1$, if the received SNR is in the interval $[\Psi_L, \Psi_{L+1})$. An L-State FSMC Channel Model is depicted in Fig. 2.

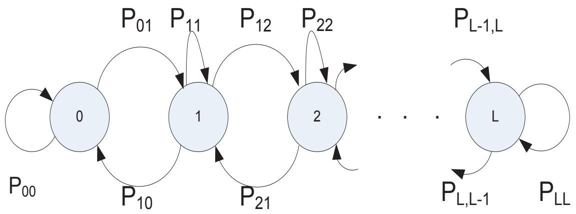

Fig. 2 2 L-State FSMC Channel Model

The steady state distribution of the FSMC is π, where $\pi \equiv [\pi_0, \pi_1, \cdots, \pi_{L-1}]$ and π_i is the probability that the underlying Markov chain is in state i given the chain is stable, and L is the number of states. In a typical multipath propagation environment, the received signal envelope has a Rayleigh distribution. With additive Gaussian noise, the received instantaneous SNR γ is distributed exponentially with Probability Density Function (PDF) [23]

$$p(\gamma) = \frac{1}{\gamma_0} \exp\left(-\frac{1}{\gamma_0}\right), \qquad \gamma \ge 0 \tag{17}$$

where γ_0 is the average SNR. In this case, the steady-state probabilities of the channel states are given by

$$\pi_l = \begin{cases} \int_{\Psi_l}^{\Psi_{l+1}} p(\gamma) d\gamma, & l = 0,1,\cdots,L-1 \\ \exp(-\Psi_l/\gamma_0) + \exp(-\Psi_{l+1}/\gamma_0) \end{cases} \tag{18}$$

while satisfying

$$\sum_{l=1}^{L} \pi_l = 1 \tag{19}$$

The packet error probability is a function of a given modulation scheme and a Forward Error Correction (FEC) code. Let p_{ei} be the channel bit error probability in the i-th state. The BER performance of uncoded BPSK scheme is given by,

$$p_{ei} = Q\left(\sqrt{\Gamma_i}\right) = Q\left(\sqrt{g_i G_i / Z_i}\right) \tag{20}$$

where Γ_i represents the average value of SIR of a BPSK in the i-th state and $Q(\cdot)$ is the Gaussian cumulative distribution function, g_i the power index and G_i the processing gain. Assuming r_i is the service bit rate and taking into account that the transmission time of each packet is specified to T_t. The number of bits per frame F_n is given by $F_n = T_t \cdot r_i$. The packet loss probability per state p_{li} is obtained as:

$$p_{li} = 1 - (1 - p_{ei})^{F_n} \tag{21}$$

The packet loss probability in a slot is given by:

$$P_l = \sum_{i=0}^{L} p_{li} \pi_i \tag{22}$$

Considering a transmission scheme where a packet is transmitted per slot and assuming there is no state transition within a slot, the probability of x packets transmitted successfully in n slots is:

$$T_n(x) = \binom{n}{x} (1 - P_l)^x (P_l)^{n-x}. \tag{23}$$

This is used in calculating the packet loss probabilities for the TCP window evolution.

3 TELETRAFFIC ANALYTICAL EVALUATIONS

3.1 *The BMAP Arrival Processes of the Model*

BMAP traffic models are used in the modeling of the network traffic as opposed to the simple Poisson models. The traffic arrival process is according to a non-homogeneous BMAP process due to the call admission control. The process is a 2-dimesional non homogeneous level dependent Markov process $\{N^p(t), J^p(t) : t \ge 0\}$ on the state space $\{(v, w : v \ge 0, 1 \le w \le m)\}$ for each $\{p > 0\}$. $N^p(t)$ is the number of arrivals and $J^p(t)$ is the phase of the process with an infinitesimal generator Q^p of the structure;

$$Q^p = \begin{bmatrix} D_0^0 & D_1^0 & D_2^0 & D_3^0 & \cdots \\ & D_0^1 & D_1^1 & D_2^1 & \cdots \\ & & D_0^2 & D_1^2 & \cdots \\ & & & D_0^3 & \cdots \\ & & & & \ddots \end{bmatrix}, \tag{24}$$

where the properties of the non-homogenous infinitesimal generator are dependent on the level p due to the effects of the CAC algorithm, a function of the number in the system. The stationary probability vector of the underlying Markov chain with generator D, denoted by π, satisfies;

$$\pi D = 0, \qquad \pi e = 1, \tag{25}$$

where e is a column vector of 1's. The fundamental arrival rate, λ gives the expected number of arrivals per unit time, is thus given by;

$$\lambda = \pi \sum_{q=1}^{\infty} q D_q e \tag{26}$$

The $BMAP_s$ can be considered as a superposition of several identical Markov Modulated Poisson Processes (MMPP's). The arrival rate matrix Λ and the infinitesimal generator matrix R, for the MMPP are given by:

$$\Lambda = \begin{bmatrix} \overline{\lambda}_0 & 0 \\ 0 & \overline{\lambda}_1 \end{bmatrix}, \tag{27}$$

$$R = \begin{bmatrix} -r_{01} & r_{01} \\ r_{10} & -r_{10} \end{bmatrix}, \tag{28}$$

where $\overline{\lambda}_i = \lambda_i P_A^p$, $i = 0,1$ and P_A^p is an admission probability dependent on the level and is a product of (2) and (16). The elements of the generator matrix D_0^p and D_1^p are functions of Λ and R [24] and incorporates CAC parameters.

3.2 The Markovian Analytical Model

The teletraffic analysis is done by a Markovian model. A full Markovian model suffers from an explosion on the matrix state space [24] and can't be used for complex networks. An approximate model has been used whose performance was tested and found to be close to the exact model [25]. The model used involves decomposition of the various classes into different priority queues at different stages with combined traffic. Consider three traffic classes with Class 1 as the highest priority and Class 3 the lowest. The arrival of traffic class k is governed by a $BMAP_k$, which is itself a combination of several processes from the sources that constitute the traffic class. An illustration of the approximate analytical model is as presented in Fig. 3.

For **Stage 1**, the traffic arrival into the high priority queue is a combination of the two high priority classes, $BMAP_{HP1} = BMAP_1 \oplus BMAP_2$. The traffic arrival into the low priority queue is simply the lowest priority traffic class, $BMAP_{LP1} = BMAP_3$. The low priority traffic does not differentiate between the two traffic types of higher priority. The total arrival into the system is $BMAP_{G1} = BMAP_1 \oplus BMAP_2 \oplus BMAP_3$. Let the total number of calls in the system be T_1, the total number of class one calls be X, the total number of class two calls be Y and the total number of class three calls be Z. The average number of calls at stage one can be represented by

$$E[T_1] = E[Z] + E[X+Y] \tag{29}$$

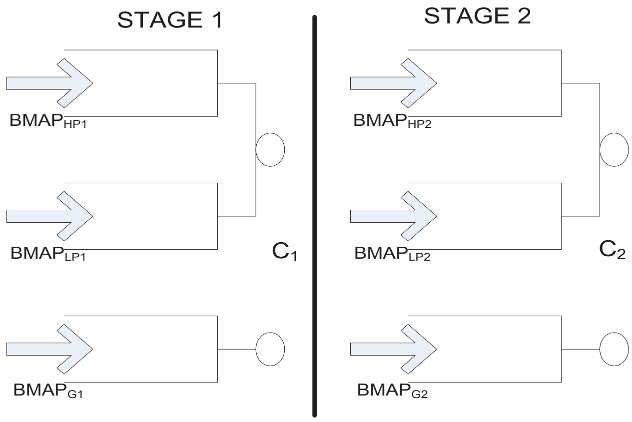

Fig. 3 Approximate Analytical Model

The expected number of calls in the high priority queue $E[X+Y]$ is calculated directly as a level dependent $BMAP/G/1$ queue with arrival $BMAP_{HP1}$. The expected number of calls in the whole system $E[T_1]$ is also calculated directly as a level dependent $BMAP/G/1$ with arrival $BMAP_{G1}$. The expected number of low priority calls $E[Z]$ can be approximated by (29).

For **Stage 2**, the traffic arrival into the high priority queue $BMAP_{HP2}$ is $BMAP_1$. The traffic arrival into the low priority queue $BMAP_{LP2}$ is simply $BMAP_2$. The total arrival into the system is $BMAP_{G2}$ and is given by $BMAP_{G2} = BMAP_1 \oplus BMAP_2$. Let the total number of Class 1 and Class 2 calls in the system be T_2. The following expression holds for the expected number of calls at Stage 2:

$$E[T_2] = E[X] + E[Y] \tag{30}$$

The expected number of calls in the high priority queue $E[X]$ is calculated directly as a level dependent $BMAP/G/1$ queue with arrival $BMAP_{HP2}$. The expected number of calls in the whole system $E[T_2] \approx E[X+Y]$ as calculated at Stage 1. The expected number of low priority calls $E[Y]$ can thus be deduced from (30). It should be noted that $E[T_2]$ can also be calculated as a level dependent $BMAP/G/1$ for accuracy comparisons. This is numerically tested and the results for the second stage normalized. The approximations are made possible by the fact that the lowest priority traffic does not differentiate between the higher priority ones. The higher priority traffic utilizes the complete network capacity. The model can be extended to several stages as an approximate model for evaluating more complex networks.

3.2 Analytical Evaluation of the Level Dependent $BMAP/G/1$ Queue

The $BMAP/G/1$ queuing system with level dependent arrivals is described by a stochastic process $\{X(t), J(t): t \geq 0\}$[24], where $X(t)$ and $J(t)$ are defined as the number of calls in the system referred to as the level and the phase of the arrival process at time t, respectively. Let τ_υ be the epoch of the υ^{th} departure from the queue, with $\tau_0 = 0$. Then $(X(\tau_\upsilon), J(\tau_\upsilon), \tau_{\upsilon+1} - \tau_\upsilon)$ is a semi-Markov process on the state space $\{(i,j): i \geq 0, 1 \leq j \leq m\}$. The state transition probability matrix of the semi-Markov is $\hat{P}_{BMAP/G/1}(x)$ where $\tau_{\upsilon+1} - \tau_\upsilon \leq x \geq 0$. The stationary vector of the Markov chain $P_{BMAP/G/1} = P_{BMAP/G/1}(\infty)$ and its steady state transitional probability matrix is given by

$$P_{BMAP/G/1} = \begin{bmatrix} F_0 & F_1 & F_2 & F_3 & \cdots \\ H_0^1 & H_1^1 & H_2^1 & H_3^1 & \cdots \\ \cdot & H_0^2 & H_1^2 & H_2^2 & \cdots \\ \cdot & \cdot & H_0^3 & H_1^3 & \cdots \\ \cdot & \cdot & \cdot & \ddots & \ddots \end{bmatrix}, \quad (31)$$

where, the elements of the matrices H_n^q and F_n are defined as follows: $[H_n^q]_{ij} =$ Pr{Given a departure at time 0, which left q customers in the system and the arrival process in phase i, the next departure occurs in a finite time x with the arrival process in phase j, and during that service there were n arrivals}, $[F_n]_{ij} =$ Pr{Given a departure at time 0, which left no customer in the system and the arrival process in phase i, the next departure occurs in a finite time x with the arrival process in phase j, leaving n customers in the system}. The matrices are

$$[H_n^q]_{ij} = P\left\{ \begin{matrix} X^q(\tau_{\upsilon+1}) = q+n, J^q(\tau_{\upsilon+1}) = j \mid X(\tau_\upsilon) \\ = q, J(\tau_\upsilon) = i, \tau_{\upsilon+1} - \tau_\upsilon \leq x \end{matrix} \right\}, \quad (32)$$

$$[F_n]_{ij} = P\left\{ \begin{matrix} X^0(\tau_{\upsilon+1}) = n, J^0(\tau_{\upsilon+1}) = j \mid X(\tau_\upsilon) = 0, \\ J(\tau_\upsilon) = i, \tau_{\upsilon+1} - \tau_\upsilon \leq x \end{matrix} \right\}. \quad (33)$$

The derivation of the state transition matrix and the steady state distribution from the state transitional matrix of (31) is done by the level dependent Markovian means [24][25][26].

4 PERFORMANCE MEASURES

The call blocking probability is used as the performance measure for the developed model. The steady state probability of the number in the system π_s is computed first. Let ψ_k be the blocking probability of call i of class k. The blocking probabilities are given by

$$\psi_k = \sum_{s \in S} \{1 - P^a(S,i)\} \pi_s, \quad (34)$$

where $P^a(S,i) = P_A^S \bullet P_A^D$, determined by (2) and (16) of the CAC algorithm.

5 DISCUSSION AND RESULTS

The performance of the analytical CAC algorithms is validated by simulations using a developed C++ discrete event simulator. Typical CDMA parameters are used in the simulation; processing gain of 128, chip rate 1.25MHz, AWGN of 10^{-18} with maximum power of 1watt is used. The call parameters were; call duration of 200 seconds, with on and off time of 0.5 seconds and 1 second respectively with an exponentially distributed service time. For the numerical results the BMAP considered is a superposition of several identical MMPPs alternating between two states. The values are chosen as follows: $\lambda_1 = 2\lambda_0$. The value of r_{01} and r_{10} used are 0.01 and 0.04, respectively. The traffic classes were selected as follows: Class 1 SNR/Delay in dB/seconds as 13/0.2, Class 2 as 8/0.3 and Class 3 as 5/0.5. For the TCP protocol; a fine timeout of 100 ms and a coarse timeout granularity of 500 ms were used. The first retransmit threshold of 3 was used and a fixed packet size of 500 bytes was used on the network. The round trip time of 200 ms was chosen and a maximum window of 20 packets was allowed.

The results of the teletraffic performance of the different TCP protocols, Tahoe and Reno, in terms of blocking probability are shown in Fig. 4. From the results, the following can be deduced; firstly, the dropping probabilities increase with an increase in the offered load as expected. Secondly, TCP Reno performs better that TCP Tahoe in terms of the call blocking probabilities as it achieves less blocking than TCP Tahoe. Finally the Poisson model achieves less blocking probabilities than the BMAP model. This can be disadvantageous in terms of overestimating the theoretical performance of the system resulting in poor network dimensioning. The analytical and simulation results tally well.

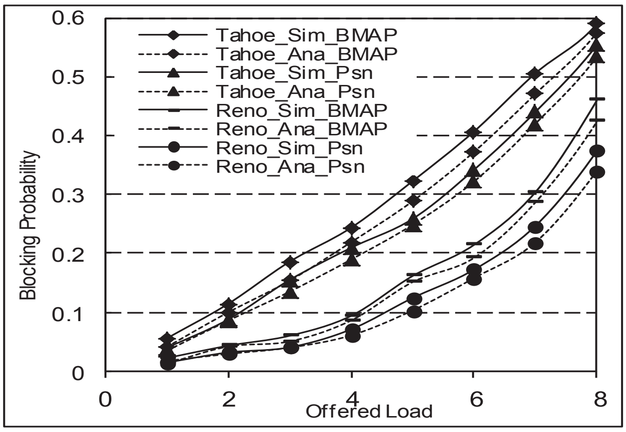

Fig. 4 Comparisons of Different Traffic Models

The analytical model was engaged to determine the behavior of TCP with different traffic classes. The results are as shown in Fig. 5. From the results, it is evident that the different traffic classes were differentiated in performance. Class 1 achieves better performance than Class 2 which performs better than Class 3. Their blocking probabilities increase with an increase in the offered load.

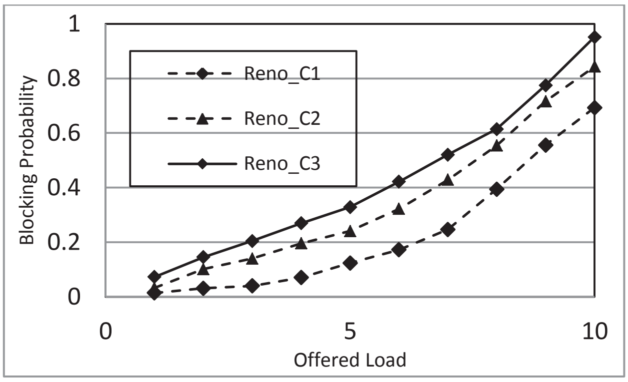

Fig. 5 Teletraffic Analysis of Different Traffic

The next test is the performance of TCP for the different wireless network protocols namely the Two State Markov Chain (TSMC) model and the Finite State Markov Chain (FSMC) model. The results are shown in Fig. 6. From the results we can deduce the following; as has been the case TCP Reno performs better than TCP Tahoe with any type of wireless model. The most important deduction is that for all the cases of TCP the TSMC wireless channel incurs less blocking than the FSMC model. This is feasible since the TSMC is an approximation of the FSMC. It could easily overestimate the losses on the network and lead to network dimensioning problems.

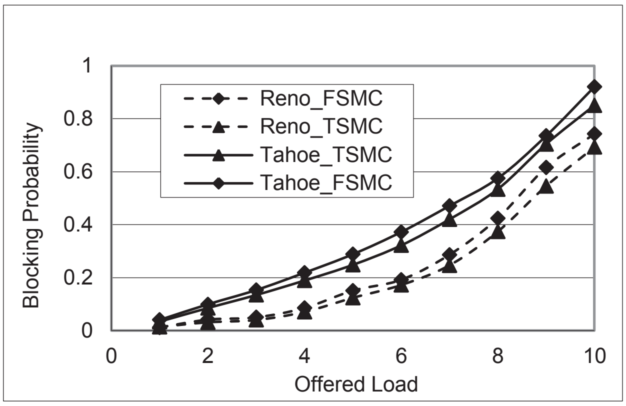

Fig. 6 Analysis for different wireless models.

The TCP window is the most significant aspect in its performance. In the next network test, the results of average window size of TCP Reno at various loads with or without CAC and for different values of timeouts of $TO\_1 = 200$ and $TO\_2 = 150$ are shown in Fig. 7. The following can be deduced: The window sizes decreases with an increase in the offered load. As the load increases, though regulated by the CAC scheme, the network traffic increases. However, the window size of TCP with CAC reduces very gradually as compared to the one without CAC. The CAC blocks incoming calls and maintains the QoS of the calls already in the system. A higher load increases the probability of packet error in the network. More timeouts occur and thus inhibit the growth of the window. TCP with a long timeout granularity performs better than that with a short time granularity.

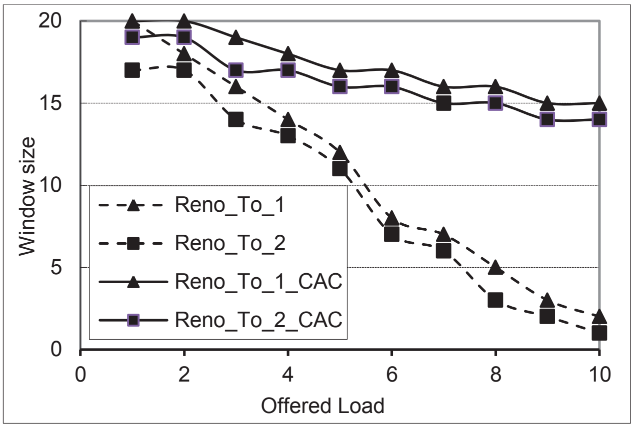

Fig. 7 TCP Reno's window for different timeout values

6 CONCLUSION

Telecommunication network protocols have found themselves being used on environments of which they were not meant to be used. Their effects need to be quantified on the new environments. CAC schemes and intelligent scheduling on the links has always been employed to alleviate the tremendous challenges of networks. These CAC schemes need to incorporate the effect of network protocols and features of future traffic. In this work, a CAC scheme featuring the following is developed, multiple traffic types, multiple admission parametersand TCP protocol issues with BMAP traffic. TCP delays are a good representative of the whole systems delay. To assess their effectiveness, the performance analysis of the solutions to alleviate network challenges (CAC) has been done. With the developed model, the effects of TCP protocol induced delays have been investigated for different TCP protocols. The results indicate that a CAC algorithm maintains the QoS for various TCP traffic in the network.

7 REFERENCES

[1] R. Prasad and T. Ojanpera, "An overview of CDMA evolution toward wideband CDMA," *IEEE Communication Surveys & Tutorials,* vol. 1, no. 1, pp. 2–29, 1998.

[2] Ka-Cheong Leung, V. O. K. Li, "Transmission control protocol (TCP) in wireless networks: issues, approaches, and challenges" *IEEE Communication Surveys & Tutorials,* vol. 8, no. 4. pp. 64-79, 2006.

[3] Y. Tian, K. Xu, and N. Ansari,"TCP in Wireless Environments: Problems and Solutions," *IEEE Communications Magazine,* vol. 43, no. 3, pp. 27–32, March 2005.

[4] Z. Liu and M. Zarki: "SIR-Based Call Admission Control for DS-CDMA Cellular Systems", *IEEE Journal on Selected Areas in Communications,* vol. 12, pp. 638-644, May 1994.

[5] W Jeon and D. Jeong: "Call Admission Control for CDMA Mobile Communications Systems Supporting Multimedia Services", *IEEE Transactions on Wireless Communications*, vol. 1, no. 4, October 2002.

[6] W. Yue and Y. Matsumoto: "Output and Delay Process Analysis for Slotted CDMA Wirelesss Communication Networks with Integrated Voice/Data Transmission," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 7, pp. 1245-1253, July 2000.

[7] C. Chang and K. "Medium Access Protocol Design for Delay-Guaranteed Multicode CDMA Multimedia Networks" *IEEE Transactions on Wireless Communications*, vol. 2, no. 6, November 2003.

[8] N. Tadayon, H. Wang, D. Kasilingam and L. Xing, "Analytical Modeling of Medium-Access Delay for Cooperative Wireless Networks Over Rayleigh Fading Channels" *IEEE Transactions on Vehicular Technology*, vol. 62, no. 1, January 2013.

[9] T. Spyropoulos, T. Turletti and K. Obraczka, "Routing in Delay-Tolerant Networks Comprising Heterogeneous Node Populations" *IEEE Transactions on Mobile Computing*, vol. 8, no. 8, August 2009.

[10] I. Cerutti, A. Fumagalli and P. Gupta, "Delay Models of Single-Source Single-Relay Cooperative ARQ Protocols in Slotted Radio Networks with Poisson Frame Arrivals" *IEEE/ACM Transactions on Networking*, vol. 16, vo. 2, April 2008.

[11]Qureshi, B. , Othman, M., Hamid, N.A.W. "Progress in various TCP variants" *2nd International Conference on Computer*, Control and Communication, 2009, pp 1-6.

[12] Balakrishan H. et. al. "A Comparison of mechanisms for improving TCP performance over wireless links," Proceedings of ACM SIGCOMM'96, August 1996.

[13] A. Klemm, C. Lindemann, M. Lohmann, "Traffic modeling of IP networks using the batch Markovian arrival process" *Performance Evaluation*, vol. 54, no. 22, pp. 149–173, 2003.

[14] X. Wang, D. Eun, and W. Wang, "A Dynamic TCP-Aware Call Admission Control Scheme for Generic Next Generation Packet-Switched Wireless Networks," *IEEE Transactions on Wireless Communications*, vol. 6, no. 8, August 2007.

[15] Georgios Theodoridis, Cesare Roseti, Niovi Pavlidou, and Michele Luglio, "TCP-Call Admission Control Interaction in Multiplatform Space Architectures," *EURASIP Journal on Wireless Communications and Networking*, 2007.

[16] Dmitri Moltchanov, "A study of TCP performance in wireless environment using fixed-point approximation," *Computer Networks*, vol. 56, pp. 1263–1285, 2012.

[17] L. Yun and D. Messerschmitt: "Power Control for variable QoS on a CDMA channel", *In Proceeding of IEEE MILCOM conference*, fort Monmouth, NJ, pp. 178-182, October 1994.

[18] A. Sampath, P.S. Kumar and J.M. Holtzman: "Power control and resource management for a multimedia CDMA wireless system", *Proceedings of PIMRC'95*, Toronto, Canada, pp. 21-25, September 1995.

[19] T. Walingo and F. Takawira, "Cross Layer Extended Parameter Call Admission Control or Future Networks" *SAIEE African Research Journal*, vol. 104, no.1, March 2013.

[20] J. B. Postel, "Transmission Control Protocol", RFC 793, September 1981.

[21] Stevens, W, 'TCP Slow Start, Congestion Avoidance, Fast Retransmit, and Fast Recovery Algorithms", RFC 2001, 1997.

[22] T. Walingo and F. Takawira, "TCP Over Wireless With Differentiated Services' *IEEE Transactions on Vehicular Technology*, vol. 53, no 6, pp. 1914-1926, November 2004.

[23] Hong Shen Wang and Nader Moayeri, "Finite-State Markov Channel—A Useful Model for Radio Communication Channels," *IEEE Transactions on Vehicular Technology*, vol. 44, no. 1, pp. 163-172, February 1995.

[24] D. M. Lucantoni, "New results on the single server queue with a batch Markovian arrival process", *Communication Statistics– Stochastic Models*, vol. 7, pp. 1-46, 1991.

[25] Tom Walingo, "Teletraffic analysis of Next Generation Networks" PhD dissertation, School of EECE, UKZN, Durban, South Africa, 2010.

[26] Hofmann J., "The BMAP/G/1 queue with Level-Dependent Arrivals - An Overview," *Telecommunication Systems*, vol. 16, pp. 347-360, 2001.