Unified Performance and Availability Model for Call Admission Control in Heterogeneous Wireless Networks

Ramesh Babu H. Siddamallaiah¹, Gowrishankar Subramanian², Piriyapatna S. Satyanarayana³

¹Department of Information Science and Engineering, Acharya Institute of Technology, Bangalore, India
²Department of Computer Science and Engineering, B.M.S. College of Engineering, Bangalore, India
³Department of Electronics and Communication Engineering, B.M.S. College of Engineering, Bangalore, India

Email: rameshbabu@acharya.ac.in, gowrishankar.cse@bmsce.ac.in, pssvittala.ece@bmsce.ac.in

Received January 31, 2010; revised February 27, 2010; accepted March 25, 2010

Abstract

The system capacity of wireless networks varies temporally. This may be due to the dynamic allocation of the channels and also the mobility of the users. The change in capacity will create greater impact on the system performance parameters. This variation of capacity particularly poses a greater challenge to the research community to ensure the Quality of Service (QoS) as it affects the call blocking probability which is one of the important QoS parameters. This paper proposes a performance model for call admission control and the availability model for a heterogeneous wireless network environment. The proposed model is able to handle three types of traffic considered for the study includes conversation traffic, interactive traffic and background traffic. The unified performance-availability model is developed using the Stochastic Area Networks (SAN). The performance of both analytical models and the SAN based performance-capacity models are verified by taking the call blocking probabilities for all the three types of traffics.

Keywords: Call Admission Control, Call Blocking Probability, Heterogeneous Wireless Networks, Stochastic Activity Networks, Quality of Service

1. Introduction

There is tremendous increase in the growth of wireless communication technologies which is evidently indicated by exponential increase in the wireless communication technologies usage. This popularity obviously demands the next stage beyond third-generation (3G) networks to include multiple wireless access technologies, all of which will coexist in a heterogeneous wireless access network environment [1,2] and use a common IP core to realize user-focused service delivery. The coexistence of Heterogeneous radio access technologies (RATs) is becoming mandatory to fulfill the needs of the growing users’ community of wireless technologies. This in turn will noticeably amplify the intensity in development of different high-speed multimedia services, such as video on demand, mobile gaming, Web browsing, video streaming, voice over IP and e-commerce etc. Seamless intersystem roaming across heterogeneous wireless access networks will be a major feature in the architecture of next generation wireless networks [3]. It is very well evident that no single RAT can provide ubiquitous coverage and continuously high quality service (QoS), the mobile users may have to roam among various radio access technologies to keep the network connectivity active to meet the applications/users requirements. With the increase in offered services and access networks, efficient user roaming and management of available radio resources becomes decisive in providing the network stability and QoS provisioning. The future users of mobile communication look for always best connected (ABC) anywhere and anytime in the Complementary access technologies like Wireless Local Area Networks (WLAN), Worldwide Inter operability for Microwave Access (WiMAX) and Universal Mobile Telecommunication Systems (UMTS) which may coexist with the satellite networks [4-6].

With this prevalent scenario of user mobility, ensuring the required radio resources to the users is a highly challenging task. In spite of failures in hardware, software, or combination of these two may still make the wireless networks to work and support the users. But there will be
a definite effect on the system capacity, i.e., the number of channels available and number of users that can be accommodated by the wireless networks may decrease and the throughput of the system will come down. This is an important aspect concerned to performance modelling. There is a strong need of performance modelling which should be able to take care of not only the pure performance but also the availability and reliability model of the system. But most of the performance models will consider this contention and generally over estimate the situation and will not consider the failure-recovery of the systems resources in turn the traditional availability models are not considering the performance metrics. There is a need for a composite model which will work on failure recovery model.

Researchers have achieved success in developing techniques for modelling the performance, availability and reliability of communication systems in a unified way. There are good number of approaches that have proved the need of unified performance model and availability models of any stochastic system under study. In this paper we have proposed composite Performance and Availability models and are evaluated using the stochastic petrinet.

The further sections of the paper are organized as follows. The Section 2 presents the performance model for call admission control. Section 3 focuses on the Availability model. The Section 4 represents SAN based performance-availability of the CAC system and Section 5 is devoted to discuss the simulation results and Section 6 concludes the paper.

2. Performance Model

In this paper we propose a novel analytic performance model admission control mechanism for reducing the call blocking probability there by increasing the resource utilization. This would achieve the objective of guaranteeing the user QoS requirements. The proposed model is able to handle three types of traffic considered for the study includes conversation traffic, interactive traffic and background traffic. All of this traffic represents different QoS service class of traffic with the following QoS parameters.

The conversational traffic is sensitive to transfer delay and jitter. It demands guaranteed bit rate and low bit error rate. The examples of the applications belonging to this category are video-conferencing and audio conferencing. The interactive traffic is a QoS class that is not sensitive to transfer delay and Jitter but demands low bit error rate. The applications of this QoS class do not need guaranteed bit rate for example web browsing, interactive chats and interactive games. The background traffic QoS class is not sensitive to transfer delay and jitter but needs low bit error rate from the network and these applications do not depend on guaranteed bit rate. The examples belonging to this group are e-mail, SMS applications. The assumption made for the design and development of analytical CAC model was type3 traffic would require three channels to be assigned in the system and type2 traffic demands two channels and type1 traffic needs one channel.

The proposed model is developed keeping in mind the WCDMA, WiFi, and WiMax. The CAC mechanism proposed is focused only on the system’s ability to accommodate newly arriving users in terms of the total channel capacity which is needed for all terminals after the inclusion of the new user. In the case when the channel load with the admission of a new call was precompiled (or computed online) to be higher than the capacity of the channel the new call is rejected, if not, the new call could be admitted. The decision of admitting or rejecting a new call in the network will be made only based on the capacity needed to accommodate the call.

We consider a heterogeneous network which comprises a set of RATs \( R_n \) with co-located cells in which radio resources are jointly managed. Cellular networks such as Wireless LAN and Wi-Max can have the same and fully overlapped coverage, which is technically feasible, and may also save installation cost. \( H \) is given as \( H \{ \text{RAT 1, RAT 2, RAT } k \} \) where \( k \) is the total number of RATs in the heterogeneous cellular network. The heterogeneous cellular network supports \( n \)-classes of calls, and each RAT in set \( H \) is optimized to support certain classes of calls.

The analytical model for call admission control mechanism in heterogeneous wireless networks is modelled using higher order Markov model and is as shown in Figure 1. In the proposed model it is assumed that, whenever a new user enters the network will originate the network request at the rate \( \lambda \), and is assumed to follow a poisson process. The service time of the different class of traffic and types of calls is \( \mu \). The mean service time of all types of users were assumed to follow negative exponential distribution with the mean rate \( 1/\mu \).

Since voice traffic is Erlang distributed, the condition that is considered for simulation is negative exponential distribution. The total number of virtual channels in the system are \( N \). When the numbers of available channels are below the specified threshold the system will drop the calls. The threshold limit is determined by three positive integers \( A_1, A_2 \) and \( A_3 \). These are called as utilization rates, where \( A \) is represented as

![Figure 1. Analytical model for CAC.](image-url)
Similarly,
\[ A = \frac{\lambda}{\mu} \]
are the utilisation rate of type 1 traffic, type 2 traffic and type 3 traffic respectively. In general the values of the utilisation rate in a steady state system will be with in 1.

When the available number of channels falls below the threshold \( A_1 \) the proposed system will accept only the voice calls and web browsing. When the available number of channels falls below the threshold \( A_2 \) the proposed system will accept only the voice calls. If the available number of channels falls below the threshold \( A_3 \) the proposed system will not accept any calls as it reaches the stage where there will be no channels available to allocate to the incoming calls and leads to system blocking. The \( P(0) \) is the probability that there are no allocated channels in the designated system. The parameters of analytical performance model are also called as Performance model parameters. The parameters are number of virtual channels \( N \), user arrival rate \( \lambda_i \) arrival rate of type 1 call \( (\lambda_1) \), arrival rate of type 2 call \( (\lambda_2) \) arrival rate of type 3 call \( (\lambda_3) \) and service time of the calls is taken as \( \mu_1 \), \( \mu_2 \) and \( \mu_3 \).

Assuming that the arrival time of all the types of traffic are equal i.e., \( \lambda_1 = \lambda_2 = \lambda_3 = \lambda \) and the service time for the types of traffic are equal i.e., \( \mu_1 = \mu_2 = \mu_3 = \mu \), the call blocking probability for type 1 traffic could be expressed as
\[ P_{i-1} = \frac{a}{3}(P_{i+1} + P_{i-2} + P_{i-3}) \]  \( (2) \)
where \( a = \lambda / \mu \) which should be generally less than one for the system stability. Similarly, the call blocking probability for type 2 traffic \( P_{i-2} \) is
\[ P_{i-2} = \frac{a}{3}(P_{i+2} + P_{i-3} + P_{i-4}) \]  \( (3) \)
And the call blocking probability for type 3 traffic \( P_{i-3} \) is represented as
\[ P_{i-3} = \frac{a}{3}(P_{i+3} + P_{i-4} + P_{i-5}) \]  \( (4) \)
The call blocking probability for the overall system traffic \( P_{ab} \) can be expressed as
\[ P_{ab} = \frac{a}{3}(P_{i} + P_{i+1} + P_{i-2}) \]  \( (5) \)

3. System Availability Model

The system availability model indicates the availability of the channels in the proposed system. When all \( N \) channels are failed then the system is unavailable for data transmission. Here the system is modelled as independent failure - repair model [7] which is also iterated in our previous work [8]. Each virtual channel is subjected to time and frequency selective fading and multipath fading then the virtual channel will be unavailable for data transmission. The individual channel is available for use by changing the mobile terminal position or by channel equalization technique [7]. The channel recovery model is developed as independent repair facility [9]. The failure rate of the channel is Poisson distribution with the rate \( \gamma \). The channel recovery is exponential repair time distribution of parameter distribution with the parameter \( r \). The \( N \) independent channel failure and recovery can be represented as a single dimension Markov Chain. The system availability can be modelled as Markov chain. The steady state probability \( P(i) \), where \( i \) is the number of available channel in the system and given by
\[ p(i) = \left( \frac{\tau}{\nu} \right)^i \frac{i!}{(N-j)!} \]  \( (6) \)

The steady state unavailability \( P(0) \) of the system is given by
\[ P(0) = \left[ \sum_{j=0}^{N-1} \frac{(\tau/\nu)^j}{(N-j)!} \right]^{-1} \]  \( (7) \)
The important observations from Performance-availability model is, increasing the number of channels will decrease the call blocking probability \( P_{ab} \). The system availability model is as shown in Figure 2.

4. Composite Performance and Availability Model

The performance-availability model is based on Stochastic Activity Networks (SAN). The SAN is a stochastic extension of Stochastic Petri Networks (SPN) in which the capacity to define temporary characteristics with statistical parameters has been added. The SAN exhibits the

Figure 2. System availability model.
innovative graphics which allows the researchers to represent a model with a high level of formal specification, expression of behaviour and dependency of the system in an uncomplicated and straightforward way. SPNs can be in general considered to constitute a method to model distributed, asynchronous concurrent systems, which have parallel characteristics. It is possible to study the performance and evolution of the system easily using Petri nets as Petri Nets combine graphic design and extensive mathematical theory to represent a system model.

A SAN performance model for CAC is represented in Figure 3 as channel usage model and the channel availability model which is also called as system capacity model is represented in Figure 4 for the aggregate traffic. The performance model of the proposed system is shown in Figure 5 and primitive components used in the proposed model are shown in Table 1. The activities $tr_{t_1}$, $tr_{t_2}$ and $tr_{t_3}$ represents the new user arrival/call arrival of traffic type1, traffic type2 and traffic type3 respectively which are timed activities and the firing distribution is a Poisson distribution. The new traffic arrivals have an inhibitory input from the input gate $ig_{nt_1}$, $ig_{nt_2}$ and $ig_{nt_3}$ when the number of virtual channels is less than $A_1$ channels, $A_2$ and $A_3$ respectively. The transition $tr_{sr_1}$, $tr_{sr_2}$ and $tr_{sr_3}$ represent user Service requests from traffic type1, type2 and type3 to system. Service requests are hyper-exponential distribution. The places AC and OC in the channel usage model indicate available channels and occupied channels/used channels.

The activity $tr_{t_1}$ represents the call arrival of traffic type1 on firing of transition $tr_{t_1}$, the output gate $og_{nt_1}$ shown in the Traffic type1 SAN performance model will function and removes single token from place AC and deposit a single token in place OC as shown in Figure 3. The transition $tr_{sr_1}$ represents user Service requests from traffic type1 to system. Service requests are hyper-exponential distribution. After the call is serviced the channel is released to the timed activity $tr_{sr_1}$ through input gate $ig_{sr_1}$. On firing $tr_{sr_1}$ the output gate $og_{sr_1}$ will draw a token from OC and deposit the token in place AC.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>SAN object</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AC</td>
<td>Place</td>
<td>Available Virtual Channels</td>
</tr>
<tr>
<td>OC</td>
<td>Place</td>
<td>Used / Consumed Virtual Channels</td>
</tr>
<tr>
<td>$tr_{t_1}$</td>
<td>Transitions</td>
<td>Type1 call arrival, type2 call arrival, type3 call arrival respectively</td>
</tr>
<tr>
<td>$tr_{t_2}$</td>
<td>Transitions</td>
<td>Service completion of type1 call/traffic, Service completion of type2 call/traffic, Service completion of type3 call/traffic</td>
</tr>
<tr>
<td>$tr_{t_3}$</td>
<td>Transitions</td>
<td>Service completion of type3 call/traffic</td>
</tr>
<tr>
<td>$ig_{nt_1}$</td>
<td>Input gate</td>
<td>Input predicate for type1 Traffic arrival, Input predicate for type2 Traffic arrival, Input predicate for type3 Traffic arrival</td>
</tr>
<tr>
<td>$ig_{nt_2}$</td>
<td>Input gate</td>
<td>Input predicate for type2 Traffic arrival, Input predicate for type3 Traffic arrival</td>
</tr>
<tr>
<td>$ig_{nt_3}$</td>
<td>Input gate</td>
<td>Input predicate for type3 Traffic arrival</td>
</tr>
<tr>
<td>$og_{nt_1}$</td>
<td>Output gate</td>
<td>Output function for type1 traffic arrival, Output function for type2 traffic arrival, Output function for type3 traffic arrival</td>
</tr>
<tr>
<td>$og_{nt_2}$</td>
<td>Output gate</td>
<td>Output function for type2 traffic arrival, Output function for type3 traffic arrival</td>
</tr>
<tr>
<td>$og_{nt_3}$</td>
<td>Output gate</td>
<td>Output function for type3 traffic arrival</td>
</tr>
<tr>
<td>$ig_{sr_1}$</td>
<td>Input gate</td>
<td>Input predicate for type1 Traffic service, Input predicate for type2 Traffic service, Input predicate for type3 Traffic service</td>
</tr>
<tr>
<td>$ig_{sr_2}$</td>
<td>Input gate</td>
<td>Input predicate for type2 Traffic service, Input predicate for type3 Traffic service</td>
</tr>
<tr>
<td>$ig_{sr_3}$</td>
<td>Input gate</td>
<td>Input predicate for type3 Traffic service</td>
</tr>
<tr>
<td>$og_{sr_1}$</td>
<td>Output gate</td>
<td>Output function for type1 traffic service, Output function for type2 traffic service, Output function for type3 traffic service</td>
</tr>
<tr>
<td>$og_{sr_2}$</td>
<td>Output gate</td>
<td>Output function for type2 traffic service</td>
</tr>
<tr>
<td>$og_{sr_3}$</td>
<td>Output gate</td>
<td>Output function for type3 traffic service</td>
</tr>
</tbody>
</table>
The performance model for type 2 call is represented in Figure 2. On firing of transition $tr_{t2}$, the output gate $og_{nt2}$ of the traffic type 2 SAN performance model will function and removes single token from place $AC$ and deposit a single token in place $OC$. The transition $tr_{sr2}$ represents user Service requests from traffic type 2 to system. Service requests are hyper-exponential distribution. After the call is serviced the channel is released to the timed activity $tr_{sr2}$ through input gate $ig_{sr2}$. On firing $tr_{sr2}$ the output gate $og_{sr2}$ will draw a token from $OC$ and deposit the token in place $AC$.

The activity $tr_{nt3}$ in Figure 3 represents the call arrival of traffic type 3 and on firing of transition $tr_{t3}$ the output gates $og_{nt3}$ in of the traffic type 3 SAN performance model will function and removes single token from place $AC$ and deposit a single token in place $OC$. The transition $tr_{sr3}$ represents user Service requests from traffic type 3 to system. Service requests are hyper-exponential distribution. After the call is serviced the channel is released to the timed activity $tr_{sr3}$ through input gate $ig_{sr3}$. On firing $tr_{sr3}$ the output gate $og_{sr3}$ will draw a token from $OC$ and deposit the token in place $AC$.

The transition $tr_{t1}$ represents an event of call arrival of type 1 traffic to the system. The transition new call/ user arrival of traffic type 1 has an inhibitory input from the input gate $ig_{nt1}$, when the total numbers of available channels are less than $A1$ the transition $tr_{t1}$ is disabled. The transition $tr_{t2}$ represents an event of call arrival of type 2 traffic to the system. The transition new user arrival of traffic type 2 has an inhibitory input from the input gate $ig_{nt2}$, when the total numbers of available channels are less than $A2$ the transition is $tr_{t2}$ is disabled. The $tr_{t3}$ is timed transition that represents the event of arrival of type 3 traffic to the system. The transition $tr_{nt3}$ is disabled when the available channel falls below $A3$.

The Figure 4 represents the fading model of the channel which gives the availability of the channel which represents the system capacity. The $AC$ and $UAC$ are the places in fading model and will represent channel availability and channel non-availability respectively in the proposed system. The transition $tr_{fad}$ represent the fading rate in wireless network and fading rate generally follows Weibull distribution. The transition $tr_{fad}$ is fired if and only if the tokens are available in the place $AC$ and this condition is implemented through input gate $ig_{fad}$. Transition $tr_{rel}$ is the channel recovery process and is assumed to be exponential distribution. When the Timed activity $tr_{rel}$ is fired the output gate $og_{rel}$ will draw a token from the place $UAC$ and send it to $AC$. This is nothing but when a channel fades then the channel will be in $UAC$ state and when channel comes out of fading state it will trigger the transition $tr_{rel}$ and place the token in $AC$. In other words the channel after coming out of fading state $UAC$ will enter the available channel state $AC$.

The SAN based performance-capacity model is represented in Figure 5, which is a composite architecture of the composite performance-availability model developed by combining the channel usage model and channel availability model/channel fading model.

5. Simulation Results and Discussion

In this section, we present the numerical results and compare the call blocking probabilities of the different types of traffic. A set of experiments were conducted varying the number of channels and the call blocking probability was compared for SAN performance- Availability model and the analytical models for all three types of traffic.

The parameters of analytic performance-capacity model are divided into performance model parameters and availability model parameters. Performance model parameters are number of channels ($N$), total user arrival rate of type 1 traffic ($\lambda_1$), total user arrival rate of type 2 traffic ($\lambda_2$), total user arrival rate of type 3 traffic ($\lambda_3$) and service time of the users ($\mu$).

The system capacity model parameters are channel failure rate ($\gamma$) and channel recovery rate ($\tau$). Parameters Values are $50(N), 7(\lambda_1), 3(\lambda_2), 1(\lambda_3), 2(\mu), 3(\gamma)$ and $2(\tau)$.

The Parameters of SAN Performance-capacity model is divided into channel usage model parameters and channel fading model parameters. The channel usage parameters are number of virtual channels represented by number of tokens in place $AC$, firing rate of type 1 traffic arrival transition ($tr_{t1}$), firing rate of type 2 traffic arrival transition ($tr_{t2}$), firing rate of type 3 traffic arrival transition ($tr_{t3}$) and channel usage service rates for type 1, type 2, type 3 traffic are $tr_{sr1}$, $tr_{sr2}$, $tr_{sr3}$ respectively. The channel usage rate is hyper-exponential distribution and classified into three types of usage such as low, medium and high with the service rate $\mu_1$, $\mu_2$ and $\mu_3$.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>SAN object</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$AC$</td>
<td>Place</td>
<td>Channel Availability</td>
</tr>
<tr>
<td>$UAC$</td>
<td>Place</td>
<td>Unavailability of Channel</td>
</tr>
<tr>
<td>$tr_{fad}$</td>
<td>Transition</td>
<td>Channel fading Rate</td>
</tr>
<tr>
<td>$tr_{rel}$</td>
<td>Transition</td>
<td>Channel recovery/release rate</td>
</tr>
<tr>
<td>$ig_{fad}$</td>
<td>Input gate</td>
<td>Input predicate for channel fading</td>
</tr>
<tr>
<td>$ig_{rel}$</td>
<td>Input Gate</td>
<td>Service completion</td>
</tr>
<tr>
<td>$og_{fad}$</td>
<td>Output gate</td>
<td>Output function for fading transition</td>
</tr>
<tr>
<td>$og_{rel}$</td>
<td>Output gate</td>
<td>Output function for recovery transition</td>
</tr>
</tbody>
</table>
μH respectively. The probability of low, medium and high usage are PL, PM and PH respectively. The channel fading model parameters are number of channels represented by number of tokens in place AC. The transition (tr_fad) represents the fading rate of the channel and recovered channel rate (tr_rel). Parameters values are 50(AC), 0.25(tr_t1), 0.5(tr_t2), 0.7(tr_t3), 1(μL), 2(μM), 3(μH), 0.5(PL), 0.3(PM), 0.2(PH), 0.34(tr_fad) and 0.5(tr_rel).

The first set of experiments is indicated by the simulation result shown in Figure 6. The call blocking probability for a system with N channels which supports three types of traffic is conducted. The experiment considers that, whenever a new user enters the network will originate the network request at the rate $\lambda_1$ for type1 traffic, $\lambda_2$ for type2 traffic and $\lambda_3$ for type3 traffic and is assumed to follow a Poisson process. The service time of the different types of traffic based calls is considered as $\mu_1$ for type1 traffic, $\mu_2$ for type2 traffic and $\mu_3$ for type3 traffic and is assumed to follow a Lognormal random Process.

For the first set of experiments we have considered the arrival rate of all the three types of traffic as $\lambda$ and service rate of all the three type of calls is same and is equal to $\mu$.

The second set of experiments conducted will present the numerical results and compare the call blocking probabilities of the different types of traffic obtained for performance-capacity model and the analytical model. The proposed a performance-availability model for call admission control mechanism in the heterogeneous RATs environment is analysed for the call blocking probability, by having variation in the number of channels. The graph obtained for the experiment setup conducted considering both the analytical model and SAN performance-availability model for the blocking probability of type1, type2, and type3 traffic is plotted. The horizontal axis shows the number of channels while the vertical axis shows the call blocking probability of all types of traffic.

The simulation results show that the call blocking probability of the different types of traffic will decrease with the increase in the number of channels in the system. The simulation results are shown in Figure 7. The simulation results show that the call blocking probability of the different types of traffic will decrease with the increase in the number of channels in the system.

The behaviour of analytical model and the performance model developed using stochastic activity network and simulated using the Mobius simulator behaves identically.

6. Conclusions

In this paper, the performance of analytical model for CAC system for next generation networks is compared and validated with the system performance-capacity model developed using SAN. The performance of both call admission control models in the heterogeneous RATs are studied pitching upon the call blocking probability by varying the number of channels. The increase in number of channels in the system decreases the call blocking probability of all traffic types. The results obtained for analytical model is in line with the performance model results where both the models behave in the similar fashion in the experiments conducted.

7. References


