Abstract– The study formulated a dynamic channel allocation model with one-level buffering in controlling congestion in Global System for Mobile Communication (GSM) network with a view to prevent call loss or degradation in quality of service of calls. The system model was implemented using object oriented programming. An algorithm was developed for accepting or rejecting of calls using ticketing scheduling. Various parameters were identified with a mathematical model to support the scheme using Markov chain technique. A simulation program using object oriented programming approach was developed to evaluate the performance of the scheme based on three performance metrics: Resource utilization, average queue length and blocking probabilities. A performance comparison with the existing scheme was tested based on the analysis of simulated model results obtained. A prototype of the scheme was tested based on the analysis of simulated model results obtained. A prototype of the scheme was tested based on the analysis of simulated model results obtained. A prototype of the scheme was tested based on the analysis of simulated model results obtained.

Index Terms– New Call, Handoff Call, Buffer, GSM, Allocation, Channel and Cell

I. INTRODUCTION

OVER the recent years, tremendous growth has been recorded in wireless networks such as global system for mobile communication (GSM) network [1]. As of today, GSM network through the use of mobile phones seem to evolve in popularity, size and diversity [8]. It is therefore of great importance to develop more advanced, as well as robust techniques to deliver manageable, stable and reliable services in order to meet the ever-growing demands for improved quality of service (QoS). Generally, as the number of users and the size of the network increases, users are likely to experience increase in response time and call loss with other performance degradation due to congestion problem.

According to [9], the network providers try to resolve and address these issues by keeping the network utilization low, which may not, seems to be a cost effective solution. Current trends towards rapid growth of wireless network communication clearly indicates that call admission control will handle congestion more efficiently to improve the network utilization, while providing a satisfactory level of service to the users. Some of the most tailored research area to handle congestion are call admission control (CAC) and network resource allocation. CAC determines the condition for accepting or rejecting a new call based on the availability of sufficient network resources to guarantee the QoS parameters without affecting the existing calls [15]. On the other hand, the network resource allocation decides how to accept incoming connection requests [17] to avoid congestion. The major call-level qualities of service parameters based on cellular concept are: new call blocking and handoff call blocking probabilities [4]. In this paper, a centralized dynamic channel allocation technique was employed with ticket scheduling to handle calls varying characteristics with a view to maximize the available resources.

In [3], authors integrated distributed channel allocation adaptive handoff for managing QoS-Sensitive cellular networks. However, all these researches are based on the number reservation method, where a predefined number of channels are reserved to serve higher priority calls but this will not perform maximally because of network changes in hot spot cells. In [16], authors developed a congestion control game with a linear pricing (me based on variations in the queuing delay experienced by the users. They used a network model based on fluid approximations, and established the existence of a unique equilibrium and the global stability of the equilibrium point for a general network topology. In [7], authors proposed an adaptive QoS management system in wireless multimedia networks. The proposed system was based on a service model designed for both connection-application-level QoS. Wireless multimedia applications are classified into different service classes in the service model by their application profiles. Based on the service model, adaptive resource allocation is performed for each service class by employing the appropriate CAC and RR schemes tailored to the QoS requirements of the service class. In [21], authors investigated the call admission control strategies for the wireless networks where the average channel holding times for new calls and handoff calls are significantly different, the traditional one-dimensional Markov chain model.
may not be suitable, two-dimensional Markov chain theory must be applied. They proposed a new approximation approach to reduce the computational complexity. However, it is important to bear in mind that the more complex and efficient a scheme is, the more computational power and time it requires to be applied. In [13], authors developed a non-preemptive prioritization scheme for access control in cellular networks. In [12], authors proposed the use of dynamic buffering to minimize congestion in mobile network. In [18], authors presented an actual call connection time characterization for wireless mobile networks under a general channel allocation scheme. In [22], authors proposed a new adaptive channel reservation scheme for handoff calls in wireless cellular networks. In [12], authors developed a distributed dynamic channel allocation algorithm for cellular network. In [14], authors demonstrated the use of multi-agent approach for dynamic channel allocation. In [6] a dynamic channel allocation mechanism for ubiquitous environment based on time constraint was developed. In [2], authors developed a centralized channel assignment and routing algorithms for multi-channel wireless mesh networks. In [11], authors discuss a non-cooperative multi-radio channel. In [3], [20], authors modeled a priority queuing systems with multi-class self-similar network traffic. However, these schemes only perform better in a light traffic condition because fixed threshold technique was employed. In a heavy traffic situation, the schemes lead to degradation in quality of service because heavy calculations in involved in choosing available channel thereby resulting to reduction in network efficiency.

The rest of the paper is structured as follows: Section II describes the system model. Section III shows numerical results and performance evaluation while section IV concludes the paper.

II. SYSTEM MODEL

The system considers two types of traffic: real-time call (e.g. voice) and non-real-time call (e.g., data). These calls are further classified as: new real-time call or real-time handoff call and new non-real-time-time call or non-real-time handoff call. The call processing entities of the system (e.g. the processing elements in the base station, the base station controller or the mobile switching center) are capable of identifying the call type at any moment. The available resources are the maximum number of channels in the cell and the buffer capacity that is used to queue real-time and non-real time handoff calls in case no channels are available. It is assumed that the number of channels in a cell is constant due to the fact that wireless resources are limited. Buffer capacity changes depending on the input traffic.

Assume users, representing requests of connection to the base station, arrive at the same time, the system first check where the requests come from (i.e. hot spot cell or cold spot cell). Hot spot calls are given priority over cold spot calls in channel allocation. If the requests are from both cells, hot spot cell request is allocated channel before cold spot request if the cold spot request is non-real time, otherwise ticket scheduling is used to serve the calls. Since real-time and non-real time handoff calls are sensitive to delay and loss respectively, they are given same priority. However, if the requests are from the same cell type, ticket-scheduling algorithm is employed and channel is allocated using centralized dynamic technique. Channels are released to the central pool if: user calls are completed, and when a mobile user crosses cell boundary. Blocked real-time and non-real-time new calls are lost, while blocked real-time and non-real-time handoff calls wait in the handoff buffer (queue) until there is an idle channel. The system model is as shown in Fig. 1.

The queuing scheme is briefly described as follows. No real-time and non-real-time new calls are queued when there is no channel. Real-time and non-real-time handoff calls are sent to the buffer, if there is no idle channel in the destination cell, and remains queued until a channel is free. In the case of high demand for real-time and non-real-time handoff, calls are denied queuing due to the limited size of the buffer. The queuing device has finite capacity. The buffer of real-time and non-real-time handoff calls lead to relatively low blocking probability.

A. System Algorithm

The basic steps of the proposed algorithm are presented in figure 2. When user calls arrive and calls are from different cells, calls are served between lines 3 and 10. However, if the calls are from same cell, ticket scheduling (each call is given a ticket and decision is made by choosing ticket at random and the call holding that ticket gets the channel) assists in serving the calls. In cases where no channel is free, since handoff calls are considered critical, real-time and non-real time handoff calls are buffered until channel to serve the call is available (lines 19-23). Handoff calls are only blocked when the buffer is full, while new calls are blocked when no channel is idle.

III. PERFORMANCE EVALUATION

A number of different scenarios were used in illustrating the behavior of the proposed scheme. Most especially, the extent of the blocking and dropping probabilities of both new call and handoff call respectively at various buffer capacities and at various transmission cycles. The total throughput of the aggregate traffic flow, both new call and handoff call, on the network was also tested. Moreover, various traffic load conditions, with different types of traffic for both calls were used in testing the buffer dimension for the scheme.

Fig. 1: System model of the proposed scheme
begin
1. While (n calls arrive for connection) do
2. if (there exist free channel) then
3. if (calls are from both cells) then
4. if (cold spot call is non-real-time) then
5. compare and serve hot spot calls before cold spot calls
6. else
7. assign ticket to each call
8. randomly pick any ticket and allocate channel to call with that ticket
9. endif
10. endif
11. elseif (calls are from same cell) then
12. if (same call type) then
13. repeat steps 7 and 8
14. else
15. allocate channel to real-time calls before non-real-time call
16. endif
17. endif
18. if (no channel is free) then
19. if (call is real-time or non-real-time handoff) then
20. is buffer full?
21. if No then
22. put call in buffer
23. repeat steps 7 and 8
24. else
25. block the call
26. endif
27. else
28. block the call
29. endif
30. endif
31. end while
end

Fig. 2: Channel allocation pseudo-code of the proposed scheme

A. Network Parameters

To study the impact of various network parameters on the proposed scheme, a basic network scenario was first considered as shown in Table 1. We then vary each network parameter in the basic scenario, one at a time, while keeping everything else fixed. The basic network scenario is a heterogeneous network with a single cell capacity of 5000 channels. We set the call service rate to 500 calls/s. Having service rate as 500 calls/s in the basic scenario allows us to evaluate the performance of the proposed scheme under heavy congestion, which are particularly problematic for congestion control using fixed threshold values. Next, we vary the maximum rate of generating handoff calls in the basic network scenario from 300 calls to 500 calls while keeping everything else fixed. This result in ranging the maximum rate of generating new calls from 250 calls to 400 calls which is the acceptable standard for most existing technologies such as GSM, UMTS, as well as higher capacity technologies, that are likely to become available in the future. Next, we vary the number of sources from 50 to 500. It is well known that the number of subscribers increase per day. Therefore, it is important to study the performance of the proposed scheme across a range of number of subscribers. We then study the impact of the heterogeneous network by picking transmission cycle from a small (10 cycles) as well as large (30 cycles) range of values. These ranges allow evaluation of the blocking and dropping probabilities of the calls. Finally, we evaluate the impact of the buffer size on the performance of the proposed scheme. This is useful because technology trends indicate that it is challenging to have large buffers in all cellular networks and therefore it is important that future schemes perform well when the buffer size is small. Considering the basic scenario, we vary the buffer size from 0.5Mb to 1.3Mb.

Table 1: Network parameters and their values used in the evaluation

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell Capacity</td>
<td>1000 calls/s</td>
</tr>
<tr>
<td>Call service rate</td>
<td>500 calls/s</td>
</tr>
<tr>
<td>Maximum rate of generating Handoff calls</td>
<td>300 – 500</td>
</tr>
<tr>
<td>Maximum rate of New calls</td>
<td>250 – 400</td>
</tr>
<tr>
<td>Number of sources</td>
<td>50 – 500</td>
</tr>
<tr>
<td>Maximum Buffer size</td>
<td>0.5 – 1.3Mb</td>
</tr>
<tr>
<td>Transmission cycle</td>
<td>0.5 – 30</td>
</tr>
</tbody>
</table>

B. Performance Analysis

A vast number of independent simulations with more complete study in which the benefits of the proposed model are evaluated to determine its performance with varying scenarios are presented. The explanation of these results and the analysis was based on the assumption that a single server is used in queuing the calls with ticket scheduling strategy to model the buffer.

The performance of the scheme is measured in heterogeneous environment. Although the network traffic composition is fairly stable in a long run, it fluctuates sometimes. In the simulation, we model the system where the portion of the non-real-time traffic in the entire traffic is changing from 30%, to 35%, and to 40%, while all the other traffic parameters are kept constant. The network stays for one-third of the simulation time, and each specific value of the traffic portion is as shown in Fig. 3(a)-(f). The results show that the scheme behaves well in heterogeneous environment because of its dynamic nature.

The throughput of each traffic type is shown in Fig. 4. In Fig. 4(a), the traffic throughput increases approximately with the arrival rate. However, when the traffic arrival rate is getting higher, or when the network is overloaded, it behaves a little differently as shown in Fig. 4(b) for non-real-time calls. It was also observed that the system throughput increases for both calls as arrival rate increases. This shows that the scheme successfully achieve a stable throughput even under heavy traffic situation.

Fig. 5 shows the channel idle probability which is defined as the probability that a released channel cannot be used by any queued calls. When the handoff call and new call load increases proportionally to the system traffic load, the channel idle probability decreases gradually as shown in the figure. This is an indication that the system resources are more wisely used, and also proves that the combined dynamic allocation and queuing policy works better than without queue. This result shows that the proposed scheme will at any time guarantee the quality of service to all the user calls without any delay.
Fig. 3(a): Call blocking probability against arrival rate at 30% of non-real-time traffic

Fig. 3(b): Call blocking probability against arrival rate at 30% of real-time traffic

Fig. 3(c): Call blocking probability against arrival rate at 35% of non-real-time traffic

Fig. 3(d): Call blocking probability against arrival rate at 35% of real-time traffic

Fig. 3(e): Call blocking probability against arrival rate at 40% of non-real-time traffic

Fig. 3(f): Call blocking probability against arrival rate at 40% of real-time traffic
Fig. 4: Throughput against arrival rate (a) Real-Time Call (b) Non-Real-Time Call

Fig. 5: Channel Idle Probability against Traffic Load

Fig. 6(a)-6(d) show the blocking probabilities of the call at various buffer sizes. The traffic arrival rate varies from 0.1 to 0.5 and its effect was tested on the network performance. This is to determine how to dimension the buffer at various network requirements (input traffic determines the size of the buffer). It can be seen that the blocking probability of calls decrease as the buffer size increases. It is an indication that the use of buffer makes fewer non-real-time calls to be admitted, but more to succeed (i.e. reduces call loss). This is an advantage over existing schemes. However, for time-sensitive traffic such as video, which cannot tolerate delay but insensitive to loss, it may be more desirable to use a smaller buffer size so as to keep the mean cell delay within the acceptable standard.

Fig. 6(a): Blocking probability against Buffer size at $\lambda = 0.1$

Fig. 6(b): Blocking probability against Buffer size at $\lambda = 0.2$

Fig. 6(c): Blocking probability against Buffer size at $\lambda = 0.3$
As shown in Fig. 7, the average waiting time in the queue for real-time calls keeps steady when that of non-real-time traffic changes. It shows that the real-time handoff calls’ performance is guaranteed in the queue. This is as a result of the priority given to the real-time handoff calls. There is no queue for real-time new calls and non-real-time new calls, so they do not need to wait. We also evaluate the delay performance of the proposed scheme with Poisson arrivals. There are 500 sources in the network and all of them are making calls with varying requirements while other parameters are fixed. The output result generated by the simulator is shown in table 3 and its graph in figure 8.

Figure 8 shows the average delay for different throughput. The minimum delay is achieved when the first few calls were received. We can see from the figure that when the system throughput is very small at around 1%, the average delay is close to the minimum. While with the increase of the throughput, the average delays increases dramatically. When the throughput in increased up to 35%, the network becomes saturated and the robustness of the proposed scheme enables the network to adapt to the fluctuation. In practice, this property shows the flexibility of the proposed scheme to traffic management. It is well known that the higher the number of subscribers the more the competition for available resources. Therefore, it is important to study the performance of the proposed scheme and its behavior on the network. We new vary the number of sources (using different calls) and study its impact.

Fig. 9(a) - Fig. 9(c) show that as we increase the number of sources, the scheme is able to maintain high utilization.
(≥90%), with negligible average queue length and near zero call loss rate. For new calls, utilization remains lower than that of handoff calls (due to higher average queue length than real-time handoff call. Loss rate for new calls, however, increases to only as high as 6%. This relatively low loss rate for non-real-time handoff call is a consequence of using one-level buffer enabled at the base station controller. When the number of sources is less than five, real-time handoff call achieves a higher average utilization than non-real-time handoff call. However, as the number of sources increased, real-time handoff call and non-real-time handoff call achieve similar average utilization. Note that real-time handoff call has a much lower average queue size compared to non-real-time handoff call even though they have similar loss rates.

1) Deployment Consideration: Wireless communication systems must not only provide high capacity and be able to coexist with other networks, they must also be possible to implement. This section considers implementation aspects such as computational complexity and installation sensitivity of the proposed scheme. The centralized algorithms cannot be too complex to perform since one of the constraints of fast allocation is time. All calculations must be done before a specified time limit and fast computers are expensive. We investigate if the computational complexity can be controlled with the ticketing scheduling and the use of one single server. To improve the capacity in a wireless network, a number of measures can be taken. One is to expand the infrastructure of the network. The last investigation of the paper looks into the expansion of the wireless network for the sake of capacity improvement.

For each cell to get information concerning the available channels centrally, it needs to be processed in some way. There is a cost involved in this process. Large channels take time and memory to handle. The complexity is often considered as one of the major drawbacks of centralization. Calculations generally require time. If the calculations are time critical, the controller has to be fast in allocating channels to the cells. The cell storing the channels grows with the size of users. For fast allocation, there is a time constraint connected to the cell length. The complexity of a selection strategy is related to the time that it takes to find a feasible channel. This time is computer dependent and a faster computer will of course find a channel in less time.

The impact of increasing the "geographical area covered by the network" and the "number of services provided by the network" on the overall performance is expected to be insignificant compared to the impact of increasing either the "number of users" or the "number of nodes and links".

In the simulation, the number of base stations is fixed while the numbers of users connected to a BTS are increased. It is assumed that the total load on the BSC is balanced and, therefore, equally divided among the available BTSs. To investigate the impact of increasing the number of users connected to one BTS, we increase the network load (calls/sec) proportionally. We also increase the call service rate of the network so as to maintain the utilization at a fixed level of 0.8. For a scalable network, the processing speed (to maintain the same utilization) should increase linearly with the network load.

![Fig. 10: Call Service rate against total load](image1)

We perform the simulation as follows: starting with the call service rate 500 calls/s and with a balanced network, we vary the service rate from 500 to 550, 600, 650, 700, 750, 800, 850, 900, 950. For each value, we adjust the network load (calls/sec) until the utilizations of all the BSC and BTSs is (approximately) equal to 0.8. We record this network load as the throughput corresponding to this given value of call service rate. The obtained results are listed in table 5. In figure 101 the call service rate required to maintain a network utilisation of 0.8 is plotted for different values of the total network load. Note that the call service rate increases with the load to maintain a fixed utilisation.

From figure 10 it was concluded that: for a traffic mix of 33.35% non-real-time call and 66.65% real-time call, after balancing the network, the throughputs increase linearly with the service rate required to maintain the same utilisation level. In practice, however, there are physical and cost limitations on the service rate which also limits the scalability of the network.

As shown in Fig. 11, it was observed that as the number of nodes increases, so also is the utilization increasing. The scheme is able to maintain this with all the call types based on the use of buffer for queuing calls that were supposed to be blocked. Although non-real-time-new call has the lowest utilization, yet it is still at an advantage.

Almost all these calls would have been blocked if the buffer is not available to queue such calls. Real-time handoff and non-real-time handoff calls enjoy the available channel much because they were given more priority. The result is in agreement with Fig. 7(a). This shows the consistency of the scheme.

![Fig. 11: Utilization against number of nodes](image2)
Moreover, figure 11 shows the call service rate against the total load. The figure shows that as the total traffic load increase, so is the service rate. This shows that no channel is likely to be idle at any point in time. It maximizes the available resources well that the network performance increases. However, the effect of this is that more computation is involved and the processing speed increases.

IV. CONCLUSION

This paper addresses congestion control in wireless network, specifically, global system for mobile communication network. A simplified but effective congestion control scheme was developed having a simulation model to study the appropriateness of the scheme in any real-world application. This model is implemented by developing a system prototype using object oriented programming approach. The simulation model thus implemented is used to study the behavior/performance of the scheme in some important wireless applications where they are believed to have better performance. The analysis of the results obtained shows that although the proposed scheme provides performance benefits over fixed threshold techniques, this is not true under all the conditions. Such conditions have been discussed. The major parameters on which the performance of the scheme depends are also presented. Furthermore, the results obtained from the simulation studies support the current trend of integrating buffering technique into the mainstream of the network rather than treating it as an exotic and specialized offshoot.

REFERENCES