

Performance Analysis of a Prioritized Multilevel Access Allocation Scheme for Data Calls in GSM Networks

Kuboye B.M.
Comp. Sci. Dept
F.U.T.A, Ondo State
Nigeria

Alese B. K.
Comp. Sci. Dept
F.U.T.A, Ondo State
Nigeria

Falaki S.O
Comp. Sc. Dept
F.U.T.A, Ondo State
Nigeria

Adewale O.S.
Comp. Sc. Dept
F.U.T.A, Ondo State
Nigeria

ABSTRACT

The demand of voice and data services on the Global System for Mobile Communications (GSM) is unprecedented in the annals of the history of telecommunication systems. As a result, grade of service (GoS) is no longer stable in the face of this high demand. This paper explores the use of prioritized multilevel Access Allocation (PMAA) scheme to stem the instability of data services when there is high demand. This paper classifies calls into Handover (HD), Real-Time (RT) and Non-Real-Time (NRT) based on the nature of their service and thereafter, priority was set among them. The HD has the highest priority followed by RT and NRT respectively. The channels in the cell were partitioned into three pools named as q_1 , q_2 , and q_3 . Data and voice calls have different characteristics, therefore, voice always have priority over data since it cannot tolerate delay in call processing. The q_1 is reserved for voice only, q_3 reserved for data only while q_3 is reserved for handoff calls for both voice and data.

General Terms

GSM Networks

Keywords

GSM, GPRS, EDGE, Realtime, Non-Realtime, Handoff

1. INTRODUCTION

Global System for Mobile Communications (GSM) was released due to the growing demand of mobile communication which the analog systems were able to cope with. GSM is more advanced than analog systems and handles more subscribers than the analog telephones. It offers high quality voice communication and low bandwidth (96 kb/sec) data connections for fax and short message service (SMS). As a result of the success of voice communication on GSM, data connection services like browsing, videos, multimedia, and e-mail demands on mobile telephone increased tremendously, thereby necessitating the need for a reliable and always available infrastructure to support these services [7]. In the quest to meet this demand, 2.5G GSM was introduced through General Packet Radio Service (GPRS) and High Speed Circuit-Switched Data (HSCSD) to give support for these demands. Thereafter, as the demand for data connection increases and 2.5G network could not support the demand, 3rd generation GSM was introduced. Enhanced Data Rates for GSM Evolution (EDGE) was created to support Internet connectivity and packet-switched services. The data rates of the 3G ranges from 144Kbps for fast moving mobiles to 2Mbps for slow moving mobile users [9]. There is always growing quest for data access on the mobile cellular radio which prompted the improvement on the 3G to 3.5G. The 4G will be the integration of the existing cellular networks and wireless LANs with added personalized mobile networks and broadband radio access networks to provide end-to-end IP connectivity.

The performance parameter in cellular radio or circuit switched depends on the Grade of Service (GOS) or Quality of Service (QoS) requirements. The main QoS parameter in circuit switched technology which GSM belongs is blocking [8]. Blocking occurs when a user wanted to use a channel for making a call and there is none available. The introduction of new data technology that doubles the speed of the previous one does not stop the congestion on the GSM network. The means of allocating the scarce channels is another issue on the overall performance of the GSM networks. Several papers have been presented on this issue and three principal categories have been identified and they are; fixed channels allocation (FCA), dynamic channel allocation (DCA) and hybrid Channel Allocation (HCA) [6]. In FCA, a set of channels is permanently allocated to each cell based on pre-estimated traffic intensity. When a user requests for a channel for communication, it searches the free channel in its own cell, if there is free channel, the communication is granted, otherwise, is blocked. DCA allows dynamic allocation of channels as new calls arrive in the system. The entire sets of free channels are kept in a central pool that will be accessible to all the cells. As soon as a call finishes using a channel, the channel is automatically returned to a central pool where it can be accessible to all the cells in the base station.

In HCA, few channels are allocated permanently to each cell and the remaining ones are in service on call basis. In other words, HCA is an integration of FCA and DCA schemes. In each of the channel allocation schemes, varying method have been proposed and investigated and more are still going to be proposed as a result of the escalating number of users always being experienced on the GSM and related networks. Due to the high demand of data on GSM networks, the channel allocation strategies are affected. New channel allocation strategies that will accommodate the data traffic on GSM network have to be implemented. Since the Radio Resource Manager (RRM) is the part of GSM Network that is charge of optimizing the usage of radio resources, new channel allocation strategies were evolved based on resource sharing algorithm implemented at RRM.

In literatures, three main static resource sharing schemes can be identified and they are Complete Sharing (CS), Complete Partition (CP) and Partial Partition (PP) [5][3][2]. The Complete Sharing (CS) allows all radio channels to be accessible for both data and voice traffics, while in the Complete Partitioning (CP), all channels are partitioned into two sets and each type of traffic is allowed to use only its dedicated set strictly. The third scheme known as Partial Partitioning (PP) is a hybrid scheme where CP and CS schemes are combined. In PP, a set of channels are shared between voice and data traffics, and the remaining timeslots are partitioned into two sets, each partition being reserved for strict usage of its dedicated type of traffic. This scheme has some advantages [2]:

- Reserving a set of time-slots for each type of traffic guarantee a minimum level of QoS for each type of traffic.
- provides a better efficiency than CP which is not suitable for maximizing radio utilization, especially in highly varying demand.

Despite all these schemes highlighted, congestion continues to be a major problem in GSM networks. This paper approaches the problem of congestion by classifying calls into different classes in order to reduced congestion and also to guarantee a good level of quality of service (QoS). Each traffic type has a degree of priority attached. A call with higher level of priority can preempt a lower priority level call when there is no free channel to make a call.

2. DATA TRAFFIC MODEL

In this paper, voice handoff traffic, handoff data traffic, real-time (RT) and non-real-time (NRT) data traffic are considered as different traffic types. The data traffic occupies pool 2 (q₂) and pool 3 (q₃) (see fig 1.0). Pool 2 will be shared between the handoff voice traffic and handoff data traffic. The complete sharing strategy has been proved to have a better system utilization among the other channel allocation strategies [4][2]. Therefore, the complete sharing strategy is adopted in each partition.

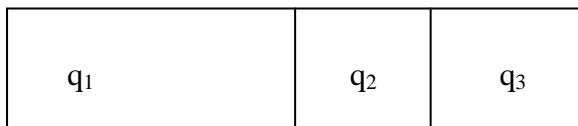


Fig 1.0: Complete sharing strategy

To guarantee the handoff voice performance, it is assumed that handoff voice traffic has preemptive priority over handoff data traffic in pool 2. Handoff voice calls is given a preemptive priority over handoff data calls because data can usually tolerate some degree of service degradation while voice is more delay sensitive [10]. When there are no channels available upon a handoff voice (hv) arrival, one of the handoff data packets in service is preempted. Likewise, to guarantee the performance of handoff data traffic (hd), when there are no channels available upon a handoff data (hd) arrival in pool 2 and pool 3, one of the NRT data packets in service in pool 3 will be preempted since hd has preemptive power over NRT and RT in pool 3. Also, RT has power to preempt non-real-time (NRT) data traffic in pool 3 if on arrival it finds no channel available for communication. When NRT calls arrives in pool 3 and no free channels available, then it will be blocked. Note that hd will be dropped if on arrival there is no RT and NRT in service and the maximum threshold to communicate with the former base station has elapsed, in the same way RT will be blocked if there is no NRT in service. Here, it is assumed that the preempted data packets are buffered in queue.

In this model, Three buffers are used to accommodate HD, RT and NRT data packets, respectively. One is for the preempted HD data packets caused by voice preemption. The other is used to accommodate preempted or blocked RT and the last is for blocked or preempted NRT data packets. Being buffered in queue, the preempted HD data packets have priority over RT data packets to obtain services. Also, the preempted packets or blocked RT have priority over NRT data packets to obtain services. All of these data calls are served in the order of priority and in case of equal priority,

first-come-first-served (FCFS) scheme is employed. The buffer size itself is regulated since there is a finite number of calls that can be accommodated and there is memorylessness in the system, therefore the size of the buffer is of the size of the capacity of the pool.

2.1 Derivation of Performance Measures

The performance analysis of radio resource allocation for multimedia traffic in this work is based on a 4-D Markov chain analysis model. The state (i, j, k, l) denote that there are i handoff voice calls, j handoff data packets, k RT data packets and l NRT data packets in the system.

$P_{i,j,k,l}$ denotes the probability of the system in state (i, j, k, l). The total number of channels in the system is C and the number of reserved channels for handoff traffic is q₂ while number of channels reserved for pool 3 traffic is q₃. The arrival of voice call request forms Poisson process with a rate of λ_v . The service time of voice calls is assumed to be exponentially distributed with a mean of $1/\mu_v$. The arrivals of HD, RT and NRT data packets are assumed to be Poisson processes with rates λ_{hd} , λ_{rt} and λ_{nrt} , respectively, and $\lambda_d = \lambda_{hd} + \lambda_{rt} + \lambda_{nrt}$ is the aggregate data arrival rate.

Assuming that λ_s^o is the mean arrival rate of new call for service type s of one user, it follows that λ_{hd}^o , λ_{rt}^o and λ_{nrt}^o are the mean arrival rates of new call for service type HD, RT and NRT of one user of those classes respectively. The handoff data arrival rate depends on the number of active users (x_s) in the system since if there is no user there cannot be any handoff. So it is given as $\lambda_{hd} = x_s \lambda_h^o$.

2.2 Service Holding Time

The service holding time of HD, RT and NRT data packets is exponentially distributed with a mean of $1/\mu_{hd}$, $1/\mu_{rt}$ and $1/\mu_{nrt}$ respectively. The aggregate mean service rate is $\frac{1}{\mu_d} = \frac{1}{\mu_{hd}} + \frac{1}{\mu_{rt}} + \frac{1}{\mu_{nrt}}$. It has been shown that it is reasonable

to assume that the service of most data are exponentially distributed, for some that are not exponentially distributed, they can be mounted on an exponentially distributed service rate that depends on state x through a bit rate b_s(x), of this service [1]. According to the model, the service rate of type s that are exponentially distributed was given as $\mu_s(x) = x_s \mu_s^o$ while the one that will be mounted on an exponentially distributed service due to low holding time was given as $\mu_s(x) = x_s \frac{b_s(x)}{m_{so}}$, m_{so} is the mean message size of type

s calls. For simplicity, in this model, the second assumption is adopted since all data that are not exponentially distributed can be mounted on exponentially distributed service. Therefore, service rate of HD, RT and NRT

$$\mu_{hd} = j_{hd} \frac{b_{hd}}{m_{hd}},$$

$$\mu_{rt} = k_{rt} \frac{b_{rt}}{m_{rt}} \text{ and } \mu_{nrt} = l_{nrt} \frac{b_{nrt}}{m_{nrt}} \text{ respectively.}$$

2.3 The Underlying Markov Chain

The proposed scheme here can be modeled with a 4-D Markov chain as shown in Fig. 2.0

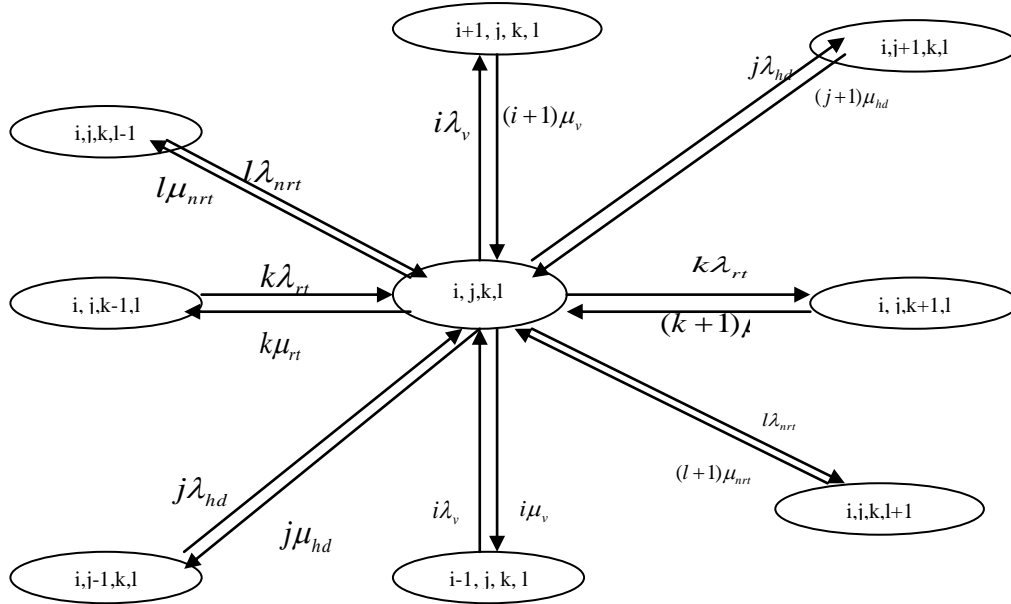


Fig 2.0: Four State Markov Chain for Data Analysis

The state space of the Markov chain is

$$P(i : j : k : l) = \lim_{t \rightarrow \infty} P((q_2(t) = i), (q_2 + q_3)(t) = j, q_3(t) = k, q_3(t) = l)$$

Let S be the set of feasible states,

$$S = \{(i, j, k, l) | 0 \leq i \leq q_2, 0 \leq j \leq q_2 + q_3, 0 \leq k \leq q_3, 0 \leq l \leq q_3\}$$

For all $i, j, k, l \in S$, the transition states of the Markov process are as follows:

(a) $S_v(i, j, k, l)$ is the transition rate from state (i, j, k, l) to $(i+1, j, k, l, m)$. A voice user is admitted into the system as long as the number of voice users in the q_2 is less than the number of available channels in q_2 or at least a data user is present in q_2 . Therefore, $S_v(i, j, k, l, m)$ can be written as

$$S_v^{i, j, k, l} = \begin{cases} \lambda_{sv}, & \text{if } 0 \leq i < q_2 \\ 0, & \text{otherwise} \end{cases}$$

(b) $H_d(i, j, k, l)$ is the transition rate from state (i, j, k, l) to $(i, j+1, k, l)$. A handoff data user is admitted into the system when the total number of S_v and H_d less than the number of available channels in q_2 or the total number of H_d q_3

Therefore, $H_d(i, j, k, l)$ can be written as

$$H_d^{i, j, k, l, m} = \begin{cases} \lambda_{hd}, & \text{if } i+j < q_2 \text{ or } j < q_3 \\ 0, & \text{otherwise} \end{cases}$$

(c) $R_T(i, j, k, l)$ is the transition rate from state (i, j, k, l) to $(i, j, k+1, l)$. A RT data user is admitted into the system when the total number of HD and RT are less than the number of available channels in q_3 . Therefore, $R_T(i, j, k, l)$ can be written as

$$R_T^{i, j, k, l} = \begin{cases} \lambda_{rt}, & \text{if } j+k < q_3 \\ 0, & \text{otherwise} \end{cases} \quad (d)$$

$N_{RT}(i, j, k, l)$ is the transition rate from state (i, j, k, l) to $(i, j, k, l+1)$. A NRT data user is admitted into the system when the total number of HD, RT and NRT are less than the number of available channels in q_3 . Therefore, $N_{RT}(i, j, k, l)$ can be written as

$$N_{RT}^{i, j, k, l} = \begin{cases} \lambda_{nrt}, & \text{if } j+k+l \leq q_3 \\ 0, & \text{otherwise} \end{cases}$$

The balance equation for the Markov process is expressed as

$$\begin{aligned} & \pi_{i-1, j, k, l} \cdot \{i\lambda_v(i, j, k, l) + i\mu_v(i, j, k, l) + j\lambda_{hd}(i, j, k, l) \\ & + j\mu_{hd}(i, j, k, l) + k\lambda_{rt}(i, j, k, l) + k\mu_{rt}(i, j, k, l) \\ & + l\lambda_{nrt}(i, j, k, l) + l\mu_{nrt}(i, j, k, l)\} \\ & = \pi_{i-1, j, k, l} \cdot i\lambda_v(i-1, j, k, l) + \pi_{i+1, j, k, l} \cdot (i+1)\mu_v(i+1, j, k, l) \\ & + \pi_{i, j-1, k, l} \cdot j\lambda_{hd}(i, j-1, k, l) + \pi_{i, j+1, k, l} \cdot (j+1)\mu_{hd}(i, j+1, k, l) \\ & + \pi_{i, j, k-1, l} \cdot k\lambda_{rt}(i, j, k-1, l) + \pi_{i, j, k+1, l} \cdot (k+1)\mu_{rt}(i, j, k+1, l) \\ & + \pi_{i, j, k, l-1} \cdot l\lambda_{nrt}(i, j, k, l-1) + \pi_{i, j, k, l+1} \cdot (l+1)\mu_{nrt}(i, j, k, l+1) \end{aligned}$$

By applying the constraint $\sum_k \pi_{i, j, k, l} = 1$ to the set of balance equations, the steady-state Probability $\pi_{i, j, k, l}$ can be

obtained to evaluate the performance metrics of the system. The steady state probability is obtained as follows:

$$\pi_{i,j,k,l} = \frac{1}{P(0,0,0,0)} \left(\frac{\rho_v^i}{i!} \cdot \frac{\rho_{hd}^j}{j!} \cdot \frac{\rho_n^k}{k!} \cdot \frac{\rho_{nrt}^l}{l!} \right) \dots\dots\dots 3.32$$

where $P(0, 0,0,0)$ is the steady-state probability of the system being idle. From the normalization equation, $\sum_s \pi_{i,j,k,l} = 1$,

$$P(0,0,0,0) = \sum_{a=0}^{C-q} \frac{(\rho_v + \rho_{hd} + \rho_n + \rho_{nrt})^a}{a!} \dots\dots\dots 3.33$$

where $a = i + j + k + l$ and $\rho = \frac{\lambda}{\mu}$. Hence the state probability is obtained as

$$\pi_{i,j,k,l} = \left(\frac{\rho_v^i}{i!} \cdot \frac{\rho_{hd}^j}{j!} \cdot \frac{\rho_n^k}{k!} \cdot \frac{\rho_{nrt}^l}{l!} \right) / \left(\sum_{a=0}^{C-q} (\rho_v + \rho_{hd} + \rho_n + \rho_{nrt})^a / a! \right)$$

Using the state probability, the following performance metrics are derived:

i Handoff Data Packet blocking: A handoff data call is blocked when upon arrival, q_2 is filled up with S_v and H_d and q_3 is filled up with H_d . The packet blocking probability of HD data traffic, P_{hdb} , is expressed as

$$P_{hdb} = \sum_{i=0}^{q_2} \sum_{j=0}^{q_2-i+q_3} \sum_{k=0}^{q_3} \sum_{l=0}^{q_3} \pi_{i,j,k,l} \dots\dots\dots 3.35$$

ii A Real-time data call is blocked when upon arrival; q_3 is filled up with H_d and R_T . The packet loss probability of RT data traffic, P_{rtb} , is expressed as

$$P_{rtb} = \sum_{i=0}^{q_2} \sum_{j=0}^{q_2+q_3} \sum_{k=0}^{q_3-j} \sum_{l=0}^{q_3} \pi_{i,j,k,l} \dots\dots\dots 3.36$$

iii The packet blocking probability of NRT traffic, P_{nrtb} , was expressed as

$$P_{nrtb} = \sum_{i=0}^{q_2} \sum_{j=0}^{q_2+q_3} \sum_{k=0}^{q_3} \sum_{l=0}^{q_3-(j+k)} \pi_{i,j,k,l} \dots\dots\dots 3.37$$

3. ANALYTICAL SIMULATION AND PERFORMANCE RESULTS FOR DATA COMMUNICATION MODEL

In this paper, the parameters used in the analytical simulation are listed in the table 1.0.

Table 1.0: Data Analytical Simulation Parameters

Total channel in the cell	22
Total channel for q_2	6
Total channel for q_3	5
Service rate $1/\mu_d$	600s, 300s
Arrival rate for voice λ_v	1.67 – 5.0 call/s \approx 100 – 250 calls
Arrival rate for handoff data	2.5 – 5.83 call/s \approx 150-250

λ_h	calls
Arrival rate for Realtime data λ_{rt}	3.33 – 6.67 call/s \approx 200-300 calls
Arrival rate for Non-realtime data λ_{nrt}	5 – 8.33 call/s \approx 300 – 500 calls
Offered load in erlangs for voice ρ_s	5.0 – 15, 3.3 – 10
Offered load in erlangs for handoff ρ_h	7.5 - 17.5, 5 – 11.7
Offered load in erlangs for RT ρ_{rt}	10 – 20, 6.7 – 13.3
Offered load in erlangs for NRT ρ_{nrt}	15 – 25, 2.5 – 9.2
P_v	Blocking Probability of voice call
P_{hb}	Dropping Probability of handoff data (HD) call
P_{rtb}	Blocking Probability of realtime (RT) data call
P_{nrtb}	Blocking Probability of non realtime (NRT) data call

Fig. 3.0 show the blocking probabilities of Data traffic for all data types in the presence of voice call when simulation was done with service time of 600s and 300s. The blocking probability under 600s was higher than the one done with 300s.

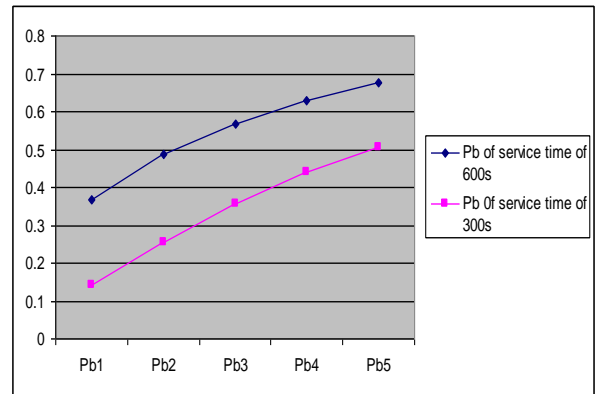


Fig 3.0: Graph representation of Data traffic probabilities in the presence of Voice call

Fig. 4.0 show the blocking probabilities of data traffic when there is no voice traffic in q_2 that shows there are more channels available for the HD users in q_2 , this account for the further reduction of blocking probabilities of data traffic in the cell.

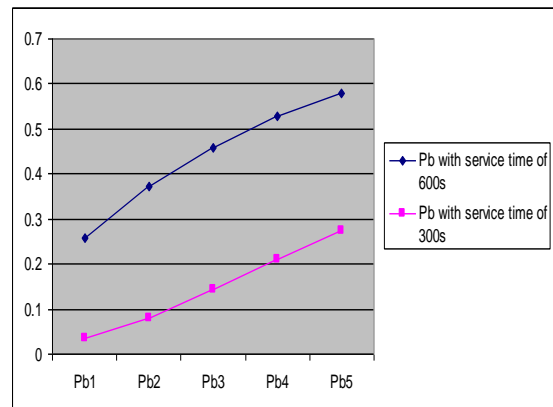


Fig 4.0: Data traffic without voice

Fig. 5.0 show a situation where we have NRT and RT in q3 only. Since there is no HD traffic to compete with the RT and NRT in q3, there is further reduction in probabilities when compared with when HD was present in Fig 6.0

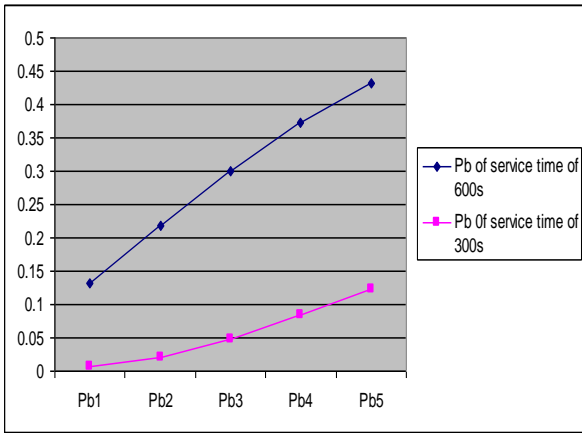


Fig 5.0: NRT and RT

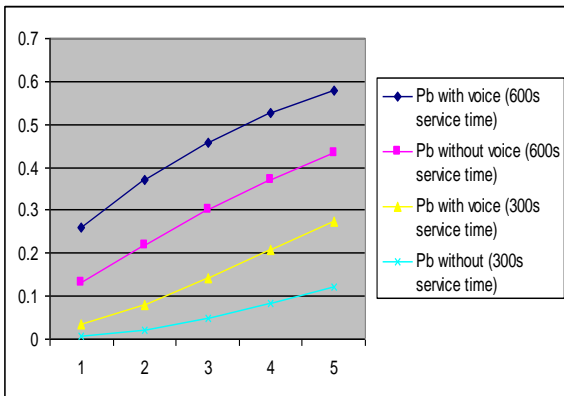


Fig 6.0: NRT and RT for Two Service Rates

Fig 7.0 show a situation when HD occupying q2 and RT occupy q3. The least blocking probabilities are achieved when te was applied to the HD.

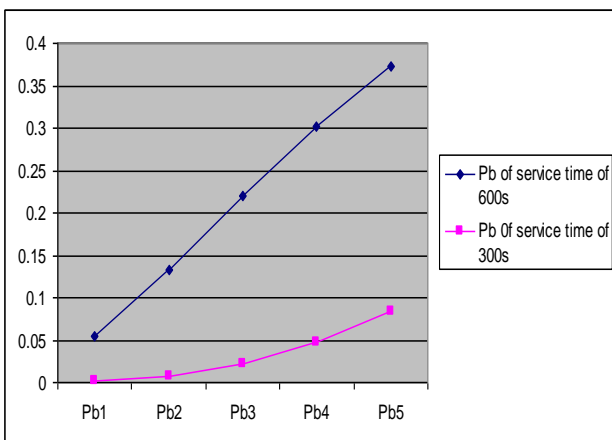


Fig 7.0: HD and RT

Fig 8.0 show a situation when HD occupying q2 and NRT occupy q3. The least blocking probabilities are achieved when te was applied to the HD.

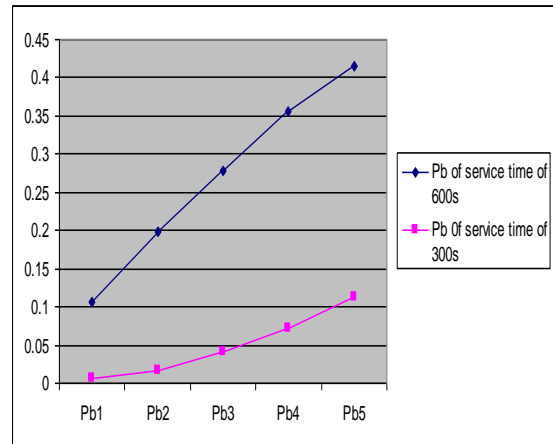


Fig 8.0: HD and NRT

4. CONCLUSION

In this paper, an analytical model of a prioritized multilevel access allocation (PMAA) scheme for data calls in cellular networks is developed. The PMAA scheme is modeled using a four-dimensional markov chain and numerical analysis is presented to estimate blocking probabilities of non-real-time and real-time data calls, dropping probabilities of handoff data calls. The scheme is tested under different time ranges in other to find a better scenerio where congestion will be minimized that still return a profit gain to the operator. PMAA aims is to increase the users satisfaction by reducing the blocking and dropping probabilities of calls as well as increase the operators' satisfaction by increasing the utilization of the bandwidth available. Hence, both users' and service providers' satisfactions are balanced in the PMAA scheme.

5. REFERENCES

- [1] Begain K., Ermel M., Muller T., Schueller J. and Schweigel M. (2000): "Analytical Call Level Model of GSM/GPRS Network", in Proceedings of Spect's00, SCS Symposium on Performance Evaluation of Computer and Telecommunication Systems, Vancouver, Canada.
- [2] Budura G. , Baling C. , Budura A. , and Marza E. , (2009), "Traffic Models and Associated Parameters in GSM/(E)GPRS Networks, WSEAS Transactions on Communications", Issue 8, Volume 8
- [3] Dahmouni H., MORIN B. and VATON S., (2005) "Performance Modeling of GSM/GPRS Cells with Different Radio Resource Allocation Strategies", Methods and Tools for Traffic Engineering, France Telecom R & D, Belfort, France.
- [4] Ermel M., Begain K., Muller T., Schuler J. , and Schweigel M., (2002). "Performance of GSM Networks wit General Packet Radio Services, Performance Evaluation, Vol. 48, Pg. 285-310, May 2002.
- [5] Fang and Ghosal (2003). "Performance Modeling and QoS Evaluation of MAC/RLC Layer in GSM/GPRS Networks". Proceeding of IEEE ICC, Alaska, USA.
- [6] Gupta and Sachan, (2007). "Distributed Dynamic Channel Allocation Algorithm for Cellular Mobile Network", Journal of Theoretical and Applied Information Technology.

- [7] Hartel, Levine R. Livingtone G. (1999). “GSM Superphones”, McGraw-Hill.
- [8] Janeviski T., (2003). Traffic Analysis and Design of Wireless IP Networks 2003 Artech House, INC ISBN 1-58053-331-0.
- [9] Prasad and Muñoz, (2003). “WLAN and WPAN towards 4g Wireless”, Norwood, Ma: Artech House
- [10] Tang L., Shensheng T. and Wei L. (2004). “A Channel Allocation Model with Preemptive Priority for Integrated Voice/Data Mobile Networks”, Proceedings of the First International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks (Qshine’04) 0-7695-2233-5/04 IEEE