A Dynamical Hybrid CAC Scheme and Its Performance Analysis for Mobile Cellular Network with Multi-Service

Jiping Li[1, 2], Shixun Wu[1] and Shouyin Liu[1]
[1] Department of Electronic and Information Engineering, Central China Normal University
Wuhan, 430079 China
[2] School of Computer and Information Science, Hubei Engineering University
Xiaogan, 432000 China
[e-mail: oucljp@yahoo.com.cn]
*Corresponding author: Shouyin Liu

Received February 2, 2012; revised April 29, 2012; accepted June 13, 2012; published June 25, 2012

Abstract

Call admission control (CAC) plays an important role in mobile cellular network to guarantee the quality of service (QoS). In this paper, a dynamic hybrid CAC scheme with integrated cutoff priority and handoff queue for mobile cellular network is proposed and some performance metrics are derived. The unique characteristic of the proposed CAC scheme is that it can support any number of service types and that the cutoff thresholds for handoff calls are dynamically adjusted according to the number of service types and service priority index. Moreover, timeouts of handoff calls in queues are also considered in our scheme. By modeling the proposed CAC scheme with a one-dimensional Markov chain (1DMC), some performance metrics are derived, which include new call blocking probability (P_{nb}) , forced termination probability (PF), average queue length, average waiting time in queue, offered traffic utilization, wireless channel utilization and system performance which is defined as the ratio of channel utilization to Grade of Service (GoS) cost function. In order to validate the correctness of the derived analytical performance metrics, simulation is performed. It is shown that simulation results match closely with the derived analytic results in terms of P_{nb} and PF. And then, to show the advantage of 1DMC modeling for the performance analysis of our proposed CAC scheme, the computing complexity of multi-dimensional Markov chain (MDMC) modeling in performance analysis is analyzed in detail. It is indicated that state-space cardinality, which reflects the computing complexity of MDMC, increases exponentially with the number of service types and total channels in a cell. However, the state-space cardinality of our 1DMC

http://dx.doi.org/10.3837/tiis.2012.06.003

model for performance analysis is unrelated to the number of service types and is determined by total number of channels and queue capacity of the highest priority service in a cell. At last, the performance comparison between our CAC scheme and Mahmoud ASH's scheme is carried out. The results show that our CAC scheme performs well to some extend.

Keywords: CAC, Markov chain, Mobile cellular network, New call blocking probability, PF, Queue scheme

1. Introduction

In recent years, mobile cellular networks have made great progress in technological aspects. Users' demand for service changes from single voice service to multimedia service which consists of voice, video, image, data or combinations of them. This requires the mobile cellular network provide reliable quality of service (QoS) in terms of new call blocking probability (P_{nb}) and handoff call dropping probability (P_{hd}). However, due to limited radio spectrum, users' mobility, decreasing cell size and ever increasing service demand with different QoS requirements, call admission control (CAC) for multi-service will become much more challenging in future mobile cellular network. During the past decade, intensive works have been done on this issue.

In [1][2][3][4][5][6], CAC schemes and related performance analysis for voice and data services are proposed. In [1], Leong C.W. et al. propose and analyze a priority based resource sharing scheme for voice/data integrate cellular networks. The proposed CAC scheme is able to simultaneously provide satisfactory QoS to both voice and data users and maintain relatively high resource utilization in a dynamic traffic load environment. In [2], Far L. M. et al. propose a CAC scheme which can keep P_{nb} and P_{hd} under a desired level under different workload. In [3], Leong C. W. et al. propose a CAC policy using limited fraction guard channel policy to reserve resources exclusively for potential handoff calls to maintain QoS for admitted users. To reduce the computing complexity, a 2-D model is proposed to approximate the 3-D model for the system with voice/data traffics and related performance analysis is carried out. In [4], Yieh-Ran Haung et al. propose an analytical model to investigate the performance in terms of voice call blocking probability, data loss probability and mean data delay for mobile network with integrated voice/data and finite data buffer. In [5], Li Yin et al. extend the traditional guard channel scheme by using two thresholds, one is used to reserve channels for voice handoff calls and the other to block data traffic into network in order to ensure voice performance in terms of P_{hd} and P_{nb}. In [6], Tat-Chung Chaua et al. propose a dual thresholds CAC scheme used for narrowband voice handoff calls and wideband data handoff calls. By using complete sharing approach, channel efficiency can be maximized meanwhile corresponding QoS can be satisfied. However, the dropping of handoff call in queues is not considered in [6], [7] and [8]. In [9], Wei Li, et al. propose two channel borrowing CAC schemes and give their performance analyses by using four-dimensional Markov chain and six-dimensional Markov chain respectively. However, the performance analyses are rather complicated though they are fairly accurate. Due to the high dimension of Markov chain modeling, the performance analysis of the CAC scheme for more than three types of traffics are rather complex, so only two types of traffic are discussed as an example in [9]. Considering the call arrival rates depend on the time of a day, Dusit Niyato et al. in [10] present a fixed channel allocation scheme and analyze the performance of the cellular wireless networks in terms of P_{nb} and handoff call blocking probability by modeling the variation of arrival rate of a day as Markov modulated Poisson process. However, due to high computational complexity of MDMC modeling, only two types of service are discussed.

In [11][12][13][14][15], CAC schemes for more than two types of service are proposed and corresponding performances are analyzed. In [11], Nidal Nasser et al. propose an adaptive framework for supporting multi-service with different QoS requirements in cellular networks. In [12], Fei Hu et al. propose a multi-class CAC mechanism based on a dynamical reservation pool for handoff request. In [13], complete sharing (CS) and complete partition (CP) scheme combined with guard channels are used for resource allocation in CAC for multi-service. The scheme extends the CP and CS scheme in [14] and use them dealing with the resource allocation problem for CAC in long term evolution (LTE) system. Stochastic Petri Net (SPN) is proposed by Bo Li et al. in [15] for N classes of traffic with different QoS requirements according to the number of needed channels, holding time and guard channels. More details for CAC schemes can refer to [16] and [17]. The major difficulty in obtaining performance metrics for CAC schemes in multi-class (more than three classes) cellular network is the high computational complexity [18]. As observed in [19], the dimension of states in the MDMC modeling increases very quickly, which results in the computing complexity due to solving a large set of flow equations. In the proposed schemes [11][12][13][14][15], few of them can support application with the number of service type more than three types. Moreover, the computing complexity of performance metrics deriving is rather complex by the MDMC modeling for the CAC schemes.

The service profiles of next generation mobile cellular networks are vastly different and the QoS demanded by these services also differs greatly. How to satisfy the diverse QoS requirements of these services of next generation mobile cellular network becomes even more challenging when facing to the reduced cell size and increased user mobility. However, to the best of the authors' knowledge, up to now there are only limited works reported in the literature regarding CAC schemes in mobile cellular network with service type more than three types. Moreover, the corresponding performance analyses for these CAC schemes are very complex due to solving a large number of flow equations based on the MDMC models. This motivates us to develop a CAC scheme for mobile cellular network with service type more than three types and to derive performance metrics for the proposed CAC scheme with low computing complexity.

In this paper, we propose a dynamical hybrid CAC scheme with integrated cutoff priority and handoff queue and derive some useful performance metrics with low computing complexity. In our CAC scheme, the higher priorities for handoff calls are realized through cutoff thresholds of occupied channels. The cutoff thresholds are dynamically adjusted according to the number of service types and service priority index. The handoff queue scheme is used to decrease P_{hd} and forced termination probability (PF), and to increase offered traffic utilization, channel utilization and system performance in terms of ratio of channel utilization to GoS cost function. By using an average service rate (the reciprocal of channel holding time), a 1DMC is used to model our CAC scheme and to derive performance metrics with low computing complexity. The main contribution of this paper lies in the following aspects. A dynamical hybrid cutoff priority CAC scheme combined with handoff queue scheme for multi services is proposed, which can support any number of service types. Moreover, some useful analytical performance metrics are derived with low computing complexity by modeling our CAC scheme with 1DMC model. Simulation is performed to verify the correctness of the derived performance metrics in terms of P_{nb} and PF. Simulation shows that the results are in good agreement with the derived performance metrics in terms of P_{nb} and PF. In addition, computing complexity of MDMC modeling for the performance analysis of our proposed CAC scheme is analyzed in detail. It is shown that the state-space cardinality of MDMC modeling increases exponentially with the number of service types and the number of total channels in a cell. However, the state-space cardinality of 1DMC modeling for the proposed CAC scheme is small and is determined by the number of channels in a cell and queue capacity of the highest priority service type.

2. System Model

This paper considers a mobile cellular network as shown in Fig.1, where each cell has C channels (i.e., time-slots for TDMA system, frequency channels for FDMA system, code sequence for CDMA system or sub-channels for OFDMA system). We focus our attention on one representative cell with equal and uniform loading across the service area [20][21]. Our goal is to design a CAC scheme for mobile cellular network which can support any number of service types and provide desired QoS requirement in terms of P_{nb} for new calls and PF for handoff calls. We assume that a representative cell has K ($K \geq 2$) types of service and each call requires the same bandwidth b_i ($i=1,2,\cdots,K$) for connection. Here it deserves to be specially noted that the number of service type K can be chosen with any positive integer as long as it is not less than two. The call of the i-th ($i=1,2,\cdots,K$) type can be further divided into new call and handoff call, denoted as n_i and h_i respectively. In our CAC scheme, we treat all the new calls fairly, i.e., all the new calls regardless of their types have the same QoS requirements in terms of P_{nb}, however, different types of handoff calls have different QoS requirements in terms of PF. Since users tend to be much more sensitive to call dropping than to call blocking, handoff calls are assigned higher priority over new calls. Moreover, the smaller PF the handoff call requires the higher priority will be provided. We assume that the priority of h_i is higher than h_j when $i > j$. We realize different priorities for different types of handoff call by reserving different number of guard channels for them. In our paper, the number of unguarded channels is called cutoff threshold. As shown in Fig.1, new calls of all types have the same cutoff threshold $\lfloor 2C/3 \rfloor$, which is only determined by the number of total channels in the representative cell. This means that once the number of occupied channels (by new or handoff calls) reaches the cutoff threshold $\lfloor 2C/3 \rfloor$, all the new calls regardless of their types will be blocked upon their arrival. However, handoff calls with different type have different cutoff thresholds. For example, handoff call h_i has the cutoff threshold equaling to $\left\lfloor \frac{2K+i}{3K}C \right\rfloor$, which is determined by the number of service type K, service type index i and the number of total channels in the representative cell C. This means that once the occupied channels (by new or handoff calls) reaches $\left\lfloor \frac{2K+i}{3K}C \right\rfloor$, handoff call of type i can not directly access to the network and be queued in Q_i waiting for released channels to continue its service when Q_i is not full upon its arrival. For the handoff call with service type index $1,2,\cdots,i,\cdots,K-1,K$, the cutoff thresholds are $\left\lfloor \frac{2K+1}{3K}C \right\rfloor$, $\left\lfloor \frac{2K+2}{3K}C \right\rfloor,\ldots,$ $\left\lfloor \frac{2K+i}{3K}C \right\rfloor,\ldots,$ $\left\lfloor \frac{2K+K-1}{3K}C \right\rfloor$ and C respectively. From the mapping of service type index to cutoff threshold, we can clearly see that the priority of h_i is higher than h_j when $i > j$.

The arrival process of the i-th type call is Poisson with arrival rate λ_{ni} for new call and λ_{hi} for handoff call ($i=1,2,\cdots,K$) respectively. In order to reduce the P_{hd} and PF, each handoff call of type i is provided with a queue Q_i with capacity L_i, however, queues are not provided for the new calls. Our CAC scheme works as follows: New calls n_i ($i=1,2,\cdots,K$) are accepted if the number of occupied channels (by new and handoff calls) is less than $\lfloor 2C/3 \rfloor$ upon their arrivals. Once the number of occupied channels is equal to or greater than $\lfloor 2C/3 \rfloor$, the new calls n_i ($i=1,2,\cdots,K$) will be blocked upon their arrival. The handoff call h_K (the K–th handoff call), which has the highest priority, is queued in Q_K if no idle channels are available upon its arrival and if the queue Q_K is not full. Once Q_K is full, the handoff call h_K will be blocked. Any other handoff call of h_i ($i=1,2,\cdots,K-1$) is queued in Q_i waiting for released channels to continue its service if the number of occupied channels is equal to or grater than $\left\lfloor \frac{2K+i}{3K}C \right\rfloor$ and if the queue Q_i is not full on its arrival. Once the Q_i is full, the handoff call h_i will be blocked. Since the handoff requests in queues are dropped as the mobile moves out of the handoff area before the handoff call is completed, the time out period will equals to the time interval that handoff requests stay in queue (within the handoff area) before it is forced to terminate. The time out period (dwell time) is assumed to have an exponential distribution with mean α_i^{-1} ($i=1,2,\cdots,K$) [8][21].

Call holding time is defined as the unencumbered conversation time between the beginning and completion of a call. The call holding time of call type i ($i=1,2,\cdots,K$) is assumed to have exponential distribution with mean μ_{ni}^{-1} for new call and μ_{hi}^{-1} for handoff call respectively. The cell residence time is defined as the amount of time that a mobile stays in a cell before handoff. The cell residence time is independent of the service type, we assume that the cell residence time of any type follows exponential distribution with the same mean η^{-1}.The channel holding time (channel occupancy time) is the minimum of cell residence time and call holding time, applying the memoryless property of the exponential distribution, the channel holding time of the *i*-th type call follows exponential distribution with mean $\sigma_{ni}^{-1}=(\eta+\mu_{ni})^{-1}$ for new call and $\sigma_{hi}^{-1}=(\eta+\mu_{hi})^{-1}$ for handoff call ($i=1,2,\cdots,K$). More detailed definition of call holding time, cell residence time and dwell time in handoff area is in [22][23].

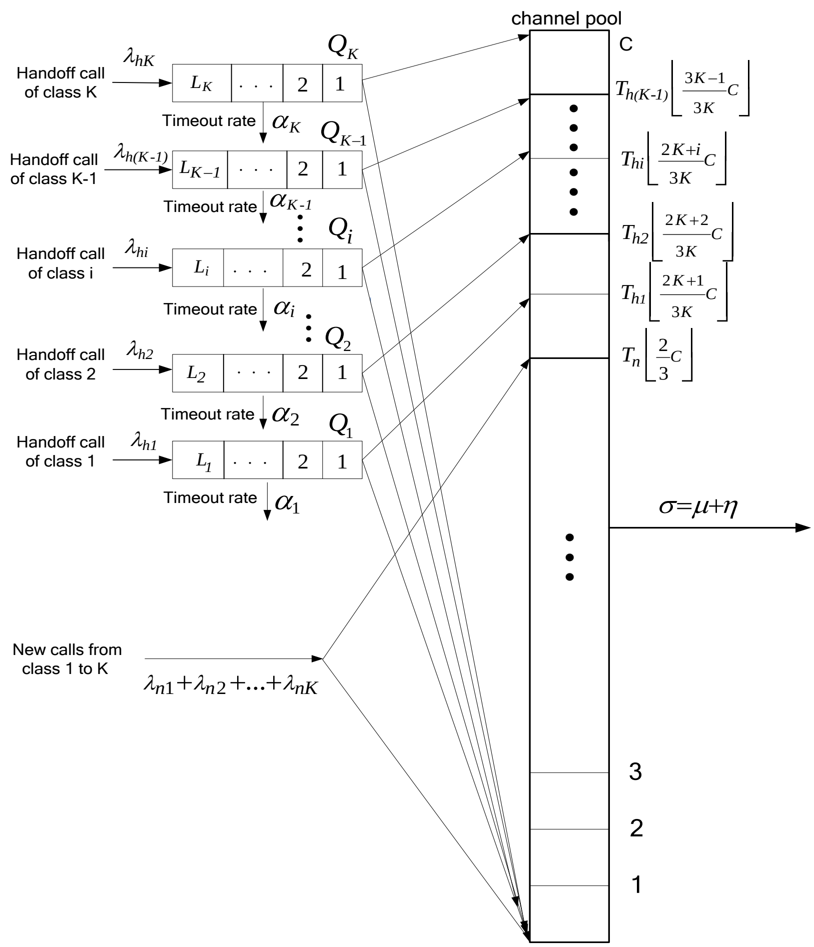

Fig. 1. A generic system model of CAC for multi-service

As described above, the proposed dynamic hybrid CAC scheme is realized through cutoff thresholds and queue scheme. The multi cutoff thresholds are used to realize QoS requirements in terms of P_{nb} for new call and PF for handoff calls. Moreover, the cutoff thresholds for handoff calls are not fixed, they can dynamically adjust according to the number of service types and service type index in a given cell which has C channels. This characteristic is very useful because the number of service type in a cell is not fixed and it is dependent on different time periods. In addition, the handoff call queue scheme can reduce P_{nb} and PF, and can increase the channel efficiency.

3. Performance Evaluation of the Dynamic Hybrid CAC Scheme

We define the sum of arrival rate for new calls from type m to K as $\lambda_n^m=\sum_{i=m}^{K}\lambda_{ni}$ and for handoff calls from type m to K as $\lambda_h^m=\sum_{i=m}^{K}\lambda_{hi}$ respectively as Wong,T.C. et al. do in [13], where $m=1,2,...,K$. Specially, we define the sum of arrival rate of all types of new call and handoff call as $\lambda=\sum_{i=1}^{K}(\lambda_{ni}+\lambda_{hi})$. Relative mobility (*rmob*) is defined as the ratio of handoff call rate to the total call rate [24] and denoted as $rmob_i=\lambda_{hi}/(\lambda_{hi}+\lambda_{ni}) \quad (i=1,2,...,K)$.

3.1 Average service time

If there are new calls from type *m* to *K*, handoff calls from type *n* to *K* access to the network, the average service time ($1/\mu_{mn}$) of multi calls can be approximted as

$$\frac{1}{\mu_{mn}}=\frac{1}{\lambda_n^m+\lambda_h^n}(\sum_{i=m}^{K}\lambda_{ni}\sigma_{ni}^{-1}+\sum_{j=n}^{K}\lambda_{hj}\sigma_{hj}^{-1}) \tag{1}$$

For a special case, if no new calls are admitted to the network, equation (1) can be simplified to

$$\frac{1}{\mu_{0n}}=\frac{1}{\lambda_h^n}\sum_{j=n}^{K}\lambda_{hj}\sigma_{hj}^{-1} \tag{2}$$

3.2 1DMC Model

In our model, we assume that the arrival process of n_i and h_i ($i=1,2,\cdots,K$) are all Poisson. Since $X_1(t)$ and $X_2(t)$ are independent Poisson process with paremeter λ_1 and λ_2 respectively, the sum of $X_1(t)$ and $X_2(t)$ is also a Poisson with parameter $\lambda_1+\lambda_2$. Due to page limitation, the proof is omitted. Given that the number of arrival service types is finite, the sum of call arrival process can be modeled as a possion process with rate equal to the sum of all the call arrival rates. Though average channel holding times are different both for different types of new calls and handoff calls , we can resort to an approximation of average channel holding time to describe the system's average service rate (the reciprocal of average service time). The transition state is defined as the number of occupied channels and queued handoff request. According to our proposed CAC scheme, with the known of average call arrival rate and average service rete at each state, obviously the transition process is 1DMC. For simplicity, it is assumed that each call (new or handoff call) of type *i* require b_i =1 channels for connection, the transition state for the 1DMC can be defined by state *i* (*i*=0,1, $\cdots$, $\lfloor 2C/3 \rfloor$, $\lfloor 2C/3 \rfloor+1$, ... $\left\lfloor \frac{2K+1}{3K}C \right\rfloor$, $\left\lfloor \frac{2K+1}{3K}C \right\rfloor+1$, $\left\lfloor \frac{2K+1}{3K}C \right\rfloor+2$, …, $\left\lfloor \frac{2K+1}{3K}C \right\rfloor+L_1$, $\left\lfloor \frac{2K+1}{3K}C \right\rfloor+L_1+1$, …, $\left\lfloor \frac{2K+2}{3K}C \right\rfloor$,

$\left\lfloor \frac{2K+2}{3K}C \right\rfloor+1$, …, $\left\lfloor \frac{2K+2}{3K}C \right\rfloor+L_2$, $\left\lfloor \frac{2K+2}{3K}C \right\rfloor+L_2+1$, …, $\left\lfloor \frac{2K+i}{3K}C \right\rfloor$, $\left\lfloor \frac{2K+i}{3K}C \right\rfloor+1$, …, $\left\lfloor \frac{2K+i}{3K}C \right\rfloor+L_i$,…, C, $C+1,\dots C+L_K$), ($L_i \le \left\lfloor \frac{2K+i+1}{3K}C \right\rfloor - \left\lfloor \frac{2K+i}{3K}C \right\rfloor$). The state transition diagram is shown in **Fig. 2**. The arc with right arrow presents average call arrival rate and the arc with left arrow represents average service rate. Due to the 1DMC modeling for performance analysis, solving a large set of flow equations can be avoided.

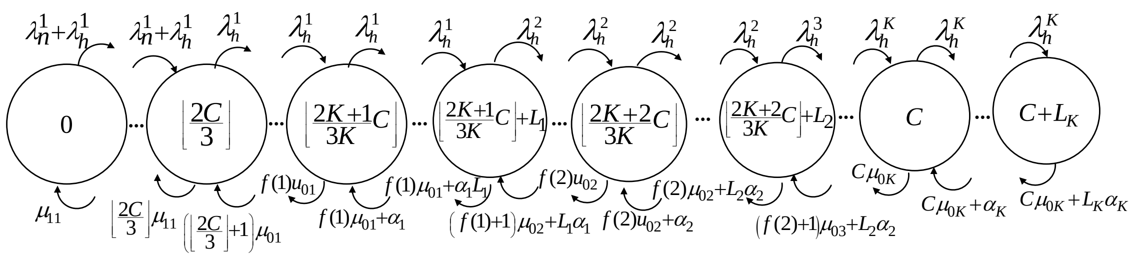

Fig. 2. State transition diagram of occupied channels. ($f(i)=\left\lfloor \frac{2K+i}{3K}C \right\rfloor, b_i=1$)

Form **Fig. 2**, the steady state probability p_i ($i=1,2,\cdots,C+L_K$) are determined by the following state balance equations.

$$iu_{11}p_i = \lambda p_{i-1} \qquad 1 \le i \le \left\lfloor \frac{2C}{3} \right\rfloor \tag{3}$$

$$\begin{cases} iu_{0m}p_i=\lambda_h^m p_{i-1} & X(m)+L_{m-1}+1\le i\le Y(m) \\ \left\{Y(m)u_{0m}+\left(i-Y(m)\right)\alpha_m\right\}p_i=\lambda_h^m p_{i-1} & Y(m)+1\le i\le Y(m)+L_m \end{cases} \tag{4}$$

where $X(m)=\left\lfloor \frac{2K+m-1}{3K}C \right\rfloor$ $Y(m)=\left\lfloor \frac{2K+m}{3K}C \right\rfloor$,m=1,2,…,K, L_0 =0，Solve (3) and (4) recursively, the steady state probability can be derived as

$$p_i=p_0\left(\frac{\lambda}{\mu_{11}}\right)^i \Big/ i! \qquad 1\le i\le \left\lfloor \frac{2C}{3} \right\rfloor \tag{5}$$

$$p_i=\begin{cases} \dfrac{p_{X(m)+L_{m-1}}\left(\lambda_h^m\right)^{i-X(m)-L_{m-1}}}{\prod_{j=1}^{i-X(m)-L_{m-1}}\left\{(X(m)+j)u_{0m}+L_{m-1}\alpha_{m-1}\right\}} & X(m)+L_{m-1}<i\le Y(m) \\ \dfrac{p_{Y(m)}\left(\lambda_h^m\right)^{i-Y(m)}}{\prod_{j=1}^{i-Y(m)}\left(Y(m)u_{0m}+j\alpha_m\right)} & Y(m)<i\le Y(m)+L_m \end{cases} \tag{6}$$

where $X(m)=\left\lfloor \frac{2K+m-1}{3K}C \right\rfloor$, $Y(m)=\left\lfloor \frac{2K+m}{3K}C \right\rfloor$, m=1,2,...K, L_0 =0, p_0 is obtain by normalization.

3.3 Performance metrics

Based on the steady state probability p_i (i=0, 1, 2, ⋯ , C, C+1, ⋯ , $C+L_K$) derived from (5) and (6), we can obtain the following performance metrics of the system. The average queue length of $Q_1, Q_2, \cdots, Q_K$ are given by

$$Q_{len}^{(m)}=\sum_{i=\left\lfloor \frac{2K+m}{3K}C \right\rfloor}^{\left\lfloor \frac{2K+m}{3K}C \right\rfloor+L_m} \left(i-\left\lfloor \frac{2K+m}{3K}C \right\rfloor\right)p_i \qquad m=1, 2, \dots, K \tag{7}$$

From **Fig. 1**, if the number of occupied channels is equal to or more than $\left\lfloor \frac{2C}{3} \right\rfloor$, new calls reguardless of their types will be blocked. So p_{nb} can be calculated as

$$p_{nb}=\sum_{j=\left\lfloor \frac{2C}{3} \right\rfloor}^{C+L_K} p_j \tag{8}$$

The blocking probability for any handoff call with type m can be expressed as

$$P_{hb}^{(m)}=\sum_{j=\left\lfloor \frac{(2K+m)C}{3K} \right\rfloor+L_m}^{C+L_K} p_j \qquad m=1, 2, \dots, K \tag{9}$$

For the service type m (m=1, 2, … , K), the average arrival rate of handoff call without blocking is $(1-P_{hb}^{(m)})\lambda_{hm}$, the mean number of dropped handoff calls in unit time is $\alpha_m Q_{len}^{(m)}$, therefore the time out probability, denoted as $P_{TO}^{(m)}$, can be given by

$$P_{TO}^{(m)}=\alpha_m Q_{len}^{(m)} \Big/ \left(1-P_{hb}^{(m)}\right)\lambda_{hm} \qquad m=1, 2, \dots, K \tag{10}$$

PF consists of two parts, the first part is P_{hb}, the second part is composed of those handoff calls which are not blocked but timeout in the queue. So PF for the handoff call of type m can be expressed as

$$PF^{(m)}=P_{hb}^{(m)}+(1-P_{hb}^{(m)})P_{TO}^{(m)} \qquad m=1, 2, \dots, K \tag{11}$$

Using Little's formula, the average waiting time in the queue for handoff call of type m, denoted as $Q_{wait}^{(m)}$, can be given as

$$Q_{wait}^{(m)}=Q_{len}^{(m)} \Big/ \left(1-P_{hb}^{(m)}\right)\lambda_{hm} \qquad m=1, 2, \dots, K \tag{12}$$

The wireless channel utilization is defined as the ratio of the number of average occupied channels to the total channels.

$$C_{utilization}=\left(\sum_{i=1}^{\left\lfloor\frac{2C}{3}\right\rfloor} ip_i+\sum_{m=1}^{K}\left(\sum_{j=\left\lfloor\frac{2K+m-1}{3K}C\right\rfloor+L_{m-1}+1}^{\left\lfloor\frac{2K+m}{3K}C\right\rfloor} jp_j+\sum_{j=\left\lfloor\frac{2K+m}{3K}C\right\rfloor+1}^{\left\lfloor\frac{2K+m}{3K}C\right\rfloor+L_m}\left\lfloor\frac{2K+m}{3K}C\right\rfloor p_j\right)\right)\Big/C \quad m=1,2,...,K \tag{13}$$

where C is the number of total channels in the representative cell .

Offered traffic, denoted as λ_O, represents the sum of arrival rate for all new calls and handoff calls of different types. λ_O can be expressed as

$$\lambda_O=\sum_{i=1}^{K}\left(\lambda_{ni}+\lambda_{hi}\right) \tag{14}$$

Total carried traffic, denoted as λ_C, represents the amount of traffic admitted to the cellular network as opposed to the offered traffic. λ_C can be expressed as

$$\lambda_C=\sum_{i=1}^{K}\left(\left(1-PF^{(i)}\right)\lambda_{hi}+\left(1-P_{nb}\right)\lambda_{ni}\right) \tag{15}$$

The offered traffic utilization, denoted as $Tuti$, is the ratio of λ_C to λ_O. $Tuti$ can be expressed as

$$Tuti=\frac{\lambda_C}{\lambda_O} \tag{16}$$

The designed goal of handoff scheme should minimize the Grade of Service (GoS) cost function. GoS for K types of service can be defined as

$$GoS=P_{nb}+r\sum_{i=1}^{K}\frac{\lambda_{hi}}{\lambda_h^1}PF^{(i)} \tag{17}$$

where $\lambda_h^1=\sum_{i=1}^{K}\lambda_{hi}$, r is the penalty factor used to reflect the effect of the handoff dropping over the new call blocking in the GoS cost function. A penalty of 5 to 20 times is commonly recommended in [25]. From the service provider's perspective, it's better to increase the channel utilization. However, from the user's point of view, it's better to minimize the GoS cost function. So in order to make a fair balance between providers' and users' satisfaction, we define the system performance as

$$Perf_{system}=\frac{C_{ultilization}}{GoS} \tag{18}$$

where $C_{ultilization}$ and GoS are defined in (13),(17) .

3.4 Comparison of computing complexity

From subsection 3.2, 1DMC is used to model our CAC scheme. If we use a MDMC to model our CAC scheme, the computing complexity is huge. In the following subsection, we compare the computing complexity between the 1DMC modeling and MDMC modeling. For MDMC, the transition diagram can be represented in a 2K-dimensional Markov chain, where each transition state is denoted in a 2K-dimensional vector $(N_1, N_2, \cdots, N_K, H_1, H_2, \cdots, H_K)$. Here N_i

and H_i are nonnegative integers and represent the number of channels occupied by n_i and h_i ($i = 1, 2, ..., K$) respectively. For convenience of expression, we define $\binom{n}{k}$ as follows

$$\binom{n}{k} = \begin{cases} \frac{n!}{k!(n-k)!} & 0 \le k \le n \\ 0 & n < 0 \text{ or } k < 0 \text{ or } n < k \end{cases} \tag{19}$$

When using MDMC, the buffered handoff in the queue is neglected since the queue size for handoff call is finite. Each element N_i and H_i ($i = 1, 2, ..., K$) in the transition state vector $(N_1, N_2, \cdots, N_K, H_1\ \ H_2, \cdots, H_K)$ should satify a set of inequalities simultaneously as following.

$$\begin{aligned} N_1 + N_2 + \cdots + N_K + H_1 + H_2 + \cdots + H_K &\le C \\ 0 \le N_1 &\le \left\lfloor \tfrac{2}{3} C \right\rfloor \\ 0 \le N_2 &\le \left\lfloor \tfrac{2}{3} C \right\rfloor \\ &\vdots \\ 0 \le N_K &\le \left\lfloor \tfrac{2}{3} C \right\rfloor \\ 0 \le H_1 &\le \left\lfloor \tfrac{2K+1}{3K} C \right\rfloor \\ 0 \le H_2 &\le \left\lfloor \tfrac{2K+2}{3K} C \right\rfloor \\ &\vdots \\ 0 \le H_K &\le \left\lfloor \tfrac{2K+K}{3K} C \right\rfloor \end{aligned} \tag{20}$$

The solution for state-space cardinality can be transfered to find all the possible solutions of inequality set (20). The solution for inequality set (20) can be partitioned into C+1 problems depicted as $N_1 + N_2 + \cdots + N_K + H_1 + H_2 + \cdots + H_K \le m$, $(m = 0, 1, 2, \cdots, C)$ with constraints $0 \le N_i \le \left\lfloor \frac{2}{3} C \right\rfloor$, $0 \le H_i \le \left\lfloor \frac{2K+i}{3K} C \right\rfloor$, where $i = 1, 2, \cdots K$ .In order to solve the $C+1$ problems,we resort to the inclusion principle [26]. The solution for each of the $C+1$ problem can be expressed as

$$N_S(m) = \binom{2K + (C-m) - 1}{C-m} - K \binom{2K + (C - m - \left\lfloor \frac{2}{3} C \right\rfloor - 1) - 1}{C - m - \left\lfloor \frac{2}{3} C \right\rfloor - 1} - \sum_{j=1}^{K} \binom{2K + (C - m - \left\lfloor \frac{2K+j}{3K} C - 1 \right\rfloor) - 1}{C - m - \left\lfloor \frac{2K+j}{3K} C \right\rfloor - 1} \tag{21}$$

where $m = 0, 1, 2, \cdots C$. Thus the state-space cardinality N_{S_MDMC} can be expressed as

$$N_{S\_MDMC} = \sum_{m=0}^{C} N_S(m) = \sum_{m=0}^{C} \left(\binom{2K + (C-m) - 1}{C-m} - K \binom{2K + (C - m - \left\lfloor \frac{2}{3} C \right\rfloor - 1) - 1}{C - m - \left\lfloor \frac{2}{3} C \right\rfloor - 1} - \sum_{j=1}^{K} \binom{2K + (C - m - \left\lfloor \frac{2K+j}{3K} C - 1 \right\rfloor) - 1}{C - m - \left\lfloor \frac{2K+j}{3K} C \right\rfloor - 1} \right) \tag{22}$$

Equation (22) indicates that the state-space cardinality of MDMC is related to both the number of total channel *C* and the number of total servie type *K*. For visual expression, state-space cardinality of MDMC related to *K* and *C* is plotted in **Fig. 3**.

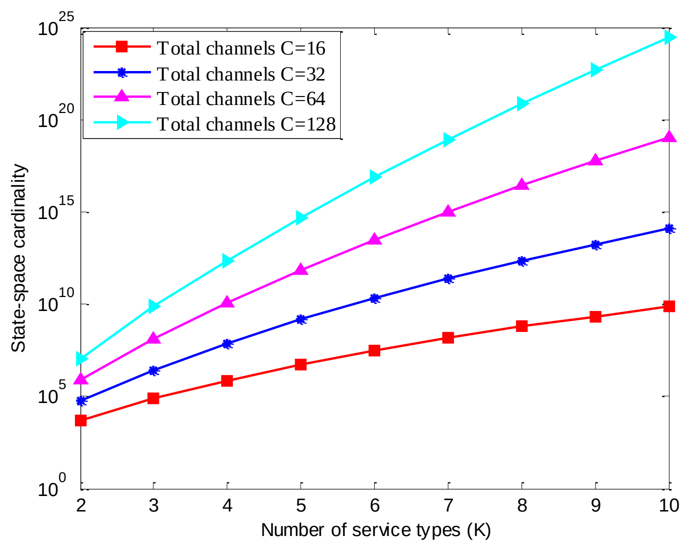

Fig. 3. The state-space cardinality of MDMC versus the number of service types K.

It is indicated that the state-space cardinality increases exponentially with *K* under given number of total channels in a cell. Besides, when *K* is fixed, the state-space cardinality also inceases exponentially with C. That means that it is hard to illustrate all the balance transition states and that it is rather impractical to write all the balance equations since the state-space cardinality of a MDMC is huge. However, in 1DMC model, the state-space cardinality is very simple and can be expressed as

$$N_{S\_1DMC}=C+L_K+1 \quad (23)$$

From (23), we can know that the state-sapce cardinality of 1DMC is only determined by *C* and queue capacity L_K for handoff call with the highest priority. Since the total available channels in the reference cell and queue capacity L_K are finite, the state-space cardinality of 1DMC model is limited to acceptable number. This makes the performance analysis of CAC scheme become more simple and tractable.

From (22) and (23), we can conclude that 1DMC modeling for cutoff priority based CAC scheme with multi cutoff thresholds is much more simpler than MDMC modeling when carrying out performance analysis. This is because the computing complexity of MDMC increase exponentially with the number of service type *K* and total channels C in a cell, however, the computing complexity of 1DMC increase linearly with *C*.

4. Numerical Results

In this section, we firstly validate the analytical performance metrics for P_{nb} and $PF^{(m)}$ through simulations, and then investigate the effect of relative mobility, traffic load and queue capacity on the system performance. Lastly, we compare the performance of our CAC scheme with that of existing CAC scheme in [7] in terms of new call blocking probability, forced termination probability of handoff call and channel utilization respectively. Though our CAC scheme can support any number of service type, for illustrating explicitly, we assume that there are four types of service application in the representative cell, i.e., K=4 and calls with different type have the same mean call arrival rate (including new and handoff call). We mark the four types of services as 1, 2, 3 and 4 with an increasing priority order. The main parameters are assumed according to [27] and are shown in **Table 1** for analysis.

Table 1. System parameters

Call Types	Call Holding Time $\mu_{ni}^{-1}/\mu_{hi}^{-1}$ (sec)	Cell Residence Time η^{-1} (sec)	Queue Capacity L_i (calls)	Dwell Time α_i^{-1} (sec)	Total Channels per Cell C
New call n_i (i=1, 2, 3, 4)	[180 150 120 90]	[180 180 180 180]	—	—	64
Handoff call h_i (i=1, 2, 3, 4)	[180 150 120 90]	[180 180 180 180]	[4 4 4 4]	[40 30 20 10]	

Fig. 4, **Fig. 5** and **Fig. 6** present respectively P_{nb} and PF in terms of analysis and simulation results versus traffic load under different $rmob$. As shown in **Fig. 4**, **Fig. 5** and **Fig. 6**, the simulation results of P_{nb} and PF match closely with their corresponding analytic results with $rmob$ equal to 0.2, 0.5 and 0.8 respectively. In addition, the higher priority the handoff call is, the smaller PF it has. This can well satisfy the QoS requirements in terms of PF for different handoff calls. That is to say, the handoff call with a high priority can preferentially occupy the idle channels. **Fig. 7** shows that the impact of combined queue for handoff calls on P_{nb} under different $rmob$. It is indicated that the impact of queue capacity on P_{nb} can be neglected when $rmob$ is small and moderate. Even with large $rmob$, the impact of queue capacity on the increase of P_{nb} is slight. The reason for the slight increase of P_{nb} is because of the decrease of PF due to the queue employing for handoff calls. With the decrease of PF, more handoff calls are accepted. Since handoff call has higher priority over new call, which is realized through cutoff thresholds of occupied channels, new calls will have less opportunity to occupy idle channels when competing with handoff calls. This leads to the increase of P_{nb} due to insufficient channels allocated to the new calls. **Fig. 8** presents the impact of queues for handoff calls on PF with a large $rmob$ equal to 0.8. As can be seen from Fig.8, PF of all service types are reduced greatly with the use of queues for handoff calls. We must point out that the decrease of PF is at the expense of slight increase of P_{nb} .The reason for the decrease of PF is

due to the decrease of handoff call blocking probability with the queue for handoff calls. **Fig. 9** shows that the channel utilization can be improved with the queue for handoff calls. In addition, under a relatively larger $rmob$, the system's channel utilization can be improved much more. The reason for this can be explained as following. In a scenario with a larger $rmob$, more handoff calls will arrive in unit time. If queues are provided for handoff calls, they can wait in the queue for idle channels to continue their services. This will result in a smaller value of PF than that with no queue for handoff calls. Though there is slight increase in P_{nb}, the decrease of PF is more sharply, which will improve the system's channel utilization. In a scenario with small $rmob$, for a new call of certain type, as long as the number of occupied channels (by new call and handoff call) reaches its cutoff threshold, the new call will be blocked though there are idle channels which are reserved for handoff calls. This means that in a scenario with small $rmob$, the queue scheme shows minor contribution to the improvement of channel utilization. As shown in **Fig. 10**, all curves can be grouped into three sets according to different $rmob$. For each group, the channel utilization increases with the traffic load, and with a fixed traffic load the channel utilization increases with the number of service type K. Moreover, with the increase of $rmob$, the channel utilization gap due to different number of service type K decreases gradually. This means that channel utilization can be improved when the system serves more types of call, especially in relative low $rmob$ scenario. **Fig. 11** shows that the system offered traffic utilization decreases with the traffic load. The reason for the decrease of offered traffic utilization is that the P_{nb} and PF increase with the traffic load and this results in the decrease of carried traffic. In addition, in a scenario with larger $rmob$, offered traffic utilization can be improved much more with the employing of queue scheme for handoff calls. However, this advantage is not prominent in a scenario with small $rmob$. The reason for this is the same as that for the increase of channel utilization with queue handoff scheme. **Fig. 12** shows the impact of queue capacity on system performance under different $rmob$. As shown in **Fig. 12**, firstly, system performance decreases with the increase of the traffic load. Secondly, under the same traffic load, the system performance decreases with the increase of $rmob$. Lastly, for a fixed $rmob$ and a given traffic load, the system performance increases with the employing of queue for handoff calls.

Since Mahmoud ASH's scheme in [7] only support two types of services, we assume that there are two types of services in each cell. For the fairness of comparison, the following assumptions are made. The number of channels C in each cell is 50. The mean cell residency time is 3 minutes, and the mean call holding time is also 3 minutes. The mean timeout for voice and data handoff calls are 10 and 15 seconds respectively. The voice and data calls (including new calls and handoff calls) have the same contribution to the offered load. New call and handoff call have the same $rmob=1/3$, i.e., $\lambda_{hi} = 0.5\lambda_{ni}$ (i=1,2). The guard channels are set at $G1 = 17$ and $G2 = C$, i.e., the threshold for new call is $M = C - G1 = 33$, queue capacity for voice handoff call Q1=5, queue capacity for data handoff call is Q2=0 and sub rating channels S=10 for Mahmoud ASH's scheme in [7]. For our scheme, the queue capacities for data and voice handoff calls are $L_1 = 5$ and $L_2 = 5$ respectively.

Fig. 13, **Fig. 14** and **Fig. 15** show the comparison of CAC performance in terms of new call blocking probability, forced termination probability for handoff calls and channel utilization between our scheme and Mahmoud ASH's scheme. As shown in **Fig. 13**, our scheme has smaller new call blocking probability than Mahmoud ASH's scheme when the offered load is below 195 Erlang. As shown in **Fig. 14**, the forced termination probability of voice handoff call ($PF\text{-}voice$) of our scheme is smaller than that of Mahmoud ASH's scheme when the offered load is beyond 150 Erlang. The forced termination probability of data handoff call ($PF\text{-}data$) of our scheme is less than that of Mahmoud ASH's scheme when the offered load is below 255 Erlang, however is larger than that of Mahmoud ASH's scheme when the offered load is beyond 255 Erlang. The reason for this can be explained as following. In our scheme, the data call has the cutoff threshold $\lfloor (2K+1)C/3K \rfloor$=41 channels when the number of service type $K=2$. However, for Mahmoud ASH's scheme, the cutoff threshold for data handoff is C=50. This means that our scheme gives high priority to voice handoff calls by restricting the admission of data calls, which results in larger value of $PF\text{-}data$ when the offered load is high. Fig.15 shows that channel utilization of our scheme is higher than that of Mahmoud ASH's scheme when the offered load is below 120 Erlang, but this advantage disappears when the offered load is beyond 120 Erlang. The reason for this is the increased value of $PF\text{-}data$ in our scheme when the offered load is heavy.

5. Conclusion

In this paper, a dynamic hybrid CAC scheme with integrated cutoff priority and handoff queue was proposed, and some useful analytical performance metrics were derived. Detailed analysis indicates that the proposed 1DMC model greatly reduces the computing complexity of performance analysis comparing with MDMC model. Extensive simulations show that the derived analytic performance metrics in terms of P_{nb} and PF match closely with the simulation results. Different from other CAC schemes with fixed cutoff thresholds, the cutoff thresholds in our CAC scheme are dynamically adjusted according to the number of service types and the service priority index. Numerical analysis shows that the employing of dynamical cutoff thresholds can improve channel utilization with the increase of the number of service types, especially in a small $rmob$ scenario. Besides, the incorporated queue scheme can further enhance the system performance metrics such as reducing PF of handoff calls, increasing offered traffic utilization, improving channel utilization and upgrading system performance. Compared with the conventional CAC schemes, our CAC scheme has no limitation on the number of service types and can be used in mobile cellular network with different multiple access techniques such as TDMA, FDMA, CDMA and OFDMA. Numerical results show that our CAC scheme performs better than Mahmoud ASH's scheme to some extend. For further research, we will focus on adaptive cutoff priority threshold design according to not only the number of service types but also the traffic characters in terms of traffic load to improve the channel utilization of the mobile cellular network. In addition, the problem that different types of calls require different bandwidth will also be discussed in our further research.

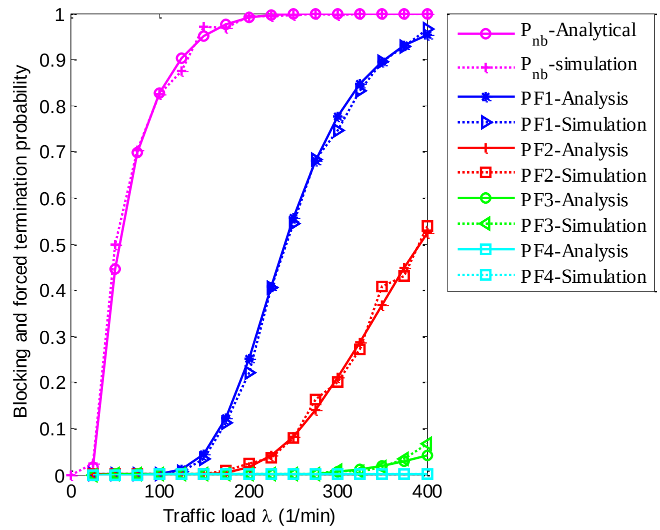

Fig. 4. P_{nb} and PF versus the traffic load λ when *rmob*=0.2.

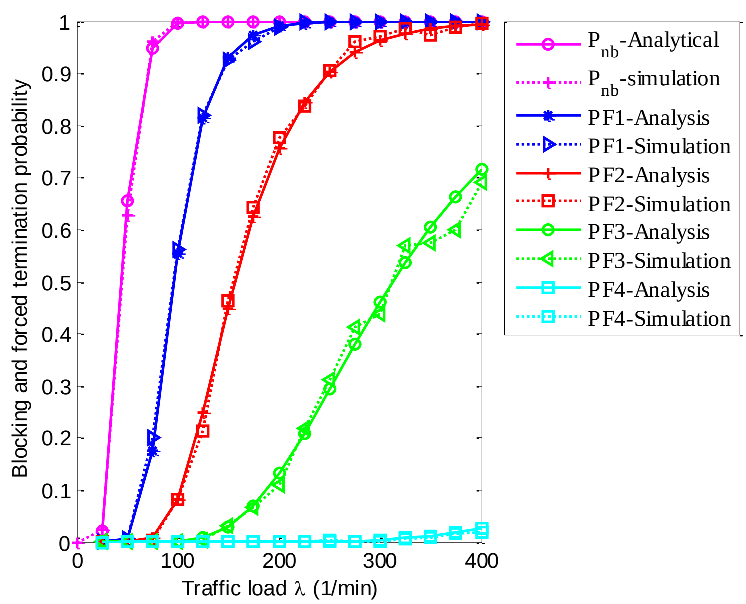

Fig. 5. P_{nb} and PF versus the traffic load λ when *rmob*=0.5.

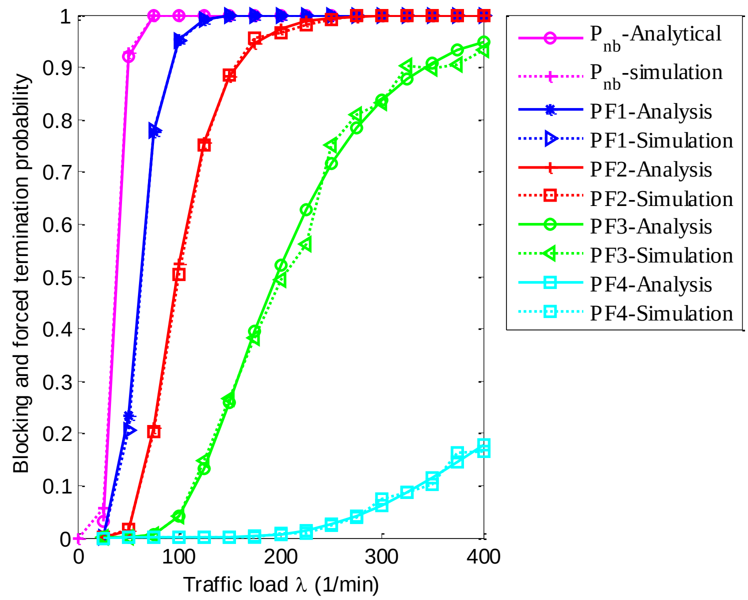

Fig. 6. P_{nb} and PF versus the traffic load λ when *rmob*=0.8.

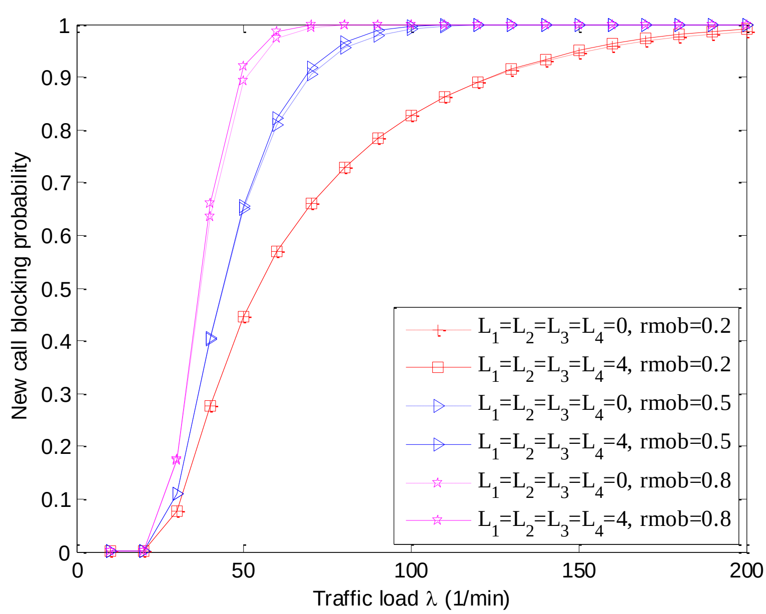

Fig. 7. The impact of queue capacity on P_{nb} under different *rmob*.

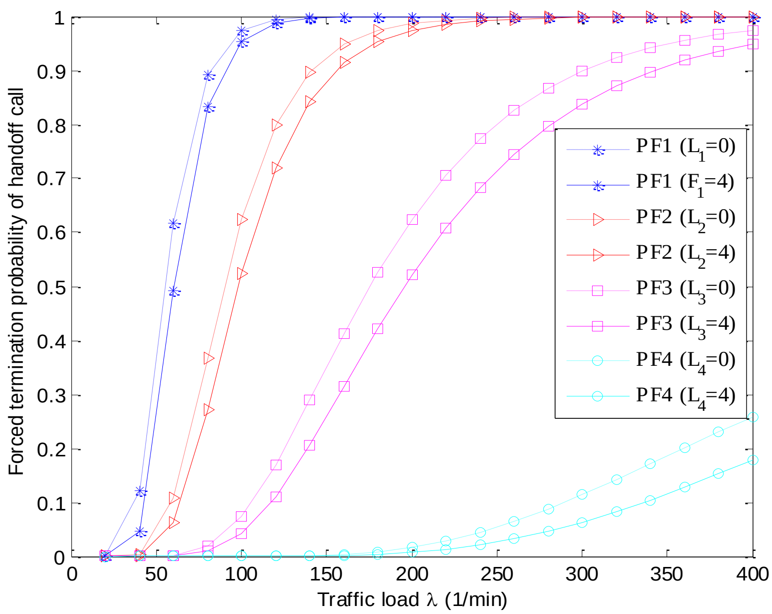

Fig. 8. The impact of queue capacity on PF when *rmob*=0.8.

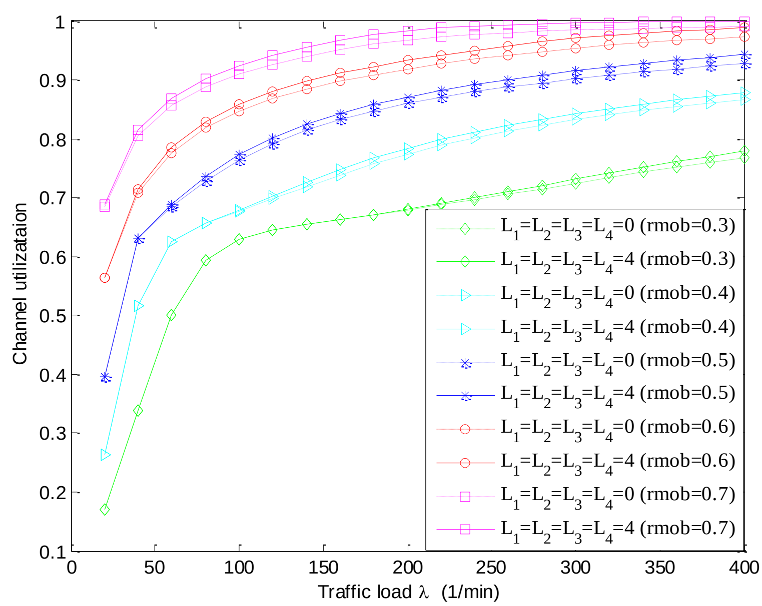

Fig. 9. The impact of queue capacity on channel utilization under different *rmob*.

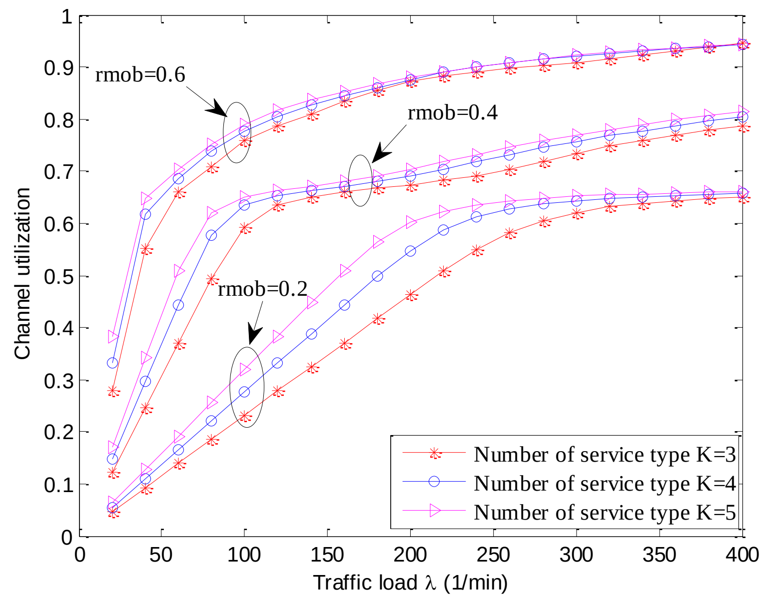

Fig. 10. The impact of the number of service type *K* on channel utilization under different *rmob*.(*C*=128)

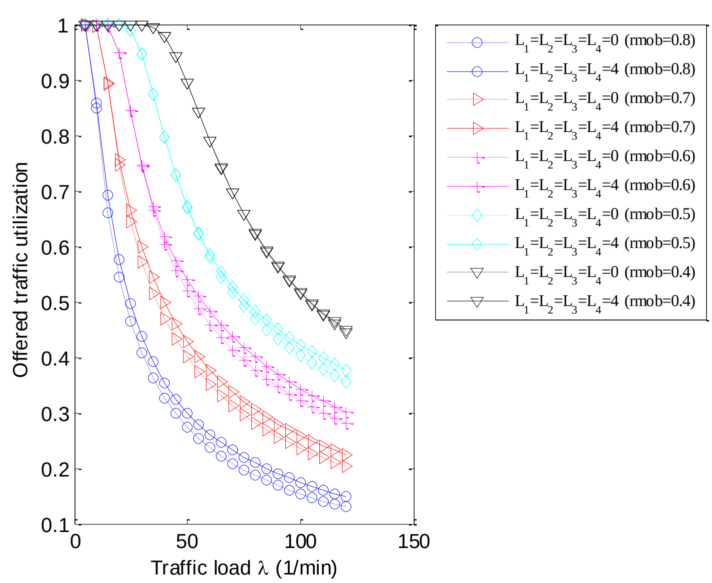

Fig.11. The impact of queue capacity on offered traffic utilization under different *rmob*.

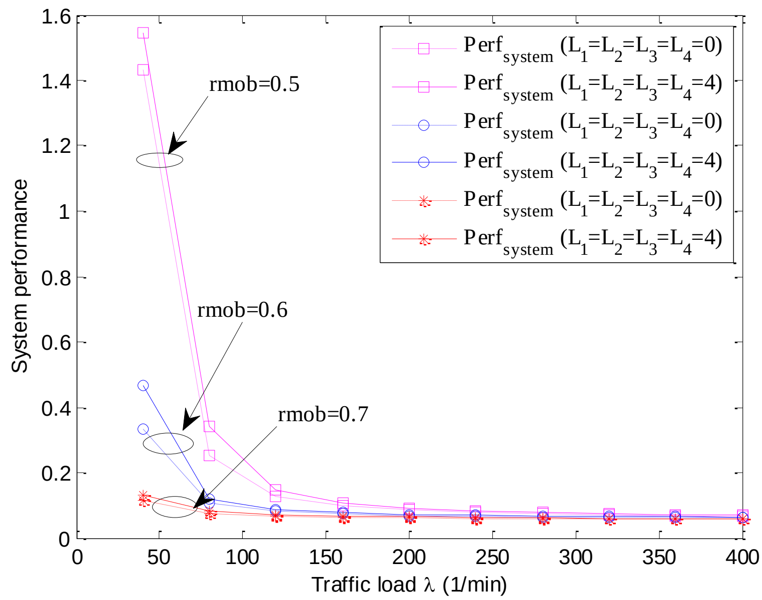

Fig. 12. The impact of queue capacity on system performance under different *rmob* (*r*=10).

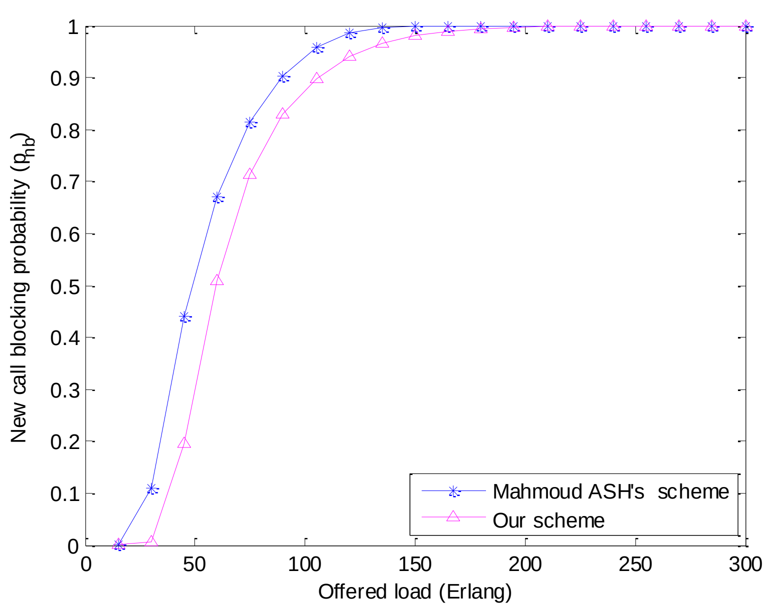

Fig. 13. New call blocking probability (P_{nb}) versus offered load (Erlang).

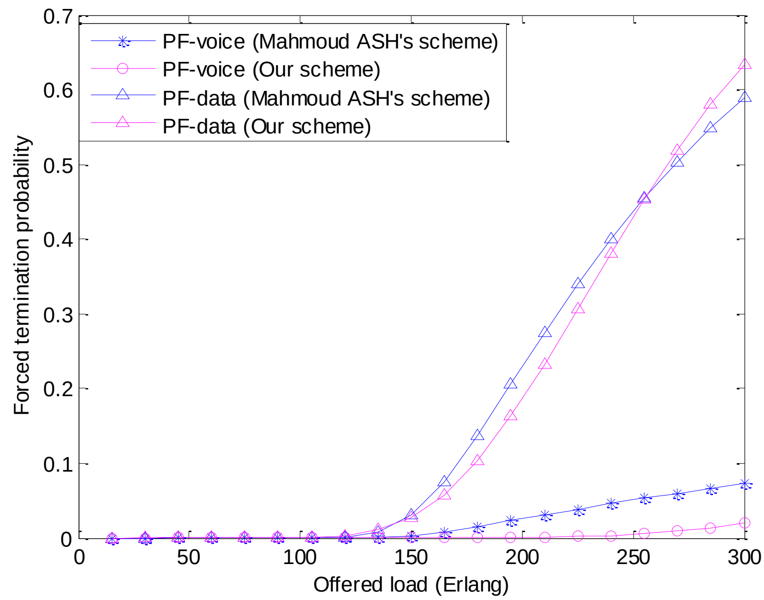

Fig. 14. Forced termination probability (PF) versus offered load (Erlang).

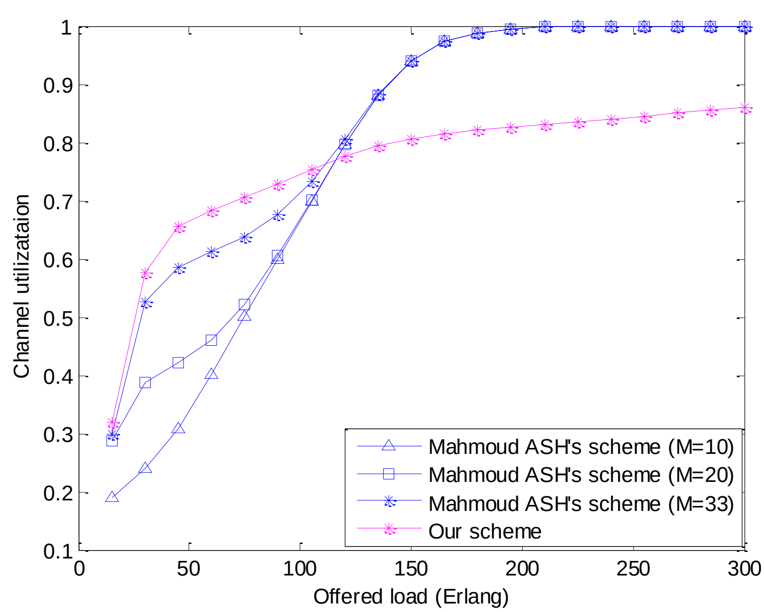

Fig. 15. Channel utilization versus offered load (Erlang).

References

[1] Leong, C. W, Weihua Zhuang and Yu Cheng; Lei Wang, "Optimal resource allocation and adaptive call admission control for voice/data integrated cellular networks". *IEEE trans. on Vehi. Tech*., vol.55, pp.654 - 669, Mar.2006. Article (CrossRef Link)

[2] Far. L. M. and Beigy. H, "A new fractional call admission control scheme in integrated cellular network". IEEE conf. EIT2009, pp.332–337, Jun.2009. Article (CrossRef Link)

[3] Leong, C. W, Weihua Zhuang, Yu Cheng and Lei Wang, "Call admission control for integrated on/off voice and best-effort data services in mobile cellular communications", *IEEE Trans. on communications*, vol.52, no.5, pp.778-790, 2004. Article (CrossRef Link)

[4] Yieh-Ran Haung, Yi-Bing Lin, and J. M. Ho, "Performance analysis for voice/data integration on a finite-buffer mobile system", *IEEE Trans. on Veh. Tech.* vol.49, no.2, Feb.2000. Article (CrossRef Link)

[5] L. Yin, B. Li, Z. Zhang and Yi-Bing Lin, "Performance analysis of a dual threshold reservation (DTR) scheme for voice/data integrated mobile wireless networks", in *Proc.of IEEE WCNC2000*, pp.258-262, Sep.2000. Article (CrossRef Link)

[6] T.-C. Chau, K. Y. M. Wong and B. Li, "Optimizing call admission control with QoS guarantee in a voice/data integrated cellular network using simulated annealing", in *Proc. IEEE Conf. GLOBECOM'02*, Nov.2002. Article (CrossRef Link)

[7] Mahmoud ASH, AlQahtani SA, "A framework of call admission control procedures for integrated services mobile wireless networks", *Arabian Journal for science and Engineering*, vol.32, No.1B, pp.115-130, Apr.2007. Article (CrossRef Link)

[8] Chung-Ju, Tian-Tsair Su and Yueh-Yiing Chiang, "Analysis of a cutoff priority cellular radio system with finite queuing and reneging/dropping", *IEEE Trans. on networking*, vol.2, no.2, pp.166-175, Apr.1994.Article (CrossRef Link)

[9] Wei Li, Hang Chen, and Agrawal, D. P. "Performance analysis of handoff schemes with preemptive and nonpreemptive channel borrowing in integrated wireless cellular networks", *IEEE Trans. on wireless communications*, vol.4, no.3, pp.1222-1233, 2005. Article (CrossRef Link)

[10] Niyato, D. Hossain, and E. Alfa, A.S. "Performance analysis of multi-service wireless cellular networks with MMPP call arrival patterns", in *Proc. of IEEE conf. GLOBECOM'04*, vol.5, pp. 3078-3082, Nov.2004. Article (CrossRef Link)

[11] Nasser, N. and Hassanein, H., "Prioritized multi-class adaptive framework for multimedia wireless networks", in *Proc. IEEE conf. on communication*, vol.7, pp.4295-4300, Jun.2004. Article (CrossRef Link)

[12] Hu, F. and Sharma, N. K., "Priority-determined multiclass handoff scheme with guaranteed mobile QoS in wireless multimedia networks". *IEEE Trans. on Vehi. Tech*, vol.53, no.1, pp.118-135, Jan.2004.Article (CrossRef Link)

[13] Wong, T. C, Mark, J. W. and Chua, K. C. , "Resource allocation in mobile cellular networks with QoS constraints", in *Proc. of IEEE conf. WCNC2002*, vol.2, pp.717-722, Mar.2002. Article (CrossRef Link)

[14] Manli Qian, Yi Huang, Jinglin Shi, Yao Yuan, Lin Tian, and Dutkiewicz,E., "A novel radio admission control scheme for multiclass services in LTE systems". in *Proc. of IEEE conf. GLOBECOM'09*, pp.1-6, Nov.2009.Article (CrossRef Link)

[15] B. Li, S. T. Chanson, and C. Lin, "Analysis of a hybrid cutoff priority scheme for multiple classes of traffic in multimedia wireless networks", *wireless networks*, vol.4, no.4, pp.279-290, Jul.1998. Article (CrossRef Link)

[16] Vassilya Abdlova and Isik Aybay, "Handoff prioritizing schemes in cellular wireless networks". in

Proc. of the seventh IEEE International Symposium on Computer Networks, pp.79-84, Jun.2006. Article (CrossRef Link)

[17] Aggeliki Sgora and Dimitrios D. vergados, "Handoff prioritization and decision scheme in wireless cellular networks: a survey", *IEEE Commnications Surveys & Tutorials*. vol.11, no.4, pp.57-77, 2009. Article (CrossRef Link)

[18] B. Epstein and M. Schwartz, "Reservation strategies for multi-media traffic in a wireless environment", in *Proc. IEEE Vehicular Technology Conference,* vol.1, pp.165-169, Jul.1995. Article (CrossRef Link)

[19] Rappaport, S. S, "Blocking, hand-off and traffic performance for cellular communication systems with mixed platforms", *in Proc. Inst. Elect. Eng*. I, vol.140, pp.389-401, 1993. Article (CrossRef Link)

[20] Song Liu, Zhisheng Niu and Daei Huang, "Performance analysis of voice message service in CDMA cellular system", *in Proc. of ICCT2003*, pp.891-895, Apr.2003. Article (CrossRef Link)

[21] D. Hong and S.S. Rappaport. "Traffic model and performance analysis for cellular mobile radio telephone systems with prioritized and non-prioritized handoff procedures", *IEEE Transactions on Vehi. Tech.*, vol.VT- 35, no.3, pp.77-92, Aug.1986. Article (CrossRef Link)

[22] Y.-B. Lin, S. Mohan, and A. Noerpel, "Queuing priority channel assignment strategies for PCS hand-off and initial access", *IEEE Trans. Vehi. Tech.*, vol.43, pp.704–712, Aug.1994. Article (CrossRef Link)

[23] Y.-B. Lin, *Introduction to Mobile Network Management*, ser. Telecommunications: National Chiao Tung University, 1997.

[24] Jens Zander, and Seong-Lyun Kim, *Radio Resource Management for Wireless Networks*, Aretch House Publishers, pp.141, 2001.

[25] F. Barcelo, "Performance analysis of handoff resource allocation strategies through the state-dependent rejection scheme", *IEEE Trans. on wireless communications*, vol.3, no.3, pp.900-909, 2004. Article (CrossRef Link)

[26] Richard A. Brualdi, *Introductory Combinatorics*, forth edition, Pearson Education, 2004.

[27] Mischa Schwartz, *Mobile Wireless Communications*, Cambridge University Press, Cambrige, 2005.

Jiping Li was born in Hubei, China. He received the B.S. and M.S. degrees in computer science from Hubei University and Ocean University of China in 2001 and 2006 respectively. He is presently pursuing the Ph. D degree in department of electronic and information engineering, Central China Normal University.
His research interests include call admission control, wireless resource allocation for OFDMA system and the application of Zigbee network.

Shixun Wu was born in Hubei, China. He received the B.S. and M.S. degrees in mathematic in 2006 and 2009 respectively. He is presently pursuing the Ph. D degree at the Department of Electrical and Computer Engineering, Central China Normal University, China.
His current research interests include wireless communication and radio positioning.

Shouyin Liu received the BS degree in physics and the MS degree in radio electronics from Central China Normal University, Wuhan, China, in 1985 and in 1988, respectively. He received the Ph. D degree from Hanyang University, Korea in 2005 in electronic communication engineering. From 2004, he has been a professor at Central China Normal University.
His current research interests include digital communication, WSN and location techniques.