We are IntechOpen,
the world's leading publisher of Open Access books
Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

154
Countries delivered to

Our authors are among the
TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

12

Providing Emergency Services in Public Cellular Networks

Jiazhen Zhou[1] and Cory Beard[2]
[1]*Howard University, Washington DC*
[2]*University of Missouri - Kansas City*
United States

1. Introduction

When emergency situations like natural disasters or terrorist attacks happen, demand in telecommunication networks will go up drastically, causing congestion in the networks. Due to the local nature of most disaster events, this kind of congestion is usually most serious at access networks, which is of special concern for cellular networks. With serious congestion in the cellular networks, it is very difficult for customers to obtain access to services. It might be possible to use reserved spectrum for national security and emergency preparedness (NS/EP) customers, such as with police and fire radio systems. However, capacity may be limited. Also, the deployment of additional equipment takes time and could not address the urgent need for communications quick enough.

On the other hand, publicly available wireless communication capabilities are pervasive and always ready to use. It would be very beneficial if NS/EP customers could use the commercially available wireless systems to respond to natural and man-made disasters (Carlberg et. al., 2005).

For several years, but especially in response to the events of September 11, 2001, the U.S. government and the wireless telecommunications industry have worked together to specify a technically and politically feasible solution to the needs of homeland security for priority access and enhanced session completion. This has resulted in definition of an end-to-end solution for national security and emergency preparedness sessions called the Wireless Priority Service (WPS) defined in the Wireless Priority Service Full Operating Capability (WPS FOC) by the FCC (FCC, 2000), (National Communications System, 2002), (National Communications System, 2003). First-responders, NS/EP leadership, and key staff are able to use this capability in public cellular networks.

To support emergency services in public cellular networks, NS/EP users should be identified and provided better guaranteed services than general customers. When NS/EP customers present access codes and have been authorized to use the emergency service, *special admission control policies* are employed in the base station to make sure their session requests get better admission. Proper methods are also deployed in the core network to provide end-to-end service for NS/EP users.

The basic requirement on the *special admission control policies* is that better admission of NS/EP customers, including both high admission probability and quick access, should be guaranteed. However, as the main purpose of public cellular networks is also to provide services for public customers, certain resources should be ensured for public users even when there are high demands of spectrum resources from NS/EP users at the same time. For this reason, we can say that NS/EP traffic has "conditional priority" — the resources allocated for the NS/EP customers must be balanced with the demand of public users. If NS/EP traffic is light, low blocking probability for them should be guaranteed; if NS/EP traffic becomes unexpectedly heavy, then public traffic should be protected through guaranteeing a certain amount of resources for public use.

In this Chapter, important parameters for NS/EP and public customers are first identified. Three candidate admission control policies are then analyzed , with two types that provide high system utilization evaluated and compared in detail. Each control policy has significant benefits and drawbacks with neither one clearly superior, so guidelines are provided for choosing best schemes that are suitable to each operator and social context.

2. Important performance metrics

There are different possible admission control strategies for providing emergency services in public cellular networks. The important performance metrics that should be considered in evaluating these strategies include: system utilization, admission probability of public and NS/EP customers, access waiting time, and termination probability.

2.1 System utilization

System utilization is a measurement of the usage of system resources, like the spectrum resources in a cellular network. In normal situations, system utilization directly determines the revenues of the operators. The higher the system utilization, the higher the revenues the operators will obtain.

Since disasters are mostly unexpected and only happen occasionally, when we are considering possible admission control strategies, we need to make sure they can help get maximum system utilization while providing priority services to NS/EP users.

2.2 Admission of NS/EP and public traffic

Priority treatment for NS/EP traffic gives high probability of admission when demand is not extremely high (for example, less than 25% of a cell's capacity). Operators will set a threshold for expected NS/EP load, and will admit sessions with high probability if demand stays within those bounds.

As an effective measurement on whether public customers are well protected in case of high NS/EP traffic, channel occupancy of public traffic is used, which measures the average amount of channel resources utilized. For example, (Nyquetek, 2002) points out that it has been agreed by government and operators that at least 75% of channel resources should be guaranteed for public use. To achieve this goal, flexible admission control strategies should be employed to cope with different load cases.

2.3 Waiting time

Queueing based methods, where session requests are put into queues so that they will be served when system resources are available later, are very common in admission control strategies since they are effective in increasing the utilization of systems. However, this means that some customers will inevitably wait some time before being allocated a channel to start their sessions. The access waiting time experienced by customers, which decides the customer satisfaction, is another important metric used to evaluate the quality of service in telecommunication systems.

For NS/EP customers, long waiting for admission is unreasonable due to the urgent need of saving life and property. An ideal admission control policy should cause minimum or even no waiting for NS/EP customers.

2.4 Session termination probability

Session termination probability is concerned with reasons why a session might be terminated before it is able to be completed. This might occur because of a failed handoff or a hard preemption.

In a cellular network, a commonly agreed upon standard is that terminating an ongoing session is much worse than blocking a new attempt. This is why channels are often reserved for handoff traffic in cellular networks. However, the cost is that the system utilization can be sacrificed. This is a dilemma also encountered in emergency situations: should we try to guarantee few ongoing public sessions are terminated, or should we make higher system utilization more important so that we can admit more NS/EP sessions? Different operators might have different opinions on this issue. However, this also reminds us that some termination of sessions, especially if this is rare, might be acceptable during specific situations like when disasters happen.

3. Candidate admission control strategies

The admission control policies discussed in this paper are all load based. This means that admission is based on whether the new session will make the load on the system too high. The appropriateness of this load based admission control model for a 3G/4G network is discussed in Section 6.

3.1 Reservation based strategies

To guarantee certain resources for a special class of traffic, reservation strategies (Guerin, 1988) are often used in cellular networks; one example is guard channel policies to hold back resources for handoffs. The main idea of a reservation based strategy is that it limits the amount of sessions that can be admitted for some classes to hold back resources in case other classes need them. The benefit is that the high priority traffic could use specially reserved resources, thus achieving better admission. Yet the disadvantage is that the reserved resources can be wasted if the high priority traffic is not as high as expected, while the low priority traffic is probably suffering blocking. As discussed previously in section 2.1, the spectrum resources are especially valuable when disasters happen, so reserving channels for NS/EP traffic will probably cause waste of resources since NS/EP traffic volume is hard to predict. This is why the strategies employing reservation schemes (including guard channel policy (Guerin, 1988) and upper limit strategy (Beard & Frost, 2001)) are not as useful in public cellular networks that support emergency services.

3.2 Pure queueing based strategies

For a pure queueing based policy, all classes of traffic can have their own queues or shared queues. Session requests that cannot get immediate service will be put into queues. When system resources become available, session requests in the queues will be scheduled according to some specific rule to determine which queue gets served next. Since pure queueing based strategies will try to serve customers waiting in the queues whenever system resources become available, the newly released resources will be immediately taken and thus *guarantee no waste of a system's capability*.

In a pure queueing based scheme, both NS/EP traffic and public originating traffic can be put into separate queues. When channels become available, these two queues will be served according to a certain probability (Zhou & Beard, 2010), or using round-robin style scheduling (Nyquetek, 2002). For the former work, the scheme is called Adaptive Probabilistic Scheduling (APS). The latter work by Nyquetek Inc. evaluated a series of pure queueing based methods, with a representative one being the Public Use Reservation with Queuing All Calls (PURQ-AC).

3.3 Preemption based strategies

As opposed to pure queueing based strategies, preemption based strategies allow high priority customers to take resources away from ongoing low priority sessions. Preempted sessions can be put into a queue so that they can be resumed later, hopefully after a very short time. It can be shown that preemption strategies can even obtain slightly higher system utilization than pure queueing strategies.

The largest benefit of preemption based strategies is that they can guarantee immediate access and assure the admission of NS/EP traffic. However, an uncontrolled preemption strategy tends to use up all channel resources and will be against the goal to protect public traffic. Furthermore, it can be annoying to preempted users. Due to these side effects, in some places like the United States it is not currently allowed, even though allowed in other places.

In (Zhou & Beard, 2009), a controlled preemption strategy was presented to suppress these side effects while exploiting the unique benefits. When the resources occupied by the NS/EP sessions surpass a threshold, preemption will be prohibited. By tuning the preemption threshold, the channel occupancy for each class can be adjusted as we prefer.

4. Analysis for the admission control schemes

The adaptive probabilistic scheduling (APS) scheme and the preemption threshold based scheduling (PTS) scheme can be both analyzed using multiple dimensional Markov chains. The main performance metrics, including admission and success probability, waiting time, and termination probability can be computed.

The main types of sessions considered are emergency sessions, public handoff sessions and public originating sessions. As shown in (Nyquetek, 2002), the current WPS is provided only for leadership and key staff, so it is reasonable to assume that most emergency users will be stationary within a disaster area. Handoff for emergency sessions is not considered here, but this current framework can be readily extended to deal with emergency handoff traffic when necessary.

Strictly speaking, the distributions of session durations, inter-arrival time, and the length of customer's patience are probably not exponentially distributed. As shown in (Jedrzycki & Leung, 1996), the channel holding times in cellular networks can be modeled much more accurately using the lognormal distribution. Consult (Mitchell et. al., 2001) where Matrix-Exponential distributions can be added to standard Markov chains like the one here to model virtually any arrival or service process. However, in reality the exponential distribution assumption for sessions is still mostly used, both in analysis-based study like (Tang & Li, 2006), and simulation-based study like Nyquetek's report (Nyquetek, 2002). For this work, similar to what is used widely and also assumed in (Nyquetek, 2002), all session durations and inter-arrival times are independently, identically, and exponentially distributed.

4.1 Adaptive probabilistic scheduling

In (FCC, 2000), the FCC provided recommendations and rules regarding the provision of the Priority Access Service to public safety personnel by commercial providers. It required that "at all times a reasonable amount of CMRS (Commercial Mobile Radio Service) spectrum is made available for public use." To meet the FCC's requirements, when emergency traffic demand is under a certain "protection threshold", high priority should be given to emergency traffic; when the emergency traffic is extremely high so that it can take most or all of the radio resources, a corresponding strategy should be taken to avoid the starvation of public traffic by guaranteeing a certain amount of radio resources will be used by public. The "protection threshold" can be decided by each operator and thus is changeable. So our strategy should be able to deal with the above requirement for any "protection threshold" value; this is why we introduce an *adaptive probabilistic scheduling* strategy instead of fixed scheduling like in (Nyquetek, 2002).

4.1.1 Description and modeling of APS

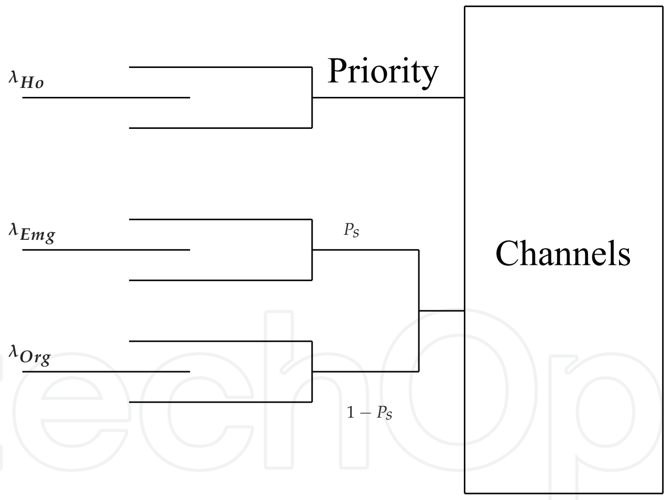


Fig. 1. Probabilistic scheduling

The basic APS scheme is illustrated in Fig. 1. In the figure, $\lambda_{Emg}, \lambda_{Ho}, \lambda_{Org}$ represent the arrival rate of emergency, public handoff, and public originating sessions respectively. When an incoming session fails to find free channels, it is put into a corresponding queue if the queue is not full. To reduce the probability of dropped sessions, the handoff sessions are assigned a non-preemptive priority over the other two classes of traffic. Note that when a disaster happens, it is uncommon for general people (who generate public traffic) to move into a disaster area. Thus, the handoff traffic into a disaster area will not be high, so setting the public handoff traffic as the highest priority will not make emergency traffic starve. If

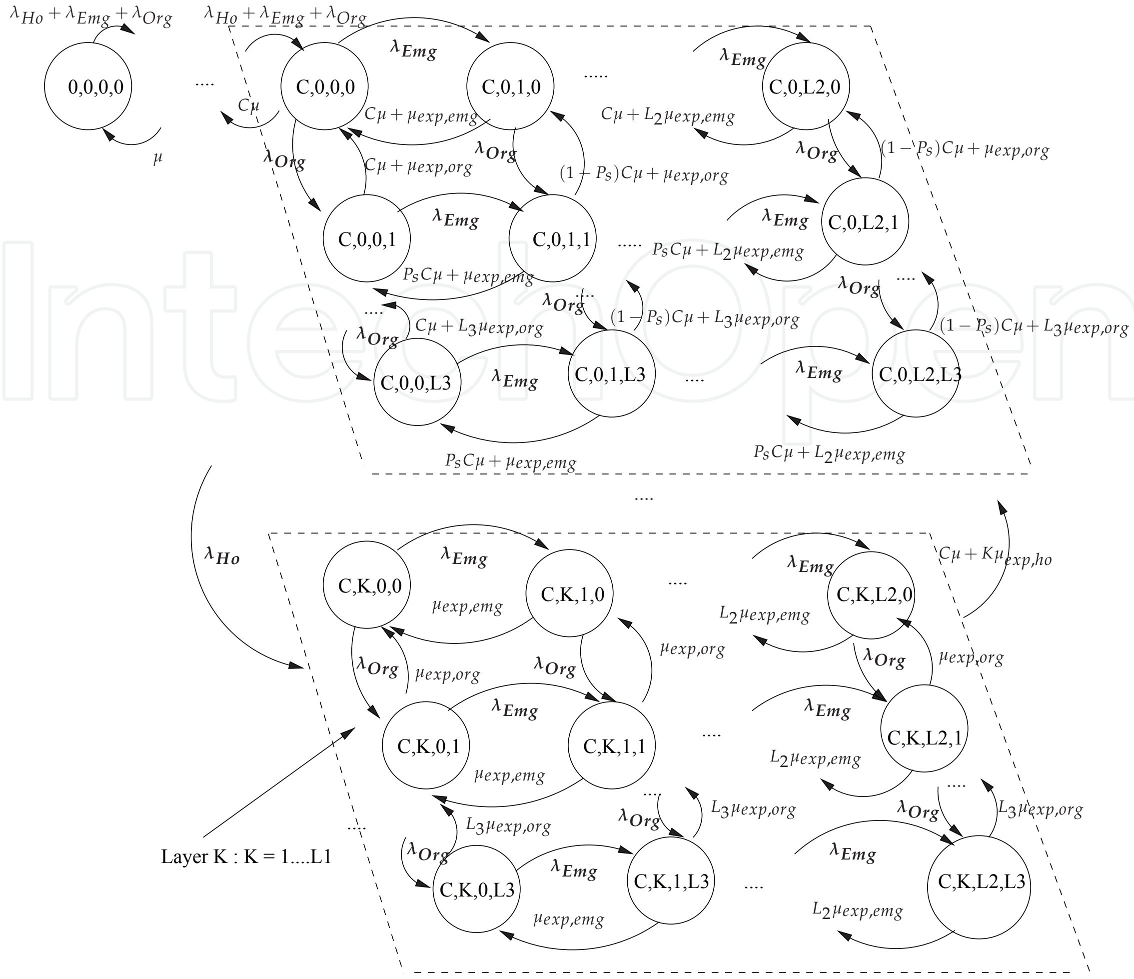


Fig. 2. State Diagram for Probabilistic Scheduling Scheme

there is no session waiting in the handoff queue when a channel is freed, a session will be randomly chosen from either the emergency queue or public originating queue according to the scheduling probability already set. The scheduling probability for emergency traffic is denoted as P_s. The algorithm to decide P_s for different cases will be introduced in Section 4.1.3. The queues are finite and customers can be impatient when waiting in the queue, so blocking and expiration are possible.

A 3-D Markov chain can be built to model the behavior of the three classes of traffic as shown in Fig. 2. Here the total number of channels is C, and the queue lengths are L_1, L_2, L_3 individually. Each state is identified as (n, i, j, k), where n is the number of channels used, i, j, k is the number of sessions in handoff queue, emergency queue, and public originating queue respectively. The arrival rate for handoff, emergency and originating sessions is λ_{Ho}, λ_{Emg}, λ_{Org}, and the service rate for all sessions is μ. The expiration times of all three classes of sessions are exponentially distributed with rates $\mu_{exp,ho}$, $\mu_{exp,emg}$ and $\mu_{exp,org}$ respectively.

The probabilistic scheduling policy is implemented using a parameter P_s, which is the probability that an emergency session will be scheduled when a channel becomes free. This can be seen in Fig. 2, for example, where the part of the departure rates from state (C,0,L2,L3) that relate to scheduling are either $P_sC\mu$ or $(1-P_s)C\mu$ for choosing to service an emergency or public session respectively.

We can see that the Markov chain in Fig. 2 does not have a product form solution. This is because the boundary (first layer) is not product form due to the probabilistic scheduling. So state probabilities will be obtained by solving the global balance equations from this Markov chain directly.

It is worthy to mention that the Markov chain size in Fig. 2 is $L_1L_2L_3$ states, thus it is not affected by the number of channels. Since the queue size employed does not need to be large (like 5) due to the reneging effect of customers waiting in the queue, the computational complexity of this Markov chain is lower than the preemption threshold strategy in (Zhou & Beard, 2009), which has a Markov chain of size CL_1L_2. Note that a typical value of C is around 50.

4.1.2 Performance evaluation

With the state probabilities solved, performance metrics can be computed for blocking, expiration and total loss probability, admission probability, average waiting time, and channel occupancy for each class.

In this system, the loss for each class of traffic consists of two parts: those sessions that are blocked when the arrivals find the queue full; and those sessions reneged (also called expired) when waiting too long in each queue. So we have: $P_{Loss} = P_B + P_{Exp}$ for each class of traffic.

(1) Blocking probability

Blocking for each class of traffic happens when the corresponding queue is full. Thus,

$$P_{B,ho} = \sum_{j=0}^{L_2} \sum_{k=0}^{L_3} P(C, L_1, j, k) \quad (1)$$

$$P_{B,emg} = \sum_{i=0}^{L_1} \sum_{k=0}^{L_3} P(C, i, L_2, k) \quad (2)$$

$$P_{B,org} = \sum_{i=0}^{L_1} \sum_{j=0}^{L_2} P(C, i, j, L_3) \quad (3)$$

(2) Expiration Probability

At a state (C, i, j, k), the arrival rates for each class are λ_{Ho}, λ_{Emg}, λ_{Org}, and the expiration rates for each class are $i\mu_{exp,ho}$, $j\mu_{exp,emg}$, $k\mu_{exp,org}$ independently. The probability of expiration is the ratio of departures due to expiration per unit time (expiration rate) over arrivals per unit time (arrival rate). Thus, we can find the overall expiration probability just based on the steady state probability, the expiration rate, and the arrival rate at each state as follows:

$$P_{Exp,ho} = \sum_{i=1}^{L_1} \sum_{j=0}^{L_2} \sum_{k=0}^{L_3} P(C, i, j, k) \frac{i\mu_{exp,ho}}{\lambda_{Ho}} \quad (4)$$

$$P_{Exp,emg} = \sum_{i=0}^{L_1} \sum_{j=1}^{L_2} \sum_{k=0}^{L_3} P(C, i, j, k) \frac{j\mu_{exp,emg}}{\lambda_{Emg}} \quad (5)$$

$$P_{Exp,org} = \sum_{i=0}^{L_1} \sum_{j=1}^{L_2} \sum_{k=0}^{L_3} P(C, i, j, k) \frac{k\mu_{exp,org}}{\lambda_{Org}} \quad (6)$$

(3) System utilization and channel occupancy

The channels are not fully used when there are still free channels available. When there are n channels being used, that means $C - n$ channels are idle, and the total portion of channels unused is thus $\frac{C-n}{C}$. So the system utilization can be calculated by considering those portion of channels unused at those possible states:

$$SysUtil = 1 - \sum_{n=0}^{C-1} \frac{(C-n)P(n,0,0,0)}{C} \tag{7}$$

We define "channel occupancy" as the proportion of channels occupied by each class of traffic. *Channel occupancy* is an important metric to measure whether the public traffic is well protected when emergency traffic is heavy. After the system utilization is obtained, with the assumption that each class of session has the same average session duration, the channel occupancy of each class can be calculated by comparing the admitted traffic of each class. The total admitted traffic rate is: $\lambda_{adm,tot} = \lambda_{Ho}(1 - P_{B,ho}) + \lambda_{Emg}(1 - P_{B,emg}) + \lambda_{Org}(1 - P_{B,org})$. Thus, we have:

$$ChOcp_{ho} = \frac{\lambda_{Ho}(1 - P_{B,ho})}{\lambda_{adm,tot}} SysUtil \tag{8}$$

$$ChOcp_{emg} = \frac{\lambda_{Emg}(1 - P_{B,emg})}{\lambda_{adm,tot}} SysUtil \tag{9}$$

$$ChOcp_{org} = \frac{\lambda_{Org}(1 - P_{B,org})}{\lambda_{adm,tot}} SysUtil \tag{10}$$

(4) Waiting time
Sessions waiting in the queue could be patient enough to wait until the next channel becomes available, or become impatient and leave the queue before being served.
If a customer needs to be put into a queue before being served or reneging, the average time staying in the queue (irrespective of being eventually served or not) can be calculated using Little's law: $\overline{T} = \frac{N_q}{\lambda(1-P_B)}$. Note that the effective arrival rate at each queue is $\lambda(1 - P_B)$, and the average queue length for each class is N_q. The mean queue length for each class can be calculated based on the steady states we compute, so we have:

$$\overline{T_{ho}} = \sum_{i=0}^{L_1} \sum_{j=0}^{L_2} \sum_{k=0}^{L_3} \frac{iP(C,i,j,k)}{\lambda_{Ho}(1 - P_{B,ho})} \tag{11}$$

$$\overline{T_{emg}} = \sum_{i=0}^{L_1} \sum_{j=0}^{L_2} \sum_{k=0}^{L_3} \frac{jP(C,i,j,k)}{\lambda_{Emg}(1 - P_{B,emg})} \tag{12}$$

$$\overline{T_{org}} = \sum_{i=0}^{L_1} \sum_{j=0}^{L_2} \sum_{k=0}^{L_3} \frac{kP(C,i,j,k)}{\lambda_{Org}(1 - P_{B,org})} \tag{13}$$

4.1.3 The adaptive probabilistic scheduling algorithm

With main performance metrics like channel occupancies computed, an algorithm for searching the best value of P_S can be obtained.
Denote the "system capacity" as the largest possible throughput of the cell. In other words, it is the total service rate of the cell - $C\mu$.
Using dynamic probabilistic scheduling, the scheduling probability for emergency traffic when there is a channel available can be adjusted according to the different arrival rates for

each class of traffic. The algorithm to find the scheduling probability for emergency traffic P_s is:

Algorithm 1: Determining the scheduling probability

Step 1: Set the initial value of P_s to 1, which means giving absolute priority to emergency traffic as opposed to public originating traffic.

Step 2: Solving the Markov chain, get the general representation about channel occupancies using equations (8) - (10) . With the current P_S value applied, if the channel occupancy of public traffic is already higher than 75%, that means the emergency traffic obviously does not affect the performance of public traffic, and thus can be accepted, stop here. Otherwise go to step 3 to search for the suitable weighting parameter.

Step 3: Use a binary search method to search for the best weighting parameter: Let $P_s = 1/2$, calculate the channel occupancy of public traffic using the general representation obtained in step 2. If it is larger than the required value, search the right half space [1/2, 1]; otherwise search the left half space [0, 1/2]. Repeat step 3 until the suitable P_s that meets the channel occupancy requirement of public traffic is found.

4.2 Preemption threshold based scheduling

4.2.1 Description and modeling of PTS

The preemption threshold based scheduling scheme (PTS) is illustrated in Fig. 3. When an incoming emergency session fails to find free capacity, and if the number of active emergency sessions is less than the preemption threshold, it will preempt resources from a randomly picked ongoing public session. The preempted session will be put into the handoff/preempted session queue. For an arriving public handoff session, it will also be buffered in the handoff/preempted session queue when no capacity is immediately available. Correspondingly, there is also an originating session queue, which is further helpful for preventing starvation of public traffic. If an incoming emergency session fails to find free resources to preempt, it will simply be dropped.

We suggest not to have a buffer for emergency users for two reasons: (1) Make sure there is no access delay for emergency sessions; (2) Guarantee the public traffic has enough system resources when emergency traffic is very heavy. If emergency traffic is queued in this case, public traffic could not be well protected even if preemption is not allowed. The reason that we use the same buffer for handoff and preempted sessions is that both of these two types of sessions are broken sessions, so they have the same urgency to be resumed. More precise configuration like using two different buffers is possible, but will not be obviously beneficial. In fact, it will make the implementation and analysis more time consuming, because it will have a much larger Markov chain state space.

When capacity becomes available later, one session from the queues is served. A priority queue based scheduling policy will be used, and it is reasonable to assume that handoff/preempted sessions have higher priority over the originating sessions. The queues are finite and customers can be impatient when waiting in the queue, so blocking and expiration are possible.

Since customers have different patience, it is reasonable to assume their impatience behavior to be random rather than deterministic like assumed in Nyquetek's study. We assume that the expiration times of traffic in the same queue are exponentially and identically distributed, and the patience of a customer is the same after each preemption.

If session durations are memoryless (i.e., exponentially distributed), this means that if at any point a session is interrupted, the remaining service time is still exponential with the same

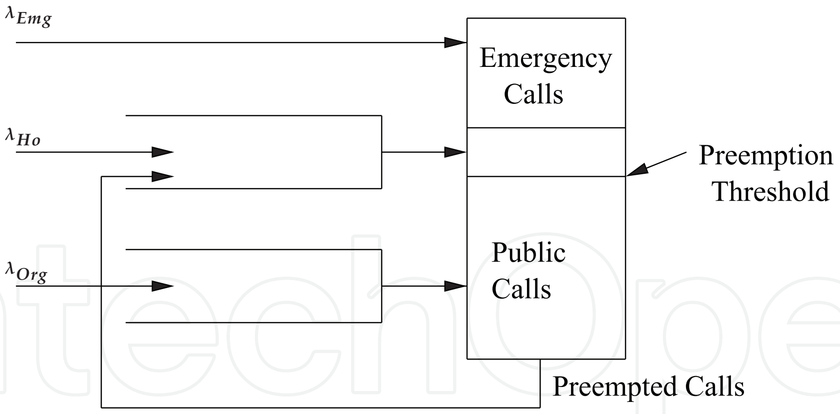


Fig. 3. Preemption threshold based scheduling

average service time as when it began. It is, therefore, reasonable to model a restarted session as a renewal process. In other words, the preempted session will be restarted with re-sampling of the exponential random variable (Conway et. al., 1967).

Same as for the APS scheme, the total number of channels is denoted as C. The length of the handoff/preempted queue is L_1 and the length of originating queue is L_2, and the *preemption threshold* is R. Each state is identified as (i, j, m, n), where i, j is the number of channels occupied by emergency and public sessions respectively, m, n represents the number of sessions in the handoff/preempted session queue and the public originating session queue individually. The arrival rates for emergency, handoff, and originating sessions are λ_{Emg}, λ_{Ho}, λ_{Org} respectively. The mean expiration rates for sessions waiting in the handoff/preempted queue and originating queue are denoted as $\mu_{exp}^{ho/prm}$ and μ_{exp}^{org}. To facilitate analysis, the average service rate for each class is assumed to be the same and denoted as μ. This also means that the session duration in a single cell is exponentially distributed with mean $1/\mu$, whether the session ends in this cell or is handed off to another cell. A Markov chain can be formed, and the state probabilities can be obtained by solving the following categories of balance equations:

(1) When the channels are not full, the typical state transition is shown in Fig. 4. Since the queues are empty in this case, in the notation we replace $P(i, j, 0, 0)$ with $P(i, j)$ for simplicity. The corresponding balance equation is:

$$\begin{aligned}&P(i,j)(\lambda_{Emg} + \lambda_{Ho} + \lambda_{Org} + (i+j)\mu)\\ &= P(i-1,j)\lambda_{Emg} + P(i,j-1)(\lambda_{Ho} + \lambda_{Org})\\ &+P(i,j+1)(j+1)\mu + P(i+1,j)(i+1)\mu. \end{aligned} \tag{14}$$

For the states on the edge, some terms of this equation will disappear.

(2) When the channels are full, queueing is involved, the typical state transition is shown in Fig. 5. The corresponding balance equation is:

$$\begin{aligned}&P(i,C-i,m,n)(\lambda_{Emg} + \lambda_{Ho} + \lambda_{Org} + C\mu + m\mu_{exp}^{ho/prm} + n\mu_{exp}^{org})\\ &= P(i-1,C-i+1,m-1,n)\lambda_{Emg} + P(i,C-i,m-1,n)\lambda_{Ho}\\ &+P(i,C-i,m,n-1)\lambda_{Org} + P(i+1,C-i-1,m+1,n)(i+1)\mu\\ &+P(i,C-i,m+1,n)((C-i)\mu + (m+1)\mu_{exp}^{ho/prm}) + P(i,C-i,m,n+1)(n+1)\mu_{exp}^{org} \end{aligned} \tag{15}$$

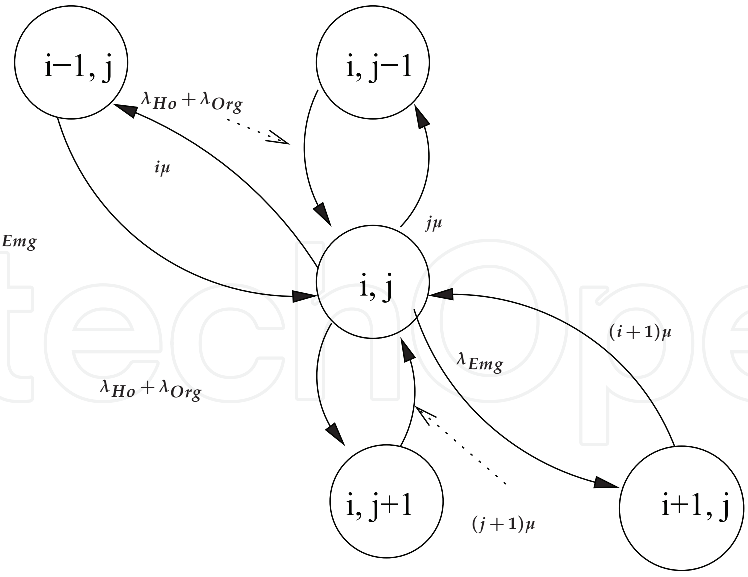


Fig. 4. The typical state change when channels are non-full

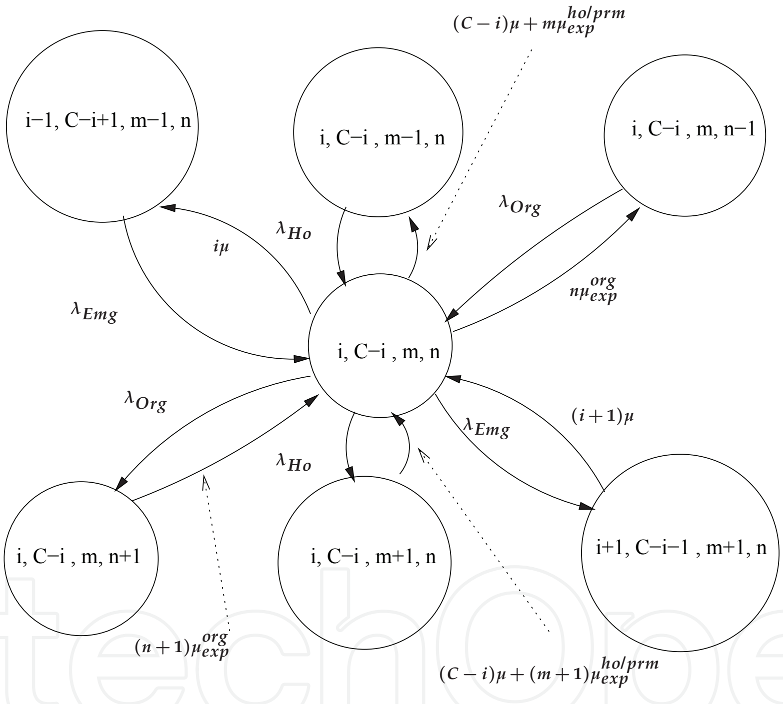


Fig. 5. The typical state change when channels are full and $i < R$

Note that when $i \geq R$, no preemption will be allowed, which will make the terms involving λ_{Emg} disappear.
With the practical consideration of expiration and preemption threshold, a product form solution for the equilibrium equations has not been found. Since we have limited the system to one buffer for handoff and preempted sessions, the computation requires operations on a matrix with size CL_1L_2, which means it depends on the number of channels and the size of the two buffers. Note that as pointed out in (Nyquetek, 2002), the buffer size need not be long (=5) because the effect will not be obvious after a certain point. Due to this fact, the computation is feasible.

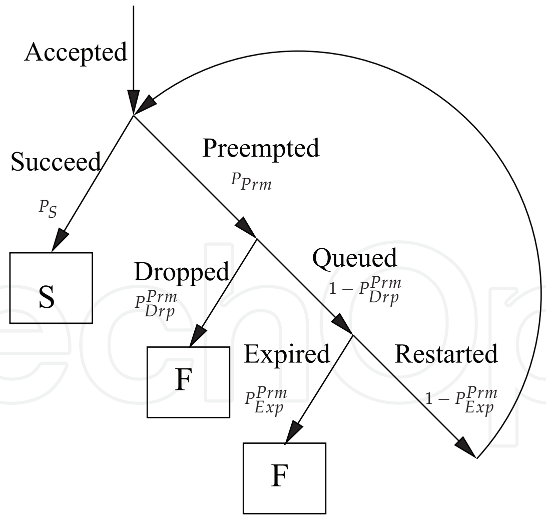


Fig. 6. Probability flow for low priority sessions

4.2.2 Performance evaluation

With the state probabilities solved, performance metrics, including average channel occupancy and the success probability, i.e., probability of finishing normally without expiring or dropping for each class can be obtained and will be shown in this subsection. Computation of related parameters, like admission probability, blocking probability of each class, the expiration probability of sessions in each queue, and preemption probability for a low priority session given that it is admitted, has been provided by (Zhou & Beard, 2006).

(1) System utilization and channel occupancy

Similar to the APS scheme, the system utilization can be computed by considering those portion of unused channels at all possible states:

$$SysUtil = 1 - \sum_{n=1}^{C-1}\sum_{i=0}^{n}\frac{(C-n)P(i,n-i,0,0)}{C} \tag{16}$$

The channel occupancies for emergency traffic and public traffic can also be computed based on steady states:

$$ChOcp^{Emg} = \textstyle\sum_{n=1}^{C}\sum_{i=1}^{n}\sum_{k=0}^{L_1}\sum_{l=0}^{L_2}\frac{iP(i,n-i,k,l)}{C} \tag{17}$$

$$ChOcp^{Pub} = \textstyle\sum_{n=1}^{C}\sum_{j=1}^{n}\sum_{k=0}^{L_1}\sum_{l=0}^{L_2}\frac{jP(n-j,j,k,l)}{C} \tag{18}$$

(2) Probability flow of low priority sessions

In Fig. 6, the probability flow of low priority sessions is shown. In the frame, "F" means failed, "S" means successful.

A session can be preempted multiple times, and with the renewal process assumption on resumed sessions, the number of preemption times will not affect the preemption probability of a session. Thus the number of preemption times is geometrically distributed with:

$$Pr(\text{Preempted } n \text{ times}) = P_{Prm}(1-A)A^{n-1}, n = 1, 2, ... \tag{19}$$

Here $A = P_{Prm}(1 - P_{Drp}^{Prm})(1 - P_{Exp}^{Prm})$ is the probability for a session to stay active; $(1 - A)$ is the probability that the session ends (succeeds, expires or is blocked after being preempted).

P_{Drp}^{Prm} is the probability for a preempted session to be dropped (due to a full queue) after being preempted, and P_{Exp}^{Prm} is the expiration probability for sessions waiting in the preempted session queue.

Thus the expected value of preempted times is $\frac{P_{Prm}}{1-A}$, or expressed in the form of preemption and expiration probability:

$$\overline{PrmTimes} = \frac{P_{Prm}}{1 - P_{Prm}(1 - P_{Exp}^{Prm})P_{Drp}^{Prm}} \tag{20}$$

(3) Success probability

For emergency sessions, all of the admitted sessions will be successfully finished, thus providing high dependability since they cannot be pre-empted. This kind of dependability cannot be assured for low priority sessions.

According to Fig. 6 we can compute the success probability given a session is admitted, which is denoted as P_{SGA}: for an admitted session, it will succeed only if it does not expire and is not blocked after being preempted. Note that $P_S = 1 - P_{Prm}$, we have:

$$\begin{aligned} P_{SGA} &= P_S \sum_{i=0}^{\infty} (P_{Prm}(1 - P_{Drp}^{Prm})(1 - P_{Exp}^{Prm}))^i \\ &= \frac{(1 - P_{Prm})}{1 - P_{Prm}(1 - P_{Drp}^{Prm})(1 - P_{Exp}^{Prm})} \end{aligned} \tag{21}$$

The successful finishing probabilities are decided by P_{SGA} and the corresponding admission probabilities :

$$P_{Succ}^{Ho} = P_{Adm}^{Ho} P_{SGA} \tag{22}$$

$$P_{Succ}^{Org} = P_{Adm}^{Org} P_{SGA} \tag{23}$$

5. Comparison of main admission control policies

In this Section, comparisons among the APS scheme (Zhou & Beard, 2010), the PURQ-AC scheme (Nyquetek, 2002), and the PTS scheme (Zhou & Beard, 2009) are provided. The main performance metrics considered include the achievable channel occupancy of public traffic, success probability, waiting time, and termination probability for each class of traffic.

The main parameters are: the number of channels in a cell is 50, and the average duration for each session is 100 seconds, so the maximum load that the system can process, called *system capacity*, is $C\mu = 0.5$ sessions/second $= 30$ sessions/minute. The load of public handoff traffic is 6 sessions/minute (20% of system capacity). Same as the parameters used in Nyquetek's report (Nyquetek, 2002), the average impatience times for handoff and originating traffic are both 5 seconds, and for emergency traffic it is 28 seconds, while the buffer sizes for all three queues are 5.

5.1 Comparison of achievable channel occupancy

As pointed out in (Nyquetek, 2002), the resource guarantee for public users is implemented through achieving at least 75% channel occupancy for public traffic. To evaluate whether all three schemes can achieve this goal, two different loads of emergency traffic are studied.

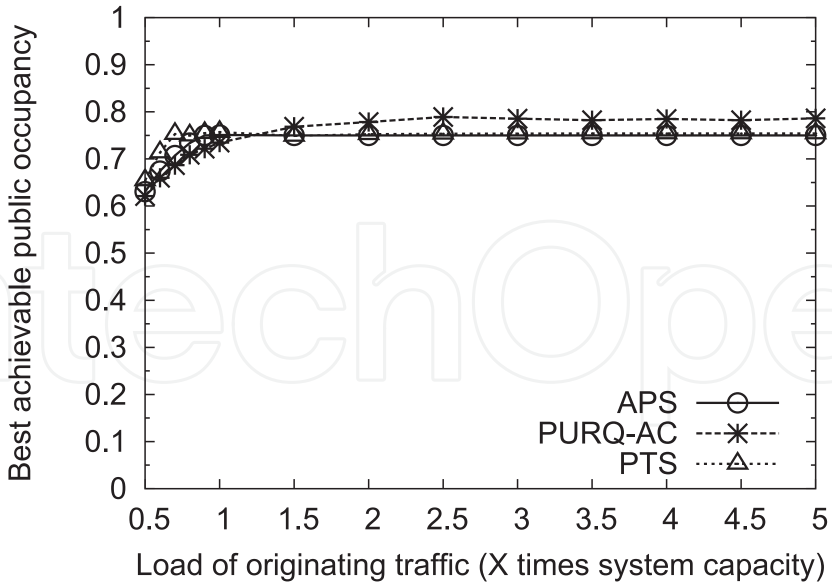


Fig. 7. Comparison of achievable channel occupancy for public traffic - Emergency traffic = 30% system capacity

When the load of emergency traffic is at 30% of the system capacity as shown in Fig. 7, all three schemes can guarantee at least 75% for public use if the load of the originating traffic is higher than the engineered system capacity. However, it can be seen that PURQ-AC will achieve even more than 75% when the public traffic is even higher. Although this protects the benefit of public traffic well, it does not achieve the guaranteed goal (25%) for emergency traffic.

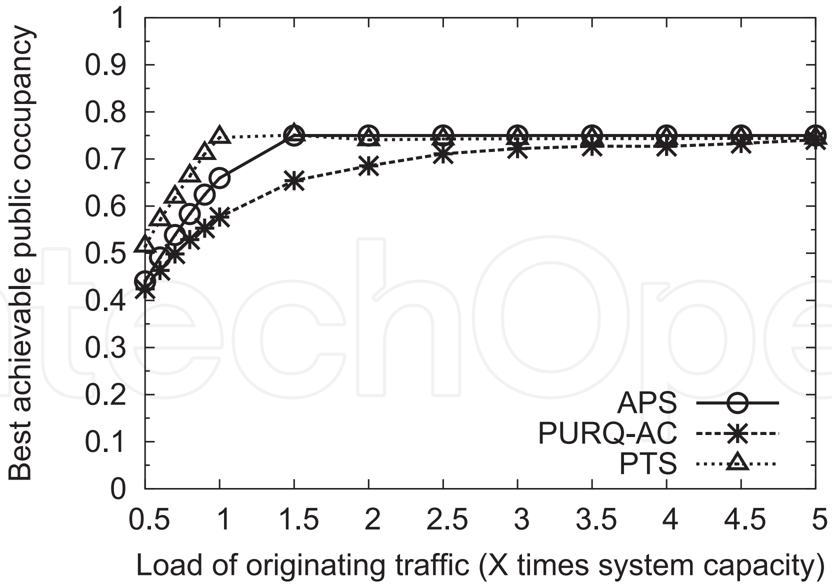


Fig. 8. Comparison of achievable channel occupancy for public traffic - Emergency traffic = 160% system capacity

Fig. 8 shows an extreme case: the emergency traffic is unexpectedly high at 160% of the system load. When the public traffic is not heavy enough, the PURQ-AC policy cannot guarantee 75% of channel occupancy for public traffic (also shown in Nyquetek (2002), Fig. 3-7). Actually, the

75% goal is not achieved until the originating traffic is 5 times of the system capacity. When the load of public originating traffic is at 100% of the system load, only 58% of channel resources are used for public traffic. In contrast, the APS scheme can achieve the goal with much lower load, at 1.5 times of the system capacity. If the originating traffic is even lower, the APS scheme can still guarantee higher channel occupancy for public traffic than what PURQ-AC can do. The PTS strategy is even better for lower load. In summary, *the APS scheme can can protect both public and emergency traffic more effectively than the PURQ-AC strategy*. The PTS scheme could be even better than the APS in this aspect.

It is worthy to note that Nyquetek also recommended a ”super count” scheme (similar to a leaky bucket scheme) to give the low load traffic better guarantees to be scheduled according to the 1/4 rule, which has been incorporated in the simulations that generate the results in Figs. 7 and 8. The main reason behind the above difference is that, with 1/4 scheduling, 75% of channel occupancy for public traffic is hard to be guaranteed with PURQ-AC because some factors, such as different impatience times for each class of customers, are not considered in (Nyquetek, 2002). For example, very short impatience times (5 seconds) in the originating traffic queue will cause a lot of customers to drop their sessions before a channel is available, thus leading to much smaller effective amounts of originating traffic to compete with emergency traffic. In contrast, the adaptive probabilistic scheduling and the preemption threshold methods are dynamic and can consider the effects of these factors and still achieve the desired channel occupancies.

5.2 Success probability for each class

As another main goal, admission of emergency traffic should be guaranteed when its volume is not unexpectedly high. To compare the effectiveness of the preemption threshold strategy and PURQ-AC in this aspect, the achieved success probability of emergency traffic and handoff traffic is shown in Fig. 9. It can be seen that the APS scheme is better than the PURQ-AC policy, and the PTS scheme is even better. APS is better than PURQ-AC because the emergency traffic can be given higher priority through a higher P_S value when the public traffic is comparatively high.

5.3 Waiting time

When the PTS scheme is applied, emergency traffic need not wait before being admitted. But for APS and PURQ-AC, emergency customers have to wait before being admitted. As shown in Fig. 10, the APS scheme can cause obviously lower waiting time than the PURQ-AC scheme when the emergency traffic load is moderate.

For the originating traffic, the waiting time in the APS scheme is almost the same as the PTS scheme, and is obviously lower than PURQ-AC (Fig. 11).

From the comparisons in Fig. 10-11 it can be concluded that the APS scheme is obviously better than the PURQ-AC method in terms of waiting time since it has lower access time for both emergency users and public users. The PTS scheme is again better than any of other two strategies.

5.4 Termination probability

As can be seen from the above comparisons, the PTS strategy is almost always the best in terms of achievable channel occupancy for public traffic, success probability for each class, and the waiting time. However, it has a significant downside: public sessions in this scheme can be

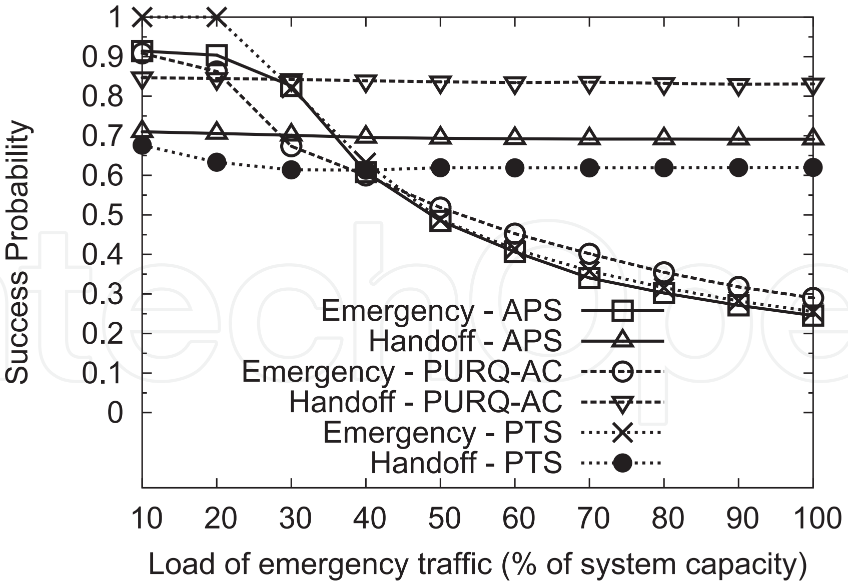


Fig. 9. Comparison of success probability

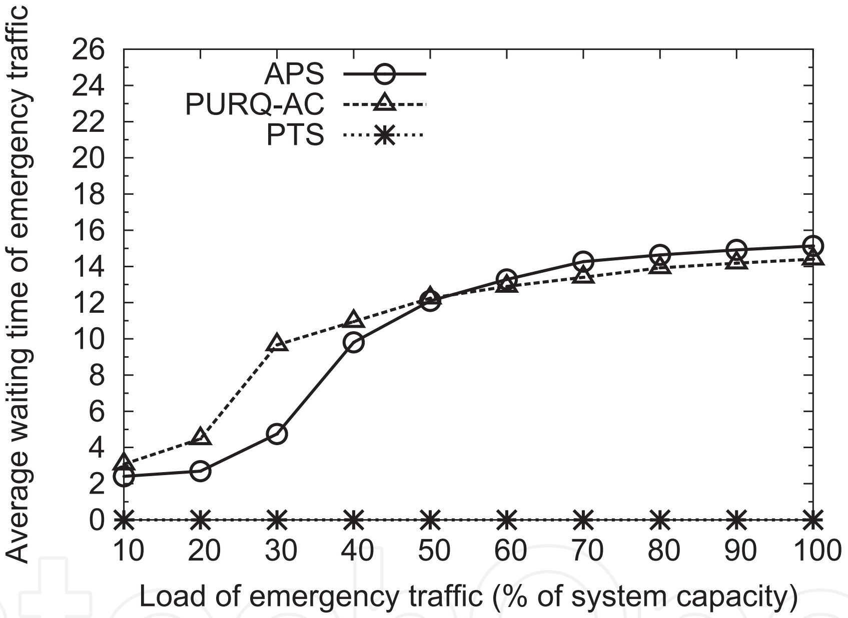


Fig. 10. Comparison of waiting time for emergency traffic

terminated due to preemption. In Fig. 12 the termination probability of public originating traffic is shown for the same scenarios considered above. For most load cases, the termination probability of public traffic for the PTS scheme is around 10%. In contrast, the queueing and scheduling based strategies APS and PURQ-AC obviously do not have such a problem. This shortcoming of the preemption based strategy is the main reason that it is not allowed in the current WPS in the United States, regardless of its other benefits.

6. Extension of the load based model to 3G/4G systems

The admission control policies discussed in this chapter are assumed to be load based. This means that admission is based on whether the new session will make the load surpass the

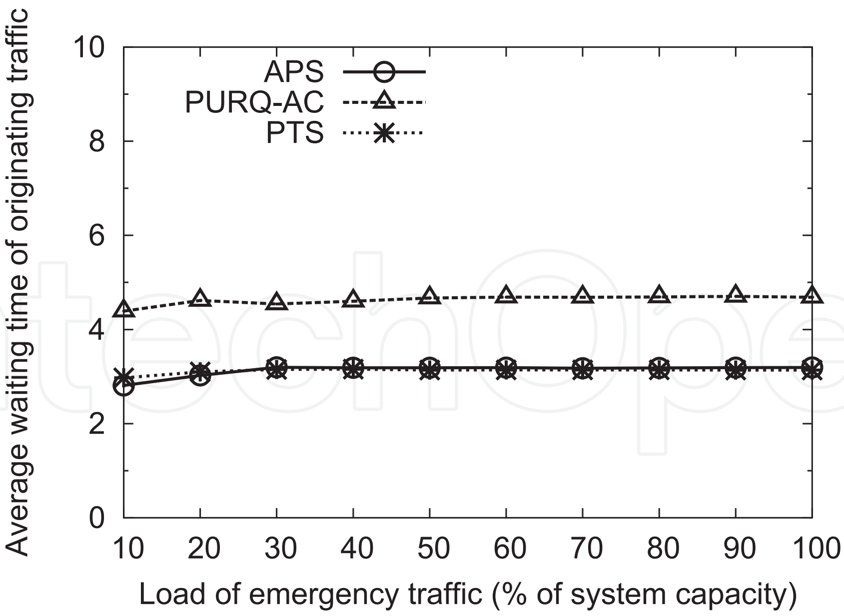


Fig. 11. Comparison of waiting time for originating traffic

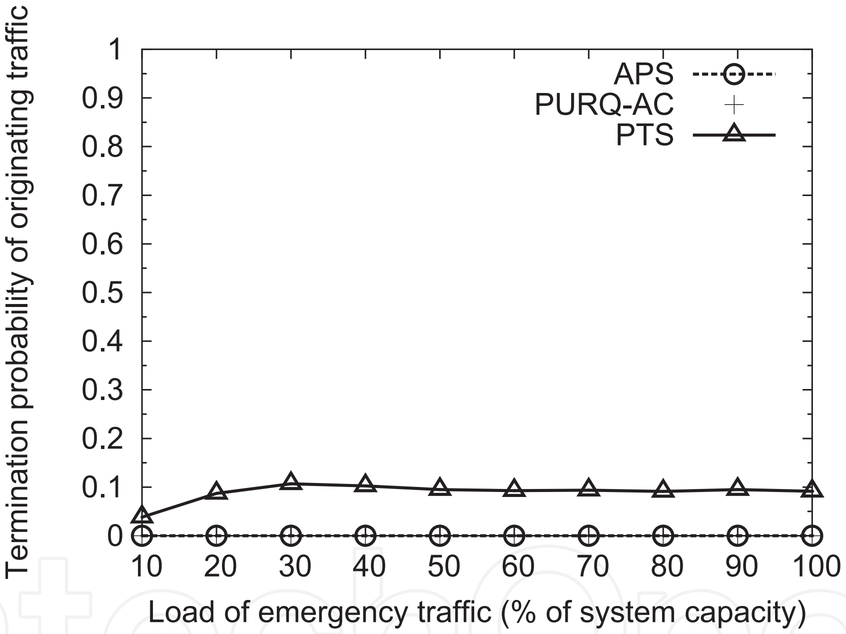


Fig. 12. Comparison of termination probabilities

capacity of the system. The load is usually measured by the number of users in a 2G system. With multiple access schemes like CDMA, WCDMA, OFDMA applied, one main difference is that interference rather than the number of users can be the main factor to be considered for the admission control problem in a 3G/4G system.

With a CDMA based access scheme, admission can be done indirectly by setting an interference-based criteria, for example a limit on CDMA Rise over Thermal (RoT), then determining ahead of time the load where a new session would cause the system to exceed the interference limit. In fact, as pointed out in (Ishikawa & Umeda, 1997), load based admission control is still suitable. In their analysis for what they call number-based CAC, the interference threshold is transferred into the maximum acceptable number of users. Then the blocking rate (measured grade of service) and the outage probability of communication quality (measured quality of service) are evaluated. The numerical results show that the number-based CAC and the interference-based CAC agree well with each other. They concluded that load-based admission is preferred because of its simplicity and ease of implementation, although interference based admission has the advantage that the threshold value has less sensitivity to other system parameters like the propagation model, traffic distribution, or the transmission rate.

As opposed to balancing blocking rate and outage probability of communication quality like in (Ishikawa & Umeda, 1997), we are mainly considering the fairness in resource use between emergency users and public users. When an emergency happens, there is much more demand than the system can handle. No matter how we try to balance capacity and quality of service, there is still blocking. So the capacity of the system, in terms of the maximum number of admitted users, can be determined according to the requirements on quality of service (QoS) only. With the capacity of the system known, the probabilistic scheduling can be tuned to achieve ideal channel occupancies for both emergency and public traffic. Note here that we assume the capacity is static for a period of time, but it can be recomputed if the SIR threshold needs to be changed, for instance, due to increased interference from neighboring cells or due to cell breathing to shift users to neighboring cells.

Another important difference is that data applications are much more common in a 3G/4G network. How would load based admission control be accomplished with both voice sessions and data sessions (emergency and public) in the same cell? Admission of voice sessions can easily be controlled based on whether a new session would go beyond the voice loading limit. Data sessions, however, can be handled in two distinctly different ways. On the one hand, if data sessions need some level of guaranteed QoS, they can be admitted similarly to voice sessions, by equating a data session to a certain number of voice sessions or a certain amount of needed bandwidth. On the other hand, service providers may treat data sessions differently by expecting them to use whatever is left over after the voice sessions are satisfied. For example, in the 3G EV-DO, Rev. A standard, "HiCap" data sessions are given different power levels and Hybrid ARQ termination targets as compared to "LoLat" voice traffic. HiCap data traffic is expected to be able to tolerate longer packet delays and to probably use TCP to adapt to the network congestion.

In conclusion, interference-based admission control can be converted into a load based admission control problem. Furthermore, the elastic property of data sessions makes it possible for us to use the same model as that of a 2G scenario. This is why we can conclude that the schemes and modeling methods shown in this chapter is suitable for all 2G, 3G, and 4G systems.

7. Summary

Due to the special requirements of emergency traffic, the reservation based admission control strategy is inappropriate due to its possible waste of resources when emergency situations do not occur. Among the schemes that can guarantee high system utilizations, the dynamic schemes like the preemption based PTS scheme and the queueing and scheduling based APS scheme demonstrate their privileges over the static schemes like PURQ-AC. The PTS scheme is almost always the best in the guaranteed protection of both public and emergency traffic and in much shorter access waiting time. However, its disadvantage is possible high termination probability. In contrast, the APS scheme is also quite good in protecting both public and emergency traffic, and it can still guarantee low termination probability for public sessions. The operators can choose the strategy that suits their specific needs.

8. References

Beard, C. (2005). Preemptive and Delay-Based Mechanisms to Provide Preference to Emergency Traffic, In *Computer Networks Journal*, Volume 47, Issue 6, pp. 801-824, April 2005.

Beard, C. & Frost, V. (2001). Prioritized Resource Allocation for Stressed Networks, In *IEEE/ACM Transactions on Networking*, Volume. 6, Issue 5,October 2001, pp. 618-633.

Carlberg, K., Brown, I., & Beard, C. (2005). Framework for Supporting Emergency Telecommunications Service (ETS) in IP Telephony, *Internet Engineering Task Force, Request for Comments 4190*, November 2005

Conway, R., Maxwell, W., & Miller L. (1967). Theory of Scheduling, published by *Addison Wesley*, 1967.

Federal Communications Commission Report and Order (2000). The Development of Operational, Technical and Spectrum Requirements for Meeting Federal, State and Local Public Safety Agency Communication Requirements Through the Year 2010, *FCC 00-242*, rel. July 13, 2000.

Guerin, R. (1988). Queueing-Blocking system with two arrival streams and guard channel, In *IEEE Transactions on Communications*, Volume 36, Issue 2, Feburary 1988, pp. 153-163.

Ishikawa, Y. & Umeda, N. (1997). Capacity Design and Performance of Call Admission Control in Cellular CDMA Systems, In *IEEE EEE Journal on Selected Areas in Communications*, Volume 15, Issue 8, 1997, pp.1627-1635

Jedrzycki, C. & Leung, V. (1996). Probability distribution of channel holding time in cellular telephony systems, In *Proc. IEEE Veh. Technol. Conf.*, May 1996, pp.247-251.

Mitchell, K., Sohraby, K., van de Liefvoort, A., & Place, J. (2001). Approximation Models of Wireless Cellular Networks Using Moment Matching, In *IEEE Journal of Selected Areas in Communications*, Volume 19, Issue 11, pp. 2177-2190, 2001.

National Communications System (2002). Wireless Priority Service (WPS) Industry Requirements for the Full Operating Capability (FOC) for GSM-Based Systems, *prepared by Telcordia*, Issue 1.0, Sept. 2002.

National Communications System (2003). Wireless Priority Service (WPS) Industry Requirements for the Full Operating Capability (FOC) for CDMA-Based Systems, *prepared by Telcordia*, Issue 1.0, Mar. 2003.

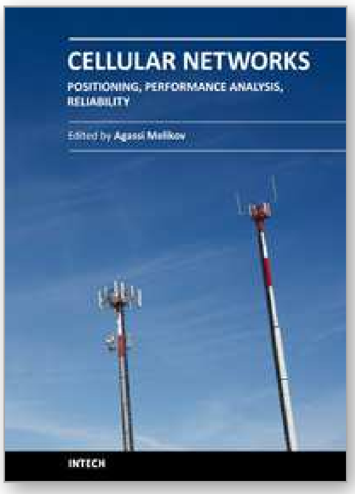
Nyquetek Inc. (2002). Wireless Priority Service for National Security / Emergency Preparedness: Algorithms for Public Use Reservation and Network Performance, August 30, 2002. Avaiable at http://wireless.fcc.gov/releases/da051650PublicUse.pdf.

Tang, S. & Li, W. (2006). An adaptive bandwidth allocation scheme with preemptive priority for integrated voice/data mobile networks, In *IEEE Transactions on Wireless Communications,* Volume 5, Issue 9, September 2006.

Zhou, J. & Beard, C. (2006). Comparison of Combined Preemption and Queuing Schemes for Admission Control in a Cellular Emergency Network, In *IEEE Wireless Communications and Networking Conference (WCNC)*, Las Vegas, NV, April 3-5, 2006.

Zhou, J. & Beard, C. (2009). A Controlled Preemption Scheme for Emergency Applications in Cellular Networks, In *IEEE Transactions on Vehicular Technology*, Volume 58, Issue 7, Sept. 2009 pp.3753-3764.

Zhou, J. & Beard, C. (2010). Balancing Competing Resource Allocation Demands in a Public Cellular Network that Supports Emergency Services, In *IEEE Journal on Selected Areas in Communications*, Volume 28, Issue 5, June. 2010.



Cellular Networks - Positioning, Performance Analysis, Reliability
Edited by Dr. Agassi Melikov

ISBN 978-953-307-246-3
Hard cover, 404 pages
Publisher InTech
Published online 26, April, 2011
Published in print edition April, 2011

Wireless cellular networks are an integral part of modern telecommunication systems. Today it is hard to imagine our life without the use of such networks. Nevertheless, the development, implementation and operation of these networks require engineers and scientists to address a number of interrelated problems. Among them are the problem of choosing the proper geometric shape and dimensions of cells based on geographical location, finding the optimal location of cell base station, selection the scheme dividing the total net bandwidth between its cells, organization of the handover of a call between cells, information security and network reliability, and many others. The book focuses on three types of problems from the above list - Positioning, Performance Analysis and Reliability. It contains three sections. The Section 1 is devoted to problems of Positioning and contains five chapters. The Section 2 contains eight Chapters which are devoted to quality of service (QoS) metrics analysis of wireless cellular networks. The Section 3 contains two Chapters and deal with reliability issues of wireless cellular networks. The book will be useful to researches in academia and industry and also to post-gradute students in telecommunication specialitiies.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following:

Jiazhen Zhou and Cory Beard (2011). Providing Emergency Services in Public Cellular Networks, Cellular Networks - Positioning, Performance Analysis, Reliability, Dr. Agassi Melikov (Ed.), ISBN: 978-953-307-246-3, InTech, Available from: http://www.intechopen.com/books/cellular-networks-positioning-performance-analysis-reliability/providing-emergency-services-in-public-cellular-networks

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821