Adaptive Probabilistic Scheduling for a Cellular Emergency Network *

Jiazhen Zhou and Cory Beard
Department of Computer Science Electrical Engineering
University of Missouri - Kansas City Kansas City, MO 64110

Abstract

Currently, wireless priority services (WPS) are provided using call queueing with a weighted round-robin schedule based on a fixed parameter. In this paper, we introduce a flexible, analytically-based scheme for dealing with admission control in a cellular network that supports emergency service. In the scheduling scheme, a priority queue is used to ensure the wise utilization of system resources while guaranteeing the priority of public handoff traffic. Also, probabilistic scheduling is considered for adaptively ensuring the high admission rate of emergency traffic and protecting public calls for different traffic demands. The admission of emergency traffic and public originating traffic can be flexibly adjusted. Numerical results about the evaluation of this strategy are presented and show an obvious improvement over related strategies.

Keywords: Admission control; emergency traffic; queuing; probabilistic scheduling; channel occupancy; handoff; expiration

1 INTRODUCTION

In a society where wireless voice communication capabilities are pervasive, emergency personnel should be able to use both government and commercially available systems to respond to natural and man-made disasters. To provide such services, a general sense of priority should be attached to emergency calls and to how they are given access to wireless resources. Such priority, however, cannot be absolute, since the needs of the general public are also very important to address.

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 call completion. This has resulted in definition of an end-to-end solution for national security and emergency preparedness (NS/EP) calls called the wireless priority service full operating capability (WPS FOC) [3], [4]. First-responders, NS/EP leadership, and key staff are able to use this capability by using allocated access codes.

This work focuses on the critical issue of tuning WPS call admission controls so that a proper relationship is enforced between admission of NS/EP traffic and public traffic. Both types of traffic have requirements for access to resources that should be balanced, especially keeping in mind that a reasonable amount of spectrum must remain available for public use even in an emergency because the public must also be able to respond to the emergency [2].

The key contribution of this work is to present an analytical modeling framework that creates flexible, adaptive algorithms operators can easily use. This work builds upon 10 years of experience in emergency communications research that has included development of call admission algorithms for wireline telephone networks [6], [7], queue management and packet scheduling policies for the general Internet [8], an Internet RFC [1], and traffic prioritization in wireless CSMA MAC protocols [12]. The work specifically related to this paper began with [7], where both delay-based and preemption-based approaches were introduced to provide "reasonable" priority for NS/EP traffic in wireless cellular networks. In [9], a weighted earliest deadline (EDD) based method was introduced to provide the flexibility of adjusting the priority between NS/EP traffic and public handoff traffic. Lower complexity methods were proposed to compute the loss probability and the average waiting time. Yet in that work the public originating traffic was not well protected when the emergency traffic was high. In [10] and [11], preemption was considered and a combined preemption and queueing method was provided to lessen the "harshness" of preemption policies so they could be adopted and tuned to practical operation in real environments.

The practical algorithms for supporting WPS services by the main cellular network operators were evaluated by Nyquetek [5] using a simulation study and a in-depth estima-

tion of traffic demands for particular cells and different traffic types (NS/EP, public, etc). The recommendations from the report have been used by operators. Our work seeks to use analytical methods for the same purpose, and the analytical methods are not dependent on the traffic load assumptions of [5], so control strategies can be adapted to actual measured traffic loads instead of general assumed traffic loads. In [5] a series of call queueing and scheduling based policies are introduced and compared for supporting emergency traffic in the commercialized wireless cellular network. In their work, they emphasized that the priority of emergency traffic can be guaranteed by queueing emergency calls when no channels are immediately available. They show that normally the emergency traffic will not be more than 10% of the normal engineered load of a cell, so the admission of public traffic will not be affected much. Furthermore, if the emergency demand is high enough that it can take much of the available resources in a cell, queueing of both public calls and emergency calls is proposed and a round-robin like scheduling policy is provided. Out of every four channels released, one will be used to serve emergency calls; in other words, a 1/4 weighted round robin schedule is proposed.

The main idea in Nyquetek's report is to guarantee that a portion of the resources will be available for public users while providing some priority for emergency traffic. But a problem with this scheme is that it is not flexible; the scheduling parameter of 1/4 is not adaptive to the possible changes of traffic volume for NS/EP traffic.

Their estimation of traffic load for NS/EP leadership and key staff at each cell is an average view, which may not be accurate in particular cases. For example, when a man-made or natural disaster happens, first-responders will commonly concentrate in certain locations, possibly making the NS/EP traffic at those cells very high. Certainly the 1/4 parameter could be adjusted by re-running Nyquetek's same simulations on-demand when traffic loads are at variance from assumptions, but this would be impractical. Many simulations would need to be run, each taking significant processing time, and a search procedure over several simulations would need to occur to find the best scheduling parameter. In this paper, we provide an analytical model that can easily be used to yield the same resulting choice of scheduling parameter, and with greater certainty since simulation results have inherent stochastic inaccuracies.

Also, in Nyquetek's report, user behavior is not considered. In their evaluation report, they just consider a timer in the equipment (at the base station) and assume that the users will be removed from the waiting queue after this fixed time limit ends. But in real life, when the users have waited too long, they will become impatient and drop the calls by themselves.

This behavior has been called reneging [13], [14]; the time before reneging is different for each user and is generally treated as exponentially distributed [14], [15]. Also consider the fact that handoff calls will move out of the handoff area if no channel is assigned in time in the new cell. We call the total behavior of leaving the waiting queue due to impatience as "expiration".

In this paper, we address the flexibility issue by providing an adaptive probabilistic scheduling strategy. When an in-progress call ends, a queued NS/EP call is chosen with a certain probability. Otherwise, a queued public call is put into service.

Also, instead of using a channel reservation scheme (also referred to as guard channels) for public handoff traffic, which was suggested in [5] and can cause low utilization of resources, we use a priority queue, where handoff calls are queued and given highest priority when a channel becomes available. This guarantees the success of handoff and accomplishes wise utilization of radio resources at the same time.

In Section II, the basic model is presented and the computation of main performance metrics is also shown. Section III provides the algorithms for assessing the handoff rate, determining the parameters for the probabilistic scheduling, and switching between two scheduling modes based on the amount of NS/EP traffic. The numerical results are shown in Section IV. Finally, Section V concludes this paper.

2 Basic Model

2.1 Basic scheme and assumptions

The main types of voice calls we deal with are emergency calls, public handoff calls and public originating calls. As shown in [5], 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. So handoff for emergency calls is not considered here, but our current work can be readily extended to deal with emergency handoff traffic when necessary.

Since the termination of calls will cause strong dissatisfaction for customers, dropping of handoff calls should be avoided as much as possible. Also, handoff traffic generally will not be high, so in our strategy we use priority queueing and give the handoff traffic the highest priority. In contrast to the resource reservation strategy used in [5] (use of guard channels), priority queueing is better at guaranteeing high system utilization because channels are not kept idle in an absence of handoff calls.

In [2], 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 the emergency traffic is under the certain "protection threshold", we should give high priority to emergency traffic; when the emergency traffic is extremely high so that it can take most or all of the radio resources, we should have a corresponding strategy 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 [5].

The basic scheme we are to use is illustrated in Fig. 1. In the figure, Class 1 is for public handoff calls, Class 2 is for emergency calls, and Class 3 is for public calls originating from within the cell. When an incoming call fails to find free channels, they will be put into individual queues until the queues become full. The calls in the handoff queue will take the channel once it become free. If there is no call in the handoff queue when a channel is freed, a call will be randomly chosen from either the emergency queue or public originating queue according to the scheduling probability already set. Note that when a disaster happens, it is uncommon for general people (who generate public traffic) to move into a disaster area. Thus the traffic handoff into a disaster area will not be high, so setting the public handoff traffic as the highest priority will not make emergency traffic starve. The scheduling probability for emergency traffic is denoted as P_s. The algorithm to decide P_s at different cases will be introduced in section 3. The queues are finite and customers can be impatient when waiting in the queue, so blocking and expiration are possible. We assume that the
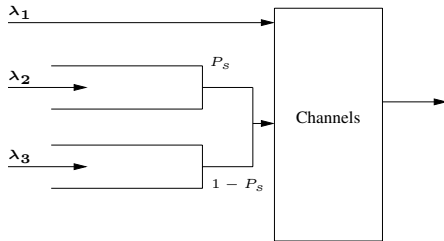

Figure 1. Probabilistic scheduling scheme

expiration times of all three classes of calls are exponentially distributed with different possible rates which can be denoted as μ_{exp1}, μ_{exp2} and μ_{exp3} separately. And similar to what's used widely and also assumed in [5], all call durations and inter-arrival times are independently, identically, and exponentially distributed. We can build a 3-D Markov chain to model the behavior of the three classes of traffic. This is shown in Fig. 2, in which the total number
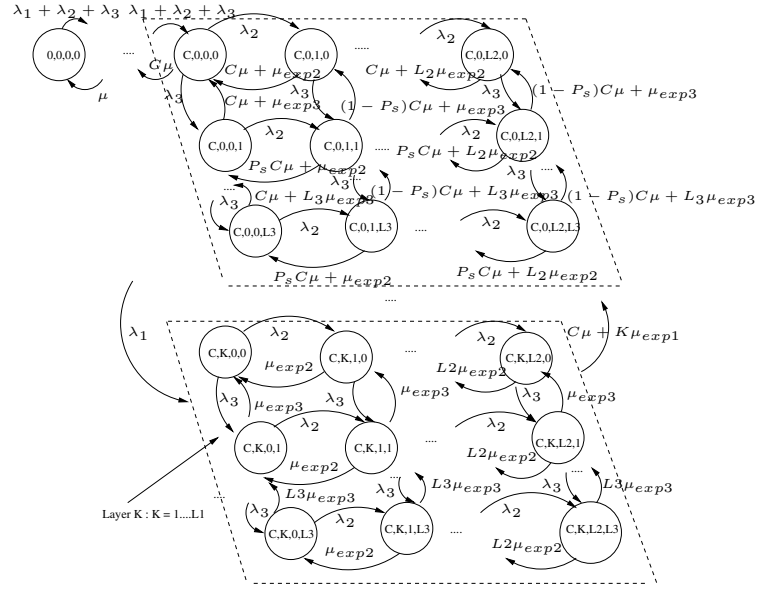

Figure 2. State Diagram for Probabilistic Scheduling Scheme

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 calls in handoff queue, emergency queue, and public originating queue respectively. The arrival rate for handoff, emergency and originating calls is λ_1, λ_2, λ_3, and the service rate for all calls is μ. We can see that the Markov chain in Fig. 2 does not have 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.

2.2 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 the system we introduced, the loss of each class consists of two parts: those calls that are blocked when the arrivals find the queue full; and those calls reneged (also called expired) when waiting too long in each queue. So we have: $P_{Loss,i} = P_{B,i} + P_{Exp,i}$ for i = 1,2,3. The admission probability can be calculated as: $P_{Adm,i} = 1 - P_{Loss,i}$.

(1) Blocking probability

Blocking for the public handoff and originating traffic happens when the corresponding queue is full. From Fig. 2

we can see that:

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

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

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

(2) Expiration Probability

Using the probability flow method in [9], we can compute the expiration probability for each class:

At a state (C, i, j, k), the arrival rate for each class is λ_1, λ_2, λ_3, and the expiration rate for each class is $i\mu_1$, $j\mu_2$, $k\mu_3$ independently. The average total number of departures per unit time from a state for expirations by class 1 is $i\mu_1$. Then, the probability of expiration is the fraction of expirations per unit time over arrivals per unit time, hence $\frac{i\mu_1}{\lambda_1}$. For class 2 calls this is $\frac{j\mu_2}{\lambda_2}$, and is $\frac{k\mu_3}{\lambda_3}$ for class 3. 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:

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

$$P_{Exp}^2 = \sum_{i=0}^{L_1} \sum_{j=1}^{L_2} \sum_{k=0}^{L_3} P(C, i, j, k) \frac{j\mu_{exp,2}}{\lambda 2} \quad (5)$$

$$P_{Exp}^3 = \sum_{i=0}^{L_1} \sum_{j=1}^{L_2} \sum_{k=0}^{L_3} P(C, i, j, k) \frac{k\mu_{exp,3}}{\lambda 3} \quad (6)$$

(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 not used, and the total portion of channels not used is thus $\frac{C-n}{C}$. So the system utilization can be calculated by considering those portion of channels not used at those possible states:

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

We define "channel occupancy" as the portion 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 call has the same average call duration, the channel occupancy of each class can be calculated by comparing the admitted traffic of each class, which is:

$$ChanOccp_i = \frac{\lambda_i(1 - P_{Loss,i})}{\sum_{i=1}^3 \lambda_i(1 - P_{Loss,i})} SysUtil \quad (8)$$

(4) Waiting time

Those calls waiting in the queue can be patient enough to wait until the next channel becomes available, or become impatient that it will leave the queue before being served. The average time staying in the queue (despite whether it is eventually served or not) can be calculated using Little's law: $\overline{T_i} = \frac{N_{q,i}}{\lambda_i(1-P_{B,i})}$. Note that the effective arrival rate at each queue will be $\lambda_i(1 - P_{B,i})$, and the average queue length for each class is $N_{q,i}$. The mean queue length for each class can be easily calculated based on the steady states we compute, so we have:

$$\overline{T_1} = \sum_{i=0}^{L_1} \sum_{j=0}^{L_2} \sum_{k=0}^{L_3} \frac{iP(C, i, j, k)}{\lambda_1(1 - P_{B,1})} \quad (9)$$

$$\overline{T_2} = \sum_{i=0}^{L_1} \sum_{j=0}^{L_2} \sum_{k=0}^{L_3} \frac{jP(C, i, j, k)}{\lambda_2(1 - P_{B,2})} \quad (10)$$

$$\overline{T_3} = \sum_{i=0}^{L_1} \sum_{j=0}^{L_2} \sum_{k=0}^{L_3} \frac{kP(C, i, j, k)}{\lambda_3(1 - P_{B,3})} \quad (11)$$

The blocking probability for each class can be calculated using equations (1) $\sim$ (3).

3 Adaptive Probabilistic Scheduling Algorithms

Denote the "system capacity" as the largest throughput of the cell. In other words, it is the total service rate of the cell, so it is $C\mu$. To facilitate analysis, we assume that the normalized system capacity is 1. So the "protection threshold" of emergency traffic - λ_{thr} will always be less than or equal to 1. The basic principle for the scheduling between emergency and public traffic is: when the emergency traffic is less than λ_{thr}, we need to guarantee at least P_{ad-em} part of the emergency calls to be admitted; when emergency traffic is higher than λ_{thr} (so it is over the "protection threshold"), we need to protect public calls (including handoff and originating calls) by guaranteeing their channel occupancy is at least $Util_{Pub}$. The parameters $\lambda_{thr}, P_{ad-em}, Util_{Pub}$ can be decided by the operators upon their respective requirements. However, those parameters are not independent of each other. On the contrary, they have dependence on each other which is restricted by the general knowledge that the system utilization should be less than 1: Since at least $Util_{Pub}$ part of radio resources should always be guaranteed for public use, to make sure this condition also holds when $\lambda_2 \leq \lambda_{thr}$, we should have:

$$\lambda_{thr}P_{ad-em} + Util_{Pub} \leq 1 \quad (12)$$

So the parameters should be carefully chosen so that equation (12) is always true. For instance, if we choose

$\lambda_{thr} = 0.3$, $P_{ad-em} = 0.9$, and $Util_{Pub} = 0.7$, then $\lambda_{thr}P_{ad-em} + Util_{Pub} = 0.97 < 1$.

In the following subsections, we introduce the ways for computing the accurate handoff rate, and the algorithms for deciding the weighting parameter for different load cases.

3.1 Computation of accurate arrival rate of handoff traffic

As mentioned in [16], based on the assumption of handoff in and handoff out in each cell are equal, the arrival rates of originating and handoff calls satisfy such condition:

$$\lambda_h = \lambda_o \frac{P_h(1 - P_L^{Org})}{1 - P_h(1 - P_L^{Ho})} \tag{13}$$

Where P_h is the probability that an ongoing call will handoff, λ_h and λ_o are the rates for handoff calls and originating calls individually, corresponding to λ_1 and λ_3 in our above denotation. P_L^{Ho} and P_L^{Org} is the failure probability for public originating calls and public handoff calls in a cell individually, corresponding to P_L^1 and P_L^3 in our denotation.

In our scheme, the handoff traffic is treated as the highest priority, and the traffic volume is not large, which makes the loss rather small, so equation (13) can be approximated as:

$$\lambda_h = \lambda_o \frac{P_h(1 - P_L^{Org})}{1 - P_h} \tag{14}$$

Given P_h is fixed, the ratio of originating calls to handoff calls are decided by the loss probability of handoff and originating calls. From above sections we can see that the loss probability is decided by the arrival rates. Obviously this is a recursive problem, so we will use an iterative algorithm to decide the accurate arrival rate of handoff calls. The loss probability of each class can be expressed in terms of arrival rate of each class after solving the Markov chain shown in Fig. 2. In each iteration, we just need to replace the value of loss probability for each class by the new values.

3.2 Algorithm for deciding scheduling weight parameter

The algorithm to find the scheduling probability for emergency traffic P_s is:

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

Step 2: Solving the Markov chain, get the general representation of loss probabilities with the method we provided in section 2.2. With the parameters we have, we can decide the admission probabilities for each class. If the admission probability of emergency traffic is less than P_{ad-em}, it means that we cannot increase the admission probability

by adjusting the weighting parameter, so we cannot satisfy the admission requirement for emergency traffic, stop here. Otherwise go to step 3 to search the suitable weighting parameter.

Step 3: Use a binary search method to search for the best weighting parameter: Let $P_s = 1/2$, using the representation of loss probabilities we described in step 2, then find the loss probability for emergency traffic. 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 this step until the suitable P_s that meets the requirement of emergency traffic is found.

When $\lambda_2 > \lambda_{thr}$, we will calculate the channel occupancy of each class using equation (8). And the algorithm for deciding scheduling probability for emergency traffic is basically the same as the algorithm above. The only difference is that we will compare the channel occupancy of public traffic instead of loss probability.

4 Numerical results

As an example, in the numerical results that follow, we set $\lambda_{thr} = 0.3$, $P_{ad-em} = 0.9$, and $Util_{Pub} = 0.7$. The handoff probability is $P_h = 0.16$. There are 100 channels in the cell, and the average service time is 100 seconds. So the capacity of this system is $C\mu = 1$. The buffer size for each queue is 5, and the average expiration time for handoff, emergency and originating traffic is 10, 20, and 20 seconds respectively.

In Fig. 3, we show scheduling probability versus the arrival rate of public originating traffic when the arrival rate of emergency traffic is fixed at 0.3. Since emergency traffic is lower than λ_{thr}, in this case we try to guarantee 90% of emergency traffic to be admitted. We can see that the scheduling probability for emergency traffic will increase as the public traffic increases to cancel the effect of a large amount of public originating traffic. But the change is less and less significant as the public originating traffic keeps increasing; this is because the traffic out of the originating queue will not change much after blocking and expirations. Since the emergency traffic rate is fixed and the admitted portion of emergency traffic almost does not change (being 90%), the channel occupancy for emergency traffic stays at 27% of the system capacity. Note that the system utilization increases when the total system load increases; that's why the channel occupancy for public traffic also increases, but is always less than 73%. It is also always greater than the 70% requirement, because λ_{thr} was chosen properly. This is shown in Fig. 4.

Fig.5 illustrates the corresponding change in waiting time for each class. The handoff traffic has the least waiting time due to its high priority, and the public originating
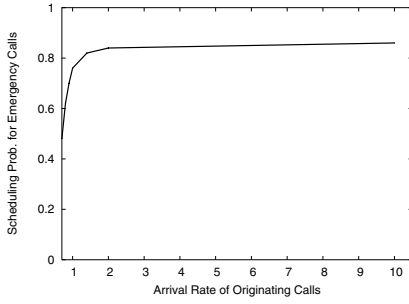
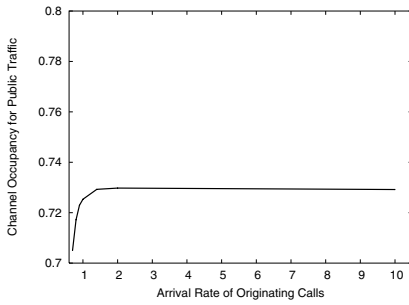**Figure 3. Scheduling prob. for emergency traffic when $\lambda_2 = 0.3$**

Figure 4. Channel occupancy for public traffic when $\lambda_2 = 0.3$

traffic has the longest waiting time, but also is less than 6 seconds despite heavy public originating traffic loads, and this is acceptable for customers. Also, we can see that as the public originating traffic increases, the originating calls need to wait longer and longer, while handoff calls and emergency calls are almost unaffected.
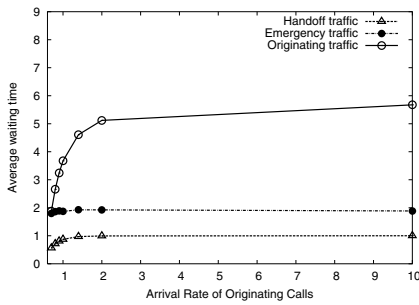

Figure 5. Average waiting time for each class when $\lambda_2 = 0.3$

Now consider the case when emergency traffic is fixed at 0.5, which is higher than λ_{thr}. We need to guarantee at least 70% of system resources are used by the public traffic. In

Fig. 6, we show the change of scheduling probability with respect to the public originating traffic; the trend is similar to what we have shown in Fig. 3. Since 70% of the system resources are used by the public traffic, the part that can be used by emergency traffic is less than 30%, so the admission probability for it will stay a little bit less than 60% (since 60% of 0.5 is 0.3). The corresponding plot for the waiting time is shown in Fig. 7. At this case the waiting time for emergency traffic basically does not change when the public originating traffic keeps increasing, but the value is much higher than the 2 seconds for the case when $\lambda_2 = 0.3$.
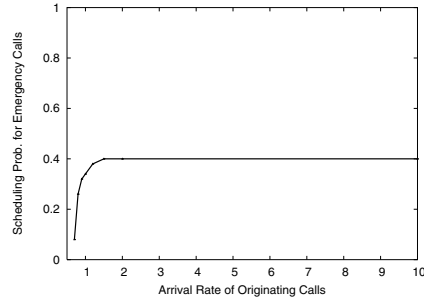

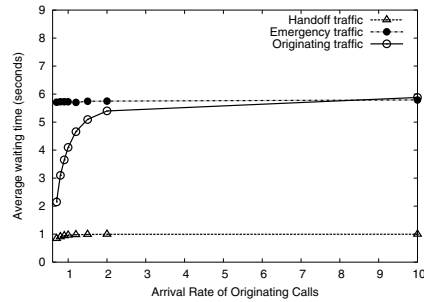**Figure 6. Scheduling prob. for emergency traffic when $\lambda_2 = 0.5$**

Figure 7. Average waiting time for each class when $\lambda_2 = 0.5$

In Fig. 8, we show the change of scheduling probability now when emergency traffic changes and when the public originating traffic is fixed at 2. We can see that the scheduling probability for emergency traffic is lower and lower as the emergency traffic increases to assure that at least 70% of the system resources are used for public traffic. As shown in Fig. 9, the average waiting time of handoff and originating calls will not be affected by the increase of emergency traffic because only a certain amount of emergency traffic (about 0.3) is admitted. But the delay in the queue for emergency traffic will become higher as the emergency traffic increases. The admission probability for emergency

traffic will be inversely proportional to its traffic rate since the admitted emergency traffic is almost fixed.
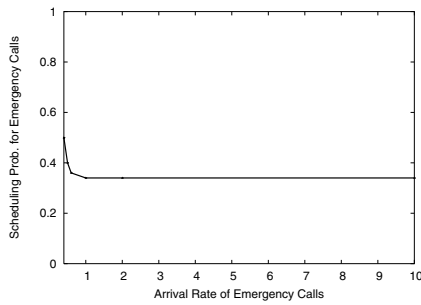

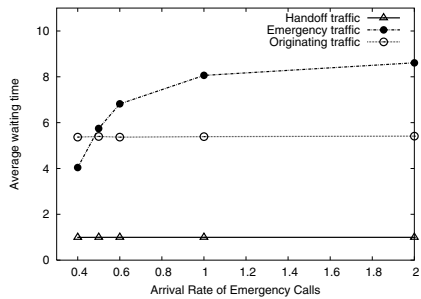**Figure 8. Scheduling prob. for emergency traffic when $\lambda_3 = 2$**

Figure 9. Average waiting time at each queue when $\lambda_3 = 2$

5 CONCLUSIONS

In this paper, an adaptive probabilistic scheduling strategy is introduced for dealing with admission control in a cellular network that supports emergency service. The scheduling probability can be dynamically adjusted with regard to the arrival rates of different classes of traffic and the requirement of operators. Correspondingly, the admission of emergency traffic and public originating traffic can be flexibly adjusted instead of being fixed as in [5]. Numerical results about the evaluation of this strategy are presented; the high flexibility to accommodate different requirements are shown; and the waiting times for each class are in reasonable ranges.

One possible future work would be to study the conditional average waiting time for those customers given that they are eventually served or dropped. Also, the combined expiration behaviors at both the backbone network and access network, and the corresponding theoretical analysis for tandem expiration and parallel expiration are worthy of studying.

References

[1] Ken Carlberg, Ian Brown, Cory Beard, "Framework for Supporting Emergency Telecommunications Service (ETS) in IP Telephony," *Internet Engineering Task Force, Request for Comments 4190*, November 2005

[2] Federal Communications Commission Report and Order, "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.

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

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

[5] Nyquetek Inc., "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.

[6] C. Beard and V. Frost, "Prioritized Resource Allocation for Stressed Networks," *IEEE/ACM Transactions on Networking*, Vol. 6, no. 5, October 2001, pp. 618-633.

[7] C. Beard, "Preemptive and Delay-Based Mechanisms to Provide Preference to Emergency Traffic." *Computer Networks Journal*, Vol. 47:6, pp. 801-824, 22 April 2005.

[8] M. Joshi, A. Mansata, S. Talauliker, and C. Beard, "Design and Analysis of Multi-Level Active Queue Management Mechanisms for Emergency Traffic," *Computer Communications Journal,* Volume 28, Issue 2, February 2005, pp. 162-173.

[9] J. Zhou and C. Beard, "Weighted Earliest Deadline Scheduling and Its Analytical Solution for Admission Control in a Wireless Emergency Network," *International Teletraffic Conference(ITC19)*, August 29-September 2, 2005, Beijing.

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

[11] J. Zhou and C. Beard, "Tunable Preemption Controls for a Cellular Emergency Network," *IEEE Wireless Communications and Networking Conference (WCNC) 2006*, Hong Kong, March 11-15, 2007.

[12] Z. Shi, C. Beard, and K. Mitchell, "Misbehavior and MAC Friendliness in CSMA Networks," *IEEE Wireless Communications and Networking Conference (WCNC)* 2007, Hong Kong, March 11-15, 2007.

[13] C. J. Ancker, Jr. and A. V. Gafarian, "Some queueing problems with balking and reneging: I", *Operations Research* Vol.11 (1963) pp. 88-100

[14] C.-J. Chang, T.-T. Su, and Y.-Y. Chiang "Analysis of a cutoff priority cellular radio system with finite queueing and reneging/dropping," *IEEE/ACM Transactions on Networking.* vol. 2, no. 2, pp. 166-175, Apr. 1994.

[15] A.E. Xhafa and O.K. Tonguz, "Dynamic priority queuing of handover calls in wireless networks: a analytical framework," *IEEE Journal on Selected Areas in Communications*, Vol.22, Iss.5, pp. 904-916, June 2004

[16] 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 Trans. on Vehicular Technology*, vol. VT-35, no. 3, pp. 77-92, Aug. 1986.