Performance Evaluation of Public Access Mobile Radio (PAMR) Systems with Priority Calls

Francisco Barceló, Josep Paradells
Dept. de Matemàtica Aplicada i Telemàtica (Unicersitat Politècnica de Catalunya)
c/ Jordi Girona 1-3, Campus Nord (Mod C3), 08034 Barcelona, SPAIN
barcelo@mat.upc.es

ABSTRACT

This paper presents a procedure for estimating the mean access delay of calls in Public Access Mobile Radio (PAMR) trunking systems. For this purpose, an approximation of the mean waiting time in the M/G/C queue with non-preemptive priority is developed and numerically tested. The proposed approximation is based on other existing approximations and has been tested for the conditions commonly found in PAMR systems with very accurate results.

I. INTRODUCTION

A. PAMR Systems

During the last decade we have witnessed a great development of mobile telecommunications systems. Private Mobile Radio (PMR) systems are oriented towards the professional market. Targeted customers are users of fleets of vehicles (transport companies, police, ambulance, fire, taxi, etc.). PMR trunked systems such as digital TETRA have replaced traditional (non-trunked) systems in Europe. PAMR are PMR systems in which the operator provides public access. With PAMR, small companies have all the advantage of the newest technology without having to afford the large investment of a private network.

The basic idea of a trunked PMR or PAMR system is the automatic sharing of a group of communication channels among a large group of users [1]. In a non-trunked system each fleet uses a dedicated channel. In a trunked system, any Mobile Station (MS) can be automatically switched to any idle channel. The consequence of grouping traffic is an improvement in the GoS or a reduction of the number of necessary channels to achieve the same GoS, i.e. delay probability and average access delay

Along this paper, the average access delay is taken as the goal performance metric of the GoS. This magnitude is easily understood outside the field of telecommunications and the customer gets a clear idea of its meaning. When designing actual systems, more specialised measures of the GoS, such as the proportion of calls delayed longer than a certain time must be guaranteed as recommended by the ITU-R [2].

Although advanced PAMR systems such as TETRA allow the hand-off of the call between adjacent cells, the hand-off feature is not considered in this paper. Hand-offs in PAMR systems are found in percentages much smaller than in PCS due to the following reasons:

- PAMR cells use to be larger than in public systems.
- PAMR calls use to be shorter: 20 seconds average in front of 120 of PCS.

B. Priority in PAMR Systems

Priority features are needed to enable important calls (from ambulances, police, fleet managers, etc.) to be subject to low delays when the system is near congestion. This advantage is obtained at the cost of non-priority calls which are subject to a higher average delay. The priority considered in this paper is of the non-preemptive type or Head of the Line (HOL): priority calls are placed in queue before all non-priority calls and never interrupt a call in progress. In systems that use preemptive priority, the preemptive call interrupts a call if there is no an available channel, thereby causing a highly annoying quality degradation on the interrupted call. In the case of preemptive priority, priority calls are never delayed so the delay probability for the priority calls is zero.

It is a common use to assign the pre-emptive feature in PAMR systems only to "emergency calls". Non-preemptive priority calls are often called "priority calls" in front of the "emergency" ones. Non-preemptive priority calls should be assigned to urgent calls that do not represent an emergency situation. The percentage of emergency calls in a PAMR system should be kept to a low level due to the high degradation that they cause to the rest of calls.

C. The System Model

A typical PAMR system is showed in Figure 1. This model can work on a stand-alone basis or connected to other switches in a cellular fashion. Unlike public cellular systems where calls are lost when blocked, in PAMR systems blocked calls are queued until a channel is released. The queueing model commonly used to evaluate PAMR systems is the well-known M/M/C queueing model [1, 3]. The infinite waiting room hypothesis is reasonable since loses (i.e. due to buffer full) should be kept to a very low level for reasonable performance. However, the M/M/C model assumes that the call holding time (CHT) follows a negative exponential distribution (n.e.d.), a situation that seldom occurs in actual PAMR systems, where the squared coefficient of variation (SCV) of the CHT is lower than one. In [5, 6] field studies for two PAMR systems report measured SCV's of the CHT between 0.5 and 0.35. The actual distribution of the CHT in PAMR systems is not as spread out as the n.e.d. assumed by the M/M/C model. Therefore the use of the M/M/C model overestimates the access delay and leads to design with system oversize (i.e. more radio channels than needed are calculated) with an unnecessary waste of radio spectrum.

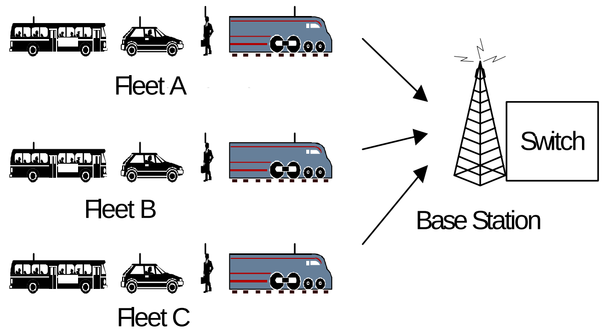

Fig. 1. A PAMR system: fleets and base station

D. *Objective*

The main contribution of this work is an approximation to the average waiting time in the M/G/C queue with priority in the form of a closed formula that is easy to compute. The result is an extension of [7] where a similar approximation for the M/D/C queue is presented. The approximation here studied suits very well the design conditions of PAMR systems. The conditions under which PAMR systems use to be evaluated are:

- Heavy traffic: When the system is not under heavy traffic, the mean access delay should be low for all calls.
- Low priority proportion: This is necessary to keep the priority strategy useful. A queue in which everybody has priority is a queue where nobody has it.

Approximations, when they are simple and accurate, are very useful in models for which exact analytical results have not yet been attained (perhaps they will never be). In teletraffic engineering design, the offered traffic is usually known or predicted within certain degree of error. This is certainly the case of PAMR systems. The 'exact result' has in this case an error of at least the same order of magnitude as the input data.

II. PAMR WITH PRIORITY AND THE M/M/C

In case of non-preemptive priority, the delay probability (*PD*) is the same for calls belonging to all priority levels. This is because the priority feature only changes the order in which calls are queued, but not the probability of finding all channels busy. In case of n.e.d. CHT, *PD* is calculated through the Erlang-C formula.

In [4], Gross and Harris give the exact formulae of *PD* and average waiting time in an M/M/C queue for the calls of each priority level assuming a generic number of priority levels. For simplicity's sake we consider in this work only two types of priority calls: priority (index 1) and non-priority or regular calls (index 2):

$$\overline{W_1}^M = \frac{PD \times d}{C(1-pr)} \quad \overline{W_2}^M = \frac{PD \times d}{C(1-pr)(1-r)}. \tag{1}$$

In Eq. (1), W_1^M and W_2^M stand for the waiting times for priority and regular calls respectively, assuming n.e.d. CHT, *PD* must be calculated according to the Erlang-C formula – $PD = \mathrm{EC}(r, C)$ –, d is the average CHT, r is the channel load and p is the priority proportion. The load due to priority calls only is then $p \times r$. It can easily be checked that the following relation holds for the mean waiting time averaged for all calls:

$$\overline{W}^M = p\overline{W}_1^M + (1-p)\overline{W}_2^M . \tag{2}$$

The priority gain is a convenient ratio. It is the quotient between the mean waiting time for all calls (as if there was no priority) and the mean waiting time for priority calls. In case of n.e.d. CHT this ratio is:

$$PG^M = \frac{\overline{W}^M}{\overline{W}_1^M} = \frac{1-pr}{1-r}. \tag{3}$$

It can be seen from Eq. (3) that the priority gain is higher for usual PAMR conditions of heavy load and low priority percentage. High priority gain is obtained when it is more necessary (congestion) and no inappropriate use is done (low priority percentage).

III. THE M/D/C QUEUE WITH NO PRIORITY

In this section we consider the M/D/C queue with no priority as an intermediate step. Among the different ways of calculating the mean waiting time in a M/D/C queue, we reject the exact computation [8] because it is highly complex and high precision is not required in the environment considered. A similar reason leads us to reject tabulated results [9], which are not available for light loads such as the load caused by considering only priority calls. In [8] we find the approximations most suitable for being used in the conditions found in a PAMR mobile radio system.

The approximation due to Cosmetatos for the M/D/C queue is excellent for medium and high loads:

$$R^D = \frac{\overline{W}^D}{\overline{W}^M} = \frac{1}{2}\left\{1 + \frac{(1-r)(C-1)\left(\sqrt{4+5C}-2\right)}{16rC}\right\}, \tag{4}$$

where W^D and W^M represent the waiting time for the deterministic and the equivalent exponential case (i.e. with the same load and number of channels) respectively. The ratio R^D represents the relative mean waiting time (i.e. relative to the n.e.d. holding time distribution case), very convenient for simplifying the notation in the rest of the paper. This approximation gives exact results for the single server case and for the asymptotic case when $r \rightarrow 1$, but it is inconsistent for light loads: if r is low or C is large, the waiting time appears to be longer for the deterministic than for the exponential call duration distribution. In reality, it must be shorter [10].

For light loads we find [8] an approximation for the M/G/C queue, which after the necessary algebra gives for the M/D/C:

$$R^D = \frac{\overline{W}^D}{\overline{W}^M} = \frac{(1-r)C}{C+1} + \frac{r}{2}. \quad (5)$$

The main advantage of this approximation compared to the others is that it does not depend on $\overline{W}^D$ (but only on $\overline{W}^M$), so one can use this approximation to find $\overline{W}^D$. The result obtained is not as accurate as Eq. (4) in general, but it continues to be valid for light loads, and gives the exact asymptotic result (notice that p should be kept low):

$$r \to 0 \qquad \Rightarrow \qquad \overline{W}^D = \frac{C}{C+1}\overline{W}^M. \quad (6)$$

In Figure 2 the ratio R_D calculated according to Eq. (4) (observe the inconsistency for light load) and Eq. (5) is plotted versus the load for a system with 30 channels.

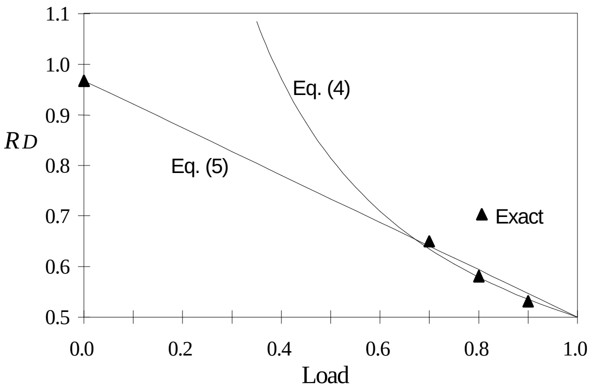

Fig. 2. Approximations of Eq. (4) and (5) for the relative mean waiting time in the M/D/30 queue

IV. INCLUSION OF PRIORITY CALLS

In this section the steps for calculating the average delay for priority and non-priority calls with D distributed holding time are proposed as an adequate combination of the results presented in Sections II and III. After running simulations and numerically testing several options, the best way to estimate the mean waiting time is to follow the steps given below:

Step 1: The mean waiting time for priority calls $\overline{W}_1^D$ should be estimated by combining Eq. (1) and Eq. (5), this latter due to the fact that priority calls offer a light load, to obtain:

$$\overline{W}_1^D = \overline{W}_1^M \left(\frac{(1-pr)C}{C+1} + \frac{pr}{2} \right). \quad (7)$$

Step 2: The mean waiting time for all calls $\overline{W}^D$ should be obtained by Eq. (4) which is very accurate, as r is not low when considering all calls in the system.

Step 3: The mean waiting time for calls with no priority $\overline{W}_2^D$ can be obtained from the following equation which is similar to Eq. (2). Note that this equation must hold for any CHT distribution:

$$\overline{W}^D = p\overline{W}_1^D + (1-p)\overline{W}_2^D. \quad (8)$$

In this Equation (8) the only unknown is the mean waiting time for non-priority calls $\overline{W}_2^D$.

After some algebra the priority gain for the M/D/C queue can be approximated by:

$$PG^D = \frac{\overline{W}^D}{\overline{W}_1^D} = \frac{1-pr}{1-r} \times \frac{R^D}{R_1^D} = PG^M \frac{R^D}{R_1^D}, \quad (9)$$

where R_1^D is calculated by Eq. (7) and R^D by Eq. (4). This priority gain is always lower than the priority gain for the exponential holding time PG^M, as R^D decreases with an increasing load [10] and the load offered by the calls with priority is lower than the total load.

This method has been tested under different conditions of load and number of channels, giving very accurate results in all cases (see [7] for details and numerical examples).

V. CALL DURATION 'G' DISTRIBUTED

A. Upper and Lower Bounds

The approximate mean access delay calculated by computing the approximation of the above section is very accurate, but the undertaken model assumes that call duration is distributed according to a D distribution. This will seldom occur in a true PAMR system. As the D distribution is the one with smaller SCV (cv^2=0) the result of the method proposed in Section IV is in fact a lower bound for the mean access delay in PAMR.

Some field studies exist on the statistical distribution of CHT in mobile radio systems. In [11] different PMR systems are measured, obtaining SCV's of call duration lower and higher than one, depending on the fleet. The work considers traditional PMR systems with one radio channel per fleet. When mixing all calls in a trunked PAMR system, the SCV becomes lower than one as reported in other field studies [5, 6]. In the PAMR case the result of applying the M/M/C model as in Section II, gives a higher bound for the mean access delay.

We know the upper and lower bounds for the mean access delay in PAMR systems in which the SCV of CHT is lower than one. The lower bound is not very useful for engineering purposes where one must guarantee something. The upper bound leads to system oversize. In this section we extend the results of the previous sections to accurately approximate the mean access delay once the SCV of the CHT is known to be lower than one.

B. Approximation for G Distributed Call Duration

Different approximations for the average queuing time in the M/G/C queue with no priority exist [8, 10]. There are reasons that induce us to select the Kimura approximation [10] as the most suitable for the PAMR environment. This approximation is a closed formula, which is very easy to compute and is more accurate than the linear approximation that leads to Eq. (5). The drawback is that we need the mean waiting time for the deterministic case (that is why we could not use it to calculate the waiting time for the M/D/C queue in Section III).

The procedure proposed in this paper to calculate the approximate mean waiting time in the M/G/C with two levels of priority consists of the following steps:

Step 1: The mean waiting time for the priority calls with general (G) holding time $\overline{W}1^G$ should be calculated by using the Kimura approximation:

$$\overline{W}1^G = \frac{1+cv^2}{\dfrac{2cv^2}{\overline{W}1^M} + \dfrac{1-cv^2}{\overline{W}1^D}}, \tag{10}$$

where cv represents the coefficient of variation of the CHT. The value of $\overline{W}_1{}^D$ must be approximated using Eq. (5) and $\overline{W}1^M$ must be computed from Eq. (1).

Step 2: The mean waiting time for all calls $\overline{W}^G$ should also be obtained by using the same Kimura approximation as follows:

$$\overline{W}^G = \frac{1+cv^2}{\dfrac{2cv^2}{\overline{W}^M} + \dfrac{1-cv^2}{\overline{W}^D}}, \tag{11}$$

where $\overline{W}^D$ must be calculated according to Eq. (4).

Step 3: The mean waiting time for calls with no priority $\overline{W}_2{}^G$ can be obtained from the equation:

$$\overline{W}^G = p\overline{W}1^G + (1-p)\overline{W}_2{}^G \tag{12}$$

where the only unknown is the mean waiting time for non-priority calls $\overline{W}_2{}^G$.

The priority gain can now be calculated approximately by manipulating the above equations:

$$PG^G = PG^D \, \frac{2cv^2 R1^D + 1 - cv^2}{2cv^2 R^D + 1 - cv^2} \tag{13}$$

As stated before, the load offered by priority calls is only a part of the total, and so, due to the increasing property of R^D when decreasing the load, $0.5 < R^D < R1^D < 1$. Then by substituting this inequality in Eq. (13) and comparing with Eq. (3) and (9) we get:

$$PG^D < PG^G < PG^M \tag{14}$$

Eq. (14) leads to the conclusion that the higher the SCV of the CHT the higher the priority gain. In other words, the advantage of using the priority feature in a PAMR system is greater when the SCV of the call holding time tends to one. This property is displayed in Figure 3.

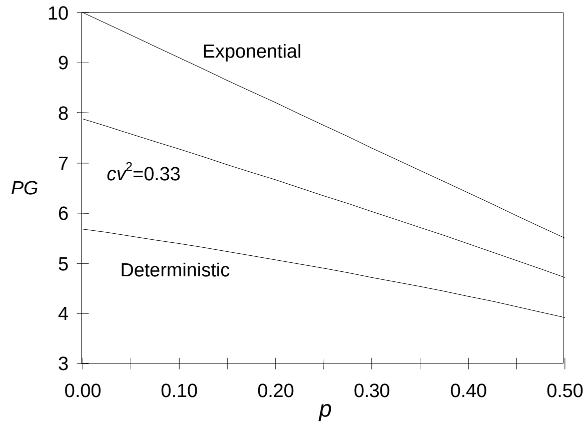

Fig. 3. *PG* for different distributions of the call duration in a PAMR system with 5 channels 90% loaded

VI. NUMERICAL RESULTS

Many numerical tests have been performed to check the proposed approximation for SCV of call duration lower than one with medium and heavy load. For *cv*=0 which represents a D distribution of CHT, the considerations made in Section IV are valid. For *cv*=1, which represents the exponential distribution of the holding time, the proposed approximation gives the exact result as can easily be concluded from Equations (1) and (10) (note that the Kimura approximation is exact for *cv*=1).

Although the approximation has not been tested for coefficients of variation larger than 1, the authors believe that it keeps its validity, especially when the SCV is not very large. This is possible mainly due the fact that the Kimura approximation continues to be valid for coefficients of variation higher than one.

In this section, the proposed approximation is compared with simulation results, in which Erlang-k distributions have been used to feed the call duration probability distribution. Erlang-k distributions are commonly used in literature to compare approximations for the general distribution with SCV lower than one [8, 10].

Figures 4 and 5 display the average access delay for priority and non-priority calls obtained by simulation (Sim) and by computing the proposed approximation (App) for a PAMR system consisting of 5 channels receiving a channel load of 90%. The mean call duration has been normalized to 1. Delay computed by using the M/M/5 model is displayed for reference (W1M and W2M). The probability density function (p.d.f.) used for the call duration is an E_3 (Erlang-3) whose SCV is 0.33. This agrees with the measurements given in [5, 6] for PAMR systems.

Figure 6 shows the same values for a PAMR system with 15 channels and an offered load of 95%, the SCV is also 0.33. In Table 1 values of the mean waiting time for priority calls calculated according to the proposed procedure are displayed for different coefficients of variation. The E_2 p.d.f. has been used in the simulations for cv^2=0.5 and E_4 for cv^2=0.25. The average delay in Table 1 is expressed in milliseconds for a mean call duration of 1 second.

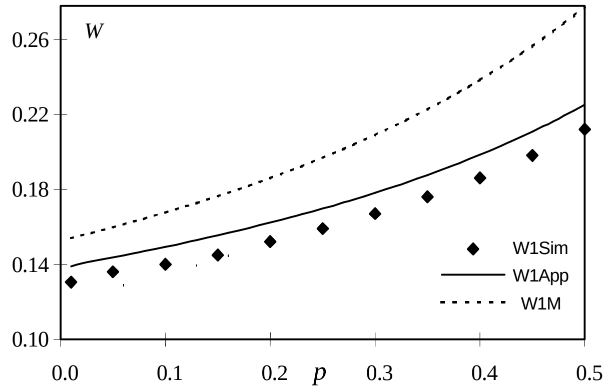

Fig. 4. Mean waiting time vs. priority proportion for priority calls in a PAMR system: 5 channels 90% loaded

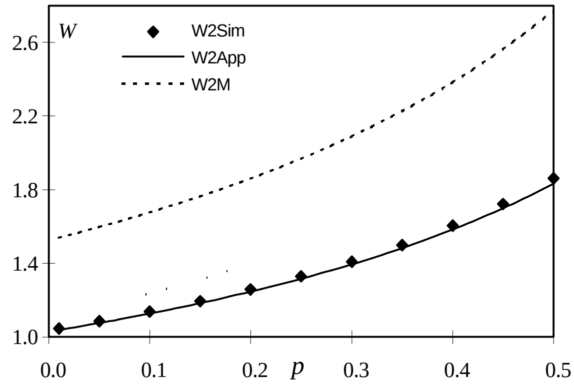

Fig. 5. Mean waiting time vs. priority proportion for regular calls in a PAMR system: 5 channels 90% loaded

Table 1. Mean access delay for priority calls under different conditions of load, priority proportion, number of available channels and SCV of call duration.
S: Simulation; A: Proposed approximation

			C=5		C=12		C=20	
SCV	*r%*	*p%*	**S**	**A**	**S**	**A**	**S**	**A**
0.25	90	10	137	146	51.3	54.3	27.5	28.6
	95	20	158	169	66.4	69.0	39.1	39.4
	90	10	148	158	56.2	58.7	30.0	30.8
	95	20	173	184	72.9	74.8	42.1	42.7
0.5	90	10	145	155	53.7	56.2	28.2	29.3
	95	20	169	179	68.5	71.3	39.1	40.4
	90	10	159	169	58.8	61.3	31.0	32.0
	95	20	186	197	75.6	78.3	43.3	44.3

VII. CONCLUSION

An adequate combination of exact results for the M/M/C queue with priority and approximated results for the M/G/C with no priority leads to an approximation for the M/G/C queue with priority. The proposed approximation has two conditions that make it useful for engineering purposes: it is both easy to compute and accurate. This method has been numerically tested in an environment similar to the one commonly found in trunked PAMR systems: heavy load, two priority levels, low priority proportion and SCV of the call duration lower than one. The accuracy in the estimation of the GoS (in this paper average access delay) in a PAMR system helps to avoid system oversize and waste of radio spectrum.

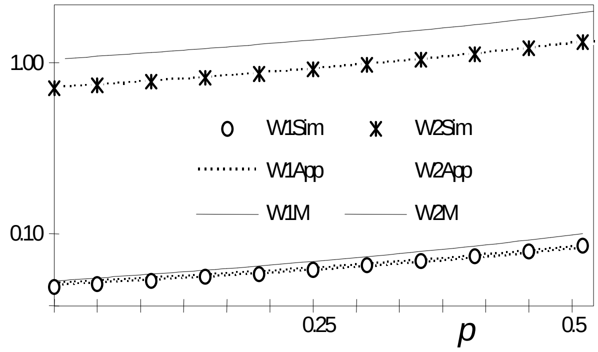

Fig. 6. Average waiting delay for a PAMR system with 15 channels 95% loaded

ACKNOWLEDGMENT

This project was funded by Spanish CICYT TIC940475.

REFERENCES

[1] E. N. Singer, *Land Mobile Radio Systems*, Prentice Hall, 1994.

[2] CCIR, *Multi-Channel Land Mobile Systems for Dispatch Traffic*, CCIR, Geneva 1990.

[3] H. Hoang, R. Malhamé, G. Chan, "Traffic Engineering of Trunked Mobile Radio Dispatch Systems", *Proc. IEEE 41st Vehicular Technology Conference*, pp. 251-256, 1991.

[4] D. Gross, C. M. Harris, *Fundamentals of Queueing Theory* , John Wiley & Sons, New York, 1974.

[5] J. Jordán, F. Barceló, Statistical Modelling of Channel Occupancy in Trunked PAMR Systems, *Proc. 15th International Teletraffic Congress ITC15*, pp. 1169-1178, Washington DC, 1997.

[6] F. Barceló, J. Jordán, "Voice Holding Time Distribution in Trunked PAMR Systems", *Proc. IEEE 47th Vehicular Technology Conference*, pp. 436-440, 1997.

[7] F. Barceló, V. Casares, J. Paradells, "The M/D/C Queue with Priority: Application to trunked Mobile Radio Systems", *IEE Electronics Letters*, Vol 32, pp. 1644-1645, 1996

[8] H. C. Tijms, *Stochastic Modelling and Analysis: A Computational Approach*, John Wiley & Sons, 1986.

[9] L. P. Seelen, H. C. Tijms, M. H. van Horn, *Tables for Multi-Server Queues*, North Holland, 1985.

[10] T. Kimura, "Approximations for multi-server queues: system interpolation", *Queuing Systems*, Vol 17, pp. 347-382, 1994.

[11] P. Cohen, H. H. Hoang, D. Haccoun, "Traffic Characterization and Classification of Users of Land Mobile Communications Channels", *IEEE Trans. on Vehicular Technology*, VT-33, pp. 276-284, 1994.