ITC19/ Performance Challenges for Efficient Next Generation Networks 11-20
LIANG X.J. and XIN Z.H.(Editors)
V.B. IVERSEN and KUO G.S.(Editors)
Beijing University of Posts and Telecommunications Press

# System Capacity Calculation for Packet-Switched Traffic in the Next Generation Wireless Systems, Part I: M/G/1 Nonpreemptive Priority Queuing Model for IP Packet Transmission

Tim Irnich and Bernhard H. Walke
Chair of Communication Networks, RWTH Aachen University
Kopernikusstr. 16, 52074 Aachen, Germany
Email: {tim|walke}@comnets.rwth-aachen.de

Hideaki Takagi
University of Tsukuba, Tsukuba-shi, 305-8573 Japan
Email: takagi@sk.tsukuba.ac.jp

**Abstract:** This paper describes a method of calculating the system capacity required to carry the traffic of multiple packet based service classes under QoS constraints mean packet delay and packet delay percentile. The approach is based on the M/G/1 queue with non-preemptive priorities. Numerical results on the sensitivity of the system capacity towards parameters offered traffic, mean and second moment of packet size distribution, required mean delay and required delay percentile are given.
**Keywords:** mobile radio systems, wireless systems, non-preemptive priority queue, modelling packet-switched traffic, capacity dimensioning, waiting time CDF approximation

## 1 INTRODUCTION

We present a method for calculating the capacity of a cell of a wireless system that is needed to fulfill service-specific *Quality of Service* (QoS) requirements of an arbitrary number of different service classes offering packet traffic. QoS requirements are considered in terms of required mean packet delay and required $r^{th}$ packet delay percentile (i.e., the percentile that corresponds to r% probability). The method is based on the concept described in [1].

This work is based on a number of contributions by the authors to the development of a new standardized methodology for estimating the future radio spectrum requirement for mobile communication, which is conducted by the *International Telecommunication Union, Radiocommunication Sector* (ITU-R). ITU-R's work on this issue will result in a new ITU-R Recommendation replacing the current ITU standard method for spectrum requirement estimation [2]. Since packet-based traffic is forecasted to be the dominant switching scheme in future wireless networks, the main motivation for developing a new spectrum requirement estimation methodology is that the current methodology recommendation [2] does not sufficiently consider packet-based traffic. ITU requires a reasonable compromise between accuracy of modelling, complexity of computation and transparency for the new method, which has been found in the form of the simple model presented in this paper. Large parts of the concept described here recently have been accepted as the basis for ITU's new spectrum estimation methodology. Alternative proposals have been considered during ITU-R's discussion. One alternative is described in a companion paper also submitted to ITC-19 [3].

The method described in this paper is based on an M/G/1 queue with non-preemptive priorities

and $\rho_{\le k}$ denotes the aggregated system load of priority $k$ and all higher priorities, i.e.,

$$\lambda_{\le k} = \sum_{i=1}^{k} \lambda_i, \qquad \beta^{(2)}_{\le k} = \sum_{i=1}^{k} \frac{\lambda_i}{\lambda_{\le k}} \beta^{(2)}_i, \quad \text{and} \qquad \rho_{\le k} = \sum_{i=1}^{k} \lambda_i \beta_i$$

We denote the mean delay of a customer of class $n$ by $D_n$, which is given by

$$D_n := E[\mathcal{D}_n] = W_n + \beta_n = \frac{\lambda_{\le N} \beta^{(2)}_{\le N}}{2\,(1-\rho_{\le n})\,(1-\rho_{\le n-1})} + \beta_n. \tag{2}$$

The second moment of the waiting time of customers of class $n$ is

$$\begin{aligned} W^{(2)}_n &:= E[\mathcal{W}^2_n] \\ &= \frac{\lambda_{\le N} \beta^{(3)}_{\le N}}{3\,(1-\rho_{\le n})\,(1-\rho_{\le n-1})^2} + \frac{\lambda_{\le N} \beta^{(2)}_{\le N} \lambda_{\le n} \beta^{(2)}_{\le n}}{2\,(1-\rho_{\le n})^2\,(1-\rho_{\le n-1})^2} + \frac{\lambda_{\le N} \beta^{(2)}_{\le N} \lambda_{\le n-1} \beta^{(2)}_{\le n-1}}{2\,(1-\rho_{\le n})\,(1-\rho_{\le n-1})^3}, \end{aligned} \tag{3}$$

where $\beta^{(3)}_{\le N}$ denotes the third moment of the weighted common service time DF, defined by

$$\beta^{(3)}_{\le N} = \sum_{i=1}^{k} \frac{\lambda_i}{\lambda_{\le k}} \beta^{(3)}_i.$$

## 3 MEAN DELAY BOUND SYSTEM CAPACITY

Assume the M/G/1 non-preemptive priority queue for modelling the downlink traffic in a cell of a packet-based wireless system that consists of $N$ different classes of *Internet Protocol* (IP) based services, where each service class $n, n = 1, ..., N$ corresponds to one customer class of the queue. Each customer corresponds to one IP packet. The size [bits] of an IP packet of class $n$ is denoted by $\mathcal{S}_n$ with moments

$$s_n := E[\mathcal{S}_n], \quad s^{(2)}_n := E[\mathcal{S}^2_n], \ \ldots \qquad 1 \le n \le N.$$

If an IP packet of class $n$ is transmitted over a channel with capacity $C$ [bits/s], the service time of the packet is $\mathcal{T}_n = {}^{\mathcal{S}_n}\!/_{C}$. Accordingly, mean and second moment of the service time DF are

$$\beta_n = \frac{s_n}{C} \quad \text{and} \quad \beta^{(2)}_n = \frac{s^{(2)}_n}{C^2}. \tag{4}$$

Substituting the expressions in (4) into (2) results in an expression for the mean IP packet delay as a function of the system capacity $C$,

$$D_n(C) = \frac{\lambda_{\le N} s^{(2)}_{\le N}}{2\left(C - \sum_{i=1}^{n} \lambda_i s_i\right)\left(C - \sum_{i=1}^{n-1} \lambda_i s_i\right)} + \frac{s_n}{C} \tag{5}$$

If the QoS requirement for IP packets of class $n$, is given in terms of a required mean delay $D_n$, the required system capacity $C_n$ is defined as the capacity satisfying the condition $D(C_n) = D_n$. Given a certain value for $D_n$, the system capacity $C_n$ required to achieve $D_n$ can be calculated by solving (5) for C. Among the three roots of this equation there is always one that satisfies the stability condition

$$C_n > \sum_{i=1}^{n} \lambda_i s_i. \tag{6}$$

The $r^{th}$ percentile $\pi_n(r)$ of service class $n$ is met if

$$P(\mathcal{D}_n \leq \pi_n(r)) \geq r. \tag{10}$$

If (10) is not satisfied, the system capacity needs to be increased, and the new $P(\mathcal{D}_n \leq \pi_n(r))$ that is valid for this increased value of the system capacity needs to be determined. This procedure needs to be repeated until (10) is satisfied. Describing the service time CDF by (8) is equivalent to assuming an n-point packet size distribution

$$P(\mathcal{X}_n = t) = \sum_{m=1}^{M_n} p_{m,n}\,\delta(x - x_{m,n}), \tag{11}$$

where $\mathcal{X}_n$ denotes the packet size random variable. This means that for packets of service class $n$ transmitted over a channel with capacity $C$ the $t_{m,n}$ values in (8) are defined by the set of possible values for the packet size $\{x_{1,n}, x_{2,n}, ..., x_{M,n}\}$, where $t_{m,n} = x_{m,n}/C$. In this case the first two moments of the packet size distribution for service class $n$ are

$$s_n = \sum_{m=1}^{M_n} p_{m,n} \cdot x_{m,n} \quad \text{and} \quad s_n^{(2)} = \sum_{m=1}^{M_n} p_{m,n} \cdot x_{m,n}^2.$$

This means that mean, second and third moment of the packet size distribution cannot be chosen arbitrarily. In this case the mean delay bound system capacity (see Sec. 3) needs to consider mean and second moment of the packet size DF according to the formulas above.

## 5 SENSITIVITY ANALYSIS

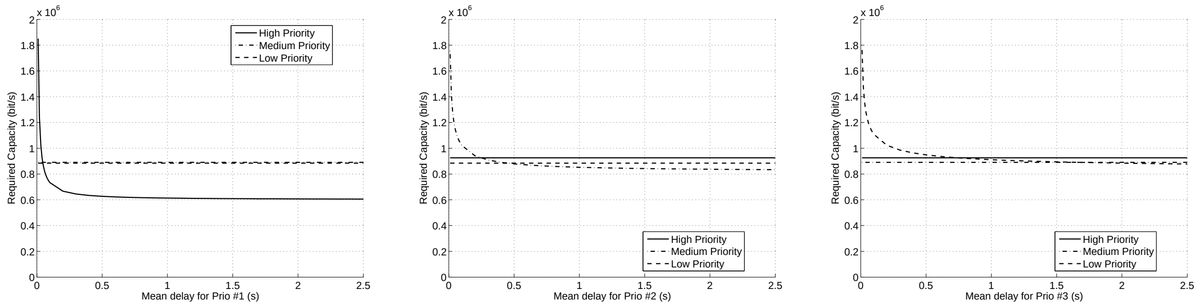
This section illustrates the sensitivity of the model towards its input parameters. In Sec. 5.1 only the mean delay bound capacity is considered. The sensitivity of the delay percentile bound capacity is examined in Sec. 5.2.

The scenario considered consists of three different service classes, called high priority class, medium priority class and low priority class. The assumed parameter values are listed in Tab. 1. For the results presented in Sec. 5.1 a negative exponential packet size distribution has been assumed, i.e., $E[\mathcal{X}^2] = 2(E[\mathcal{X}])^2$ and $E[\mathcal{X}^3] = 6(E[\mathcal{X}])^3$. To illustrate the influence of a certain parameter on the required system capacity, only one parameter is varied at a time, and the other parameters are constantly set to the values shown in Tab. 1. Since the model allows different values for each class, the influence of each parameter is evaluated separately for each class, i.e. the influence of for example the mean packet size of the high priority class is evaluated under constant mean packet size of the medium and low priority class, etc.

### 5.1 Sensitivity of Mean Delay bound System Capacity

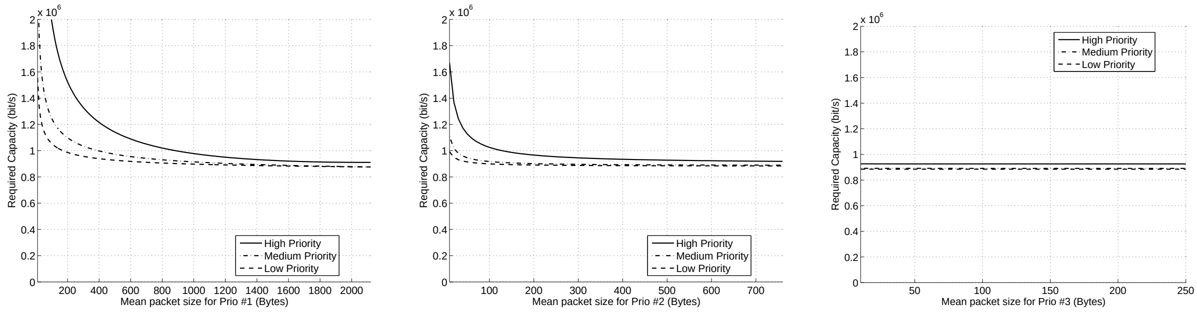
In Fig. 2 the influence of the offered traffic on the required system capacity is shown. The offered traffic is increased from 1kbit/s to 2Mbit/s in steps of 1kbit/s. In Fig. 2(a) the traffic offered for the high priority class is increased, Fig. 2(b) illustrates the dependency of the system capacity on the offered traffic for the medium priority class, while Fig. 2(c) shows the impact of the traffic offered on the low priority class.

Except for very low traffic the required system capacity exhibits a linear dependency on the offered traffic. Increasing traffic of the high priority class leads to increasing load for all classes, increasing the required system capacity for all classes; see Fig. 2(a). If the traffic for the medium priority class is increasing, the high priority class is affected by packets of medium or low priority




(a) High Priority (b) Medium Priority (c) Low Priority

Figure 3: Mean delay bound system capacity vs. mean delay requirement




(a) High Priority (b) Medium Priority (c) Low Priority

Figure 4: Sensitivity of system capacity required for mean delay to mean packet size

asymptote is defined by the stability condition of the queue. The capacity requirement of a particular priority class does not take into account the delay requirements of other priority classes. Thus, the required capacity of the priority classes not being varied is constant.

In Fig. 4 the influence of the mean packet size on the required system capacity is shown. The mean packet size is increased from 100byte in steps of 10byte for all priorities. The second moment is kept constant, so that this case is equivalent to varying the coefficient of variation. Under constant second moment of the packet size the maximum values for the mean packet size are limited by the conditions

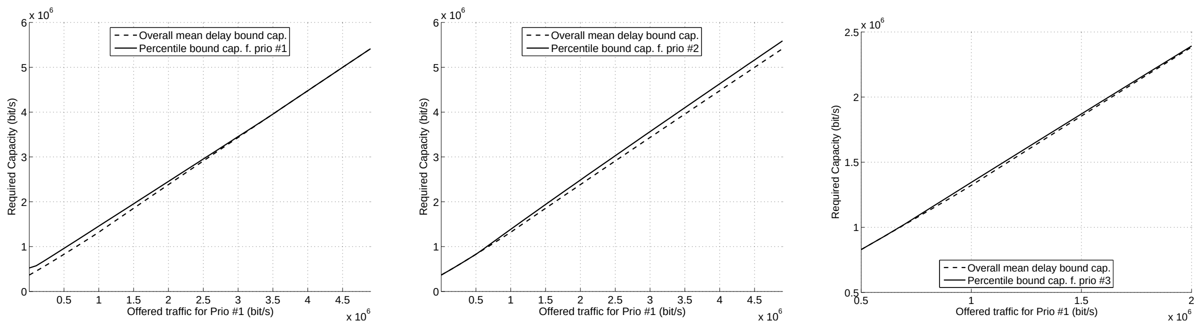
$$\begin{aligned} s_1 &\leq \sqrt{s_1^{(2)}} = \sqrt{4.5 \cdot 10^6} = 2120 \\ s_2 &\leq \sqrt{s_2^{(2)}} = \sqrt{5.832 \cdot 10^5} = 760 \\ s_3 &\leq \sqrt{s_3^{(2)}} = \sqrt{6.48 \cdot 10^4} = 254 \end{aligned} \quad (12)$$

In Fig. 4(a) the mean packet size of the high priority class is increased, Fig. 4(b) illustrates the dependency of the required system capacity on the mean packet size for the medium priority, while Fig. 4(c) shows the impact of the mean packet size for the low priority. The results shown here are obtained for constant offered traffic, i.e. for increasing mean packet size the packet arrival rate is lowered so that the product of arrival rate and mean packet size is constant. The priority with highest offered traffic (i.e., the high priority class) shows the strongest dependency on the mean packet size. For classes with low offered traffic the influence of mean packet size becomes very small, but the slope is not equal to zero (see Fig. 3(c)).

In Fig. 5 the influence of the second moment of the packet size distribution on the required

Table 2: Values for $p_{1,n}$, $x_{1,n}$ and $x_{2,n}$ leading to $s_n$, $s_n^{(2)}$ and $s_n^{(3)}$ equal to values in Tab. 1

| Parameter | Priority | Value |
|---|---|---|
| $x_{1,n}$ | High | 879 |
| unit: byte | Medium | 316 |
| | Low | 105 |
| $x_{2,n}$ | High | 5121 |
| unit: byte | Medium | 1844 |
| | Low | 615 |
| $p_{1,n}$ | High | 0.8536 |
| | Medium | 0.8536 |
| | Low | 0.8536 |




(a) High Priority (b) Medium Priority (c) Low Priority

Figure 6: Sensitivity of percentile bound system capacity offered traffic

In Fig. 6 the influence of the offered traffic for the high priority service class on the delay percentile bound system capacity is shown. The behavior of the delay percentile bound capacity towards variation of offered traffic for medium and low priority service is similar. The required delay percentile is varied in Fig. 6(b) for the high priority, in Fig. 6(b) for the medium priority and in Fig. 6(c) for the low priority.

Fig. 7 visualizes the sensitivity of the delay percentile bound system capacity towards the required delay percentile. It is obvious that the influence of the required delay percentile on the required system capacity can be significant when the percentile requirement is close to or lower that the mean delay requirement.

## 6 Conclusion

This paper presents a new method for calculating the capacity of a wireless system that is needed to fulfill QoS requirements given in terms of required mean delay and required delay percentile. The method is based on a novel way of utilizing the M/G/1-FCFS queue with non-preemptive priorities. The presented approach for calculating the required system capacity has been accepted as a core concept within ITU's new spectrum requirement estimation methodology. It fulfills ITU's requirements of technology neutrality and reasonable trade-off between complexity and accuracy in modelling packet traffic. The sensitivity analysis shows that the offered traffic is the primary influence to the required capacity, and that the model's behaviour in general is in line