Blocking of dynamic multicast connections in a single link

Jouni Karvo, Jorma Virtamo, Samuli Aalto, Olli Martikainen
Helsinki University of Technology
P.O.Box 1100, FIN-02015 HUT, Finland
Email: `{jouni.karvo,jorma.virtamo,samuli.aalto, olli.martikainen}@hut.fi`

Abstract

In this paper, a method for calculating blocking experienced by dynamic multicast connections in a single link is presented. A service center at the root of a tree-type network provides a number of channels distributed to the users by multicast trees which evolve dynamically as users join and leave the channels. We reduce this problem to a generalized Engset system with nonidentical users and generally distributed holding times, and derive the call and channel blocking probabilities as well as the link occupancy distribution.

Keywords

Multicast, blocking, queueing systems, generalized Engset system

1 INTRODUCTION

Call blocking probabilities in a circuit switched network carrying multiple traffic classes can be calculated with exact algorithms, such as the recursion of Kaufman (1981) and Roberts (1981), or with approximative methods, such as the normal type approximation (Naoumov 1995). These algorithms are applicable for point to point connections, such as telephone calls or ATM connections. They apply also for static multicast connections, where the structure of each multicast tree is fixed in advance. In a more dynamic environment, where the trees evolve with arriving and departing customers, these models are not adequate.

Multicast connections have a bandwidth saving nature. This means that a multicast connection – in taking the form of a tree where streams merge at the nodes – requires much less capacity from the network links than a bunch of separate point to point connections from the root node to the leaf nodes of the tree (see figure 1). This effect poses a problem when dimensioning a network. With multicast trees, less bandwidth is needed near the root of the tree than the sum of the bandwidths used near the leaves.

Diot *et al*. (1997) give a review on the work in the area of multicast traffic. Most

Broadband Communications P. Kühn & R. Ulrich (Eds.)

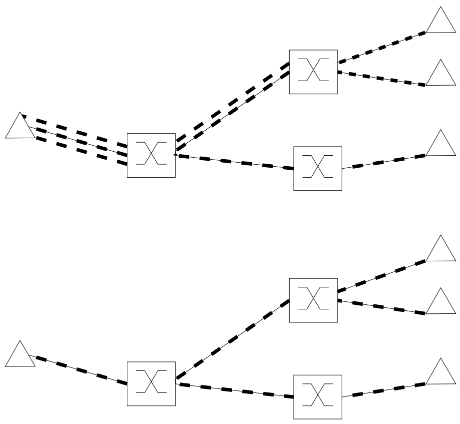

Figure 1 Point to point (top) vs. point to multipoint, or multicast connections (bottom). Data streams are presented with a thick broken line.

work in the sense of blocking in ATM networks with multicasting has been made on blocking in multicasting capable switches, see, for example, (Giacomazzi and Trecordi 1995) or (Kim 1996). Shacham and Yokota (1997) propose call admission control algorithms for real time multicast transmission. We use a similar setting and device an algorithm for calculating call blocking probabilities.

The model of dynamic multicast connections used in this paper pertains, e.g., to the case in which TV or radio station provides several programs to viewers or listeners via a telecommunication network. The model consists of a tree-type distribution network. The service center located at the root node offers the users at the leaves a set of programs delivered to the subscribing users by multicast channels. The programs run independently of their subscribers, who can join and leave the channel any time. Thus, each channel forms a dynamic tree.

A joining user is assumed to choose a channel probabilistically according to a channel preference distribution, which is the same for all the users. When joining, the user, U, creates a new branch to the tree extending from the leaf to the nearest node, A, already connected to the channel (see figure 2). Blocking may occur on any link of the new branch. On the other hand, there is no blocking on the links upstream from the connecting node A, since on that path the channel is already on. Note, however, that the joining user may extend the time the channel remains switched on.

Several multicast trees may use the same link. The required capacity in the links varies accordingly, sometimes leading to blocking when the requested capacity is not available. The probability that there is not sufficient capacity available for a channel in a specific link is called channel blocking probability on that link. By the call blocking probability we mean the probability that the user's request to subscribe to a channel is blocked. The latter is smaller, since the user's subscription is always

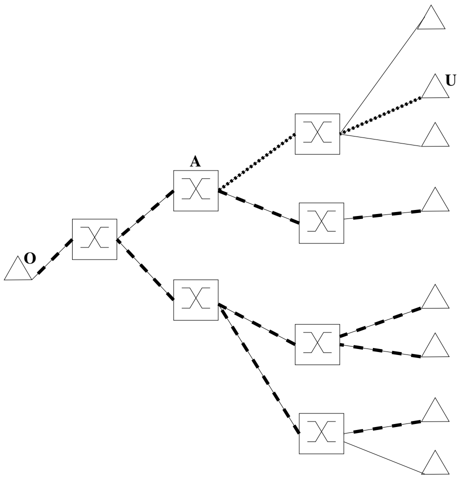

Figure 2 The tree of a channel with a new branch created by a joining user U.

accepted when the channel is already on. Both these blocking probabilities can be calculated on link or end-to-end basis. In this paper we show how to calculate the blocking probabilities of dynamic multicast connections in a specific link with finite capacity assuming that all the other links have infinite capacity.

The paper is organized as follows. In section 2 we present preliminary considerations related to a network with all the links having an infinite capacity. The case of a finite capacity link is considered in section 3. The correspondence with the generalized Engset system is given, and the required blocking probabilities are derived. An example of link occupancy and call blocking probability calculations is presented in section 4. Section 5 gives a brief summary.

2 LINK OCCUPANCY IN AN INFINITE SYSTEM

In this section we consider a link in a system, which has infinite capacity on all its links, and where, consequently, no blocking occurs. First we set the mathematical model up including the assumptions needed later on. Then we determine the mean times that an individual channel traversing the link is on and off. Finally we show how to calculate the distribution for the link capacity usage. The results will be utilized in the next section, where we focus on the blocking problem of a link with finite capacity.

Consider a link in an infinite system. The multicast channel population is denoted by I, i.e., I is the set of channels ('programs') provided by the service center. Let $c_i \in Z_+$ denote the capacity requirement of channel $i \in I$. We assume that the users downstream of the considered link subscribe to these channels according a Poisson process with intensity λ. This is a model for an infinite user population, which is a reasonable assumption in networks with a large number of users, such as TV or radio multicasting in a network (for a link not too close to the leaves of the multicast tree). Further, we assume that each user chooses the channel independently of others and from the same preference distribution, α_i being the probability that channel i is chosen. As a result, the subscriptions to channel i arrive according to a Poisson process with intensity $\lambda_i = \alpha_i\lambda$. We assume that the users' holding times are generally distributed with mean $1/\mu_i$. Finally, let a_i denote the offered traffic intensity for channel i,

$$a_i = \lambda_i/\mu_i. \tag{1}$$

Consider then the on and off times of a single channel. Let $T_{i,\mathrm{on}}^{(\infty)}$ and $T_{i,\mathrm{off}}^{(\infty)}$ denote their means, respectively. As mentioned above, no blocking occurs in an infinite system. Thus, if the channel is off, it is turned on every time a new subscription arrives. The channel remains in the on state (and occupies the link) as long as there are users connected to the channel. Thus, the probability p_i that channel i is on equals the probability that there is **at least one** user connected to the channel. The probability q_i that channel i is off is then the same as the probability that there are **no** users connected to the channel. On the other hand, under the assumptions made above, the number of users simultaneously connected to channel i is distributed as the number of customers in an $M/G/\infty$ queue, i.e. according to the Poisson distribution with mean a_i. Thus,

$$p_i = 1 - e^{-a_i}, \tag{2}$$

$$q_i = e^{-a_i}. \tag{3}$$

Another implication is that on and off times of the channel considered are distributed as busy and idle periods, respectively, in the corresponding $M/G/\infty$ queue. Thus,

$$T_{i,\mathrm{on}}^{(\infty)} = \frac{e^{a_i} - 1}{\lambda_i}, \tag{4}$$

$$T_{i,\mathrm{off}}^{(\infty)} = \lambda_i^{-1}. \tag{5}$$

The former equation follows from the fact that

$$p_i = \frac{T_{i,\mathrm{on}}^{(\infty)}}{T_{i,\mathrm{on}}^{(\infty)} + T_{i,\mathrm{off}}^{(\infty)}}. \tag{6}$$

We see that the mean on time of the most popular channels as a function of the offered traffic intensity grows extremely rapidly because of the exponential term in the numerator. This indicates that there is likely to be a set of channels that are almost constantly carried on the link.

Let X denote the number of channels in use, and X_i indicate whether channel i is on ($X_i = 1$) or off ($X_i = 0$). Since

$$X = \sum_{i \in I} X_i, \tag{7}$$

where the X_i are independent Bernoulli variables with mean p_i, we have

$$\mathrm{E}[X] = \sum_{i \in I} p_i, \tag{8}$$

$$\mathrm{Var}[X] = \sum_{i \in I} p_i q_i. \tag{9}$$

Let then Y denote the number of capacity units simultaneously occupied in the link,

$$Y = \sum_{i \in I} c_i X_i. \tag{10}$$

Its distribution $(\pi_j)_{j=0}^{\infty}$, called the link occupancy distribution, can be calculated by the convolution algorithm (Iversen 1987), or, equivalently, from the probability generating function:

$$P(z) = \prod_{i \in I} (q_i + p_i z^{c_i}) = \sum_{j=0}^{\infty} \pi_j z^j. \tag{11}$$

As regards the mean and variance of Y, it follows from (10) and the independence of the Bernoulli variables X_i that

$$\mathrm{E}[Y] = \sum_{i \in I} c_i p_i, \tag{12}$$

$$\mathrm{Var}[Y] = \sum_{i \in I} c_i^2 p_i q_i. \tag{13}$$

All the results in this section are valid in a system in which all the links have infinite capacity. When multicast connections are carried on a link which has finite capacity, blocking may occur. This is studied in the next section.

3 BLOCKING IN A LINK WITH FINITE CAPACITY

In this section we show how to calculate blocking probabilities in a link with finite capacity, C, assuming that all the other links have infinite capacity.

It is important to make a distinction between various types of blocking. The **channel blocking probability** B_i^c of channel i is defined to be the probability that an attempt to turn channel i on fails due to lacking capacity, whereas the **call blocking probability** b_i^c of channel i (seen by a user subscribing to channel i) refers to the probability that a user's attempt to subscribe to channel i fails. These are different, since the user's subscription is always accepted when the channel is already on. Finally we define the **time blocking probability** B_i^t of channel i to be the probability that at least $C - c_i + 1$ capacity units of the link are occupied.

Consider a single channel $i \in I$. Denote by $T_{i,\text{on}}$ and $T_{i,\text{off}}$ the mean on and off periods, respectively, in this finite system. By considering a cycle consisting of an on period and the following off period, we deduce that the call blocking probability of channel i is

$$b_i^c = \frac{\lambda_i T_{i,\text{off}} - 1}{\lambda_i T_{i,\text{on}} + \lambda_i T_{i,\text{off}}}, \tag{14}$$

where $\lambda_i T_{i,\text{off}} - 1$ is the mean number of failed attempts to subscribe to channel i during the cycle (the last subscription arriving in the off period will be accepted), and the denominator represents the mean total number of attempts during the cycle. The frequency of accepted calls when the channel is off is clearly $\lambda_i(1 - B_i^c)$. Thus,

$$T_{i,\text{off}} = \frac{1}{\lambda_i(1 - B_i^c)}. \tag{15}$$

On the other hand, we observe that in this finite system (where the capacities of all the other links are assumed to be infinite) the on period of a channel is independent of the evolution of the other channels: once the channel is turned on all the incoming subscriptions will be accepted. This implies that the on periods are distributed as those of an infinite system. Thus,

$$T_{i,\text{on}} = T_{i,\text{on}}^{(\infty)} = \frac{e^{a_i} - 1}{\lambda_i}. \tag{16}$$

By combining equations (14), (15) and (16), we obtain the following expression for the call blocking probabilities of channel i:

$$b_i^c = \frac{B_i^c}{(1 - B_i^c)(e^{a_i} - 1) + 1}. \tag{17}$$

Thus, the only item that still remains to be determined is the channel blocking probability B_i^c. We start the derivation by observing that our finite system can be described as a generalized Engset system.

By an Engset system we refer to the well known $M/M/m/m/K$ system with a finite user population, see (Kleinrock 1975). In a generalized Engset system the users are nonidentical, that is their mean holding and interarrival times as well as the requested resources can be different. Moreover, we allow the holding times to have a general distribution.

The channels in our system represent the users in the Engset system. When the channel is on, the 'user' is active, and when the channel is off, the 'user' is idle. Thus, the holding time of user i in the generalized Engset system is generally distributed with mean $T_{i,\mathrm{on}}$, and the interarrival time is exponentially distributed with mean λ_i^{-1}. As a consequence, we deduce that the channel blocking probability B_i^c equals the call blocking probability of user i in the corresponding generalized Engset system. Similarly, the time blocking probability B_i^t equals that of the generalized Engset system.

The time blocking probability of user i in the generalized Engset system can be calculated from the following formula:

$$B_i^t = \frac{\sum_{j=C-c_i+1}^{C} \pi_j}{\sum_{j=0}^{C} \pi_j}, \tag{18}$$

where π_j is the probability that j capacity units are occupied in an infinite system as defined in equation (11). In the special case that $c_i = 1$ for all i, this follows from the result of Cohen (1957). In the general case, it can be shown to follow from the insensitivity property of the product form probabilities of a multirate loss system (Hui 1990). It is also known that the call blocking of user i equals the time blocking (of user i) in a system where user i is removed. Thus the channel blocking probability is as follows:

$$B_i^c = \frac{\sum_{j=C-c_i+1}^{C} \pi_j^{(i)}}{\sum_{j=0}^{C} \pi_j^{(i)}}, \tag{19}$$

where $\pi_j^{(i)}$ is the probability that j capacity units are occupied in an infinite system with user i removed. These occupancy probabilities can be identified from the probability generating function

$$\sum_{j=0}^{\infty} \pi_j^{(i)} z^j = \prod_{k \in I - \{i\}} (q_k + p_k z^{c_k}), \tag{20}$$

where $I - \{i\}$ denotes the reduced set of users. Alternatively, in order to save computational effort, one may use deconvolution, which means

$$\prod_{k \in I-\{i\}} (q_k + p_k z^{c_k}) = \frac{P(z)}{q_i + p_i z^{c_i}}, \tag{21}$$

but numerical problems may arise for large systems.

To summarize, the call blocking b_i^c can be calculated from formula (17) by using (19). Note that the denominator in (17) is always greater than 1. Thus, the call blocking b_i^c seen by a user subscribing to channel i is always smaller than the corresponding channel blocking B_i^c. This reflects the fact that the users subscribing to a channel while the channel is on do not experience any blocking. We see also that, for the most popular channels, blocking seen by a user drops practically to zero, since the exponential term in the denominator grows rapidly with a_i ($b_i^c \approx B_i^c e^{-a_i}$). For a channel with $a_i \ll 1$, the channel blocking and the call blocking seen by a user are approximately the same.

Since $b_i^c \leq B_i^c \leq B_i^t$, an upper limit for the call blocking is the time blocking in a system with all channels present. No call blocking seen by a user can be higher than this, but call blocking for user approaches it for channels with channel preferences α_i near zero.

4 AN EXAMPLE

As an example, a truncated geometric distribution is used for channel choosing preferences:

$$\alpha_i \propto (1-p)^{i-1}, \qquad i = 1, 2, \ldots, |I|, \tag{22}$$

where the index i is the channel number. With this numbering the channels are in a descending order according to the usage. For numerical calculations we choose $p = 0.2$ and $|I| = 200$. We choose the average viewing time $1/\mu = 900$ s to be the same for all channels, and $\lambda/\mu = 3.5 \cdot 10^6$. The link capacity is $C = 70$ channels, and each channel has identical capacity requirement $c_i = 1$.

First, we consider an infinite capacity link. The channel usages p_i are depicted in figure 3 as a function of channel index i. The mean and variance of the number of used channels are calculated from equations (8) and (9). They are 63.4 and 3.11, respectively. In this example, where all capacity requirements are equal to 1, these same values hold for mean and variance of capacity requirement for the channels. The link occupancy distribution, π_j, equation (11), is presented in figure 4. The mean channel on times, $T_{i,\mathrm{on}}$, equation (16), for this example are presented in figure 5. As expected, the most popular channels have practically infinite on times.

Then we consider a system with finite link capacity in a link, $C = 70$. Call and channel blocking probabilities, equations (14) and (19) respectively, for each channel

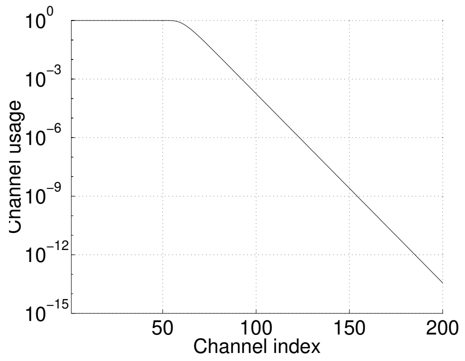

Figure 3 Channel usage versus channel index for geometrically distributed user preferences.

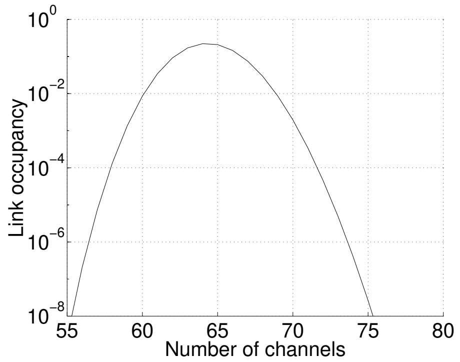

Figure 4 Link occupancy distribution for geometrically distributed user preferences.

are shown in figure 6. This shows clearly that the call and channel blocking probabilities differ for the most popular channels for which there is a high probability that the channel is already on when the user subscribes the channel.

5 SUMMARY

In this paper, we have presented a method for calculating the call and channel blocking probabilities in a link carrying multicast traffic. Multicast traffic has the property of requiring less link capacity than a set of point to point connections providing the same connectivity.

The blocking calculation presented gives us a grip of TV or radio delivery on a circuit switched multicast system, such as an ATM network with virtual circuits. We

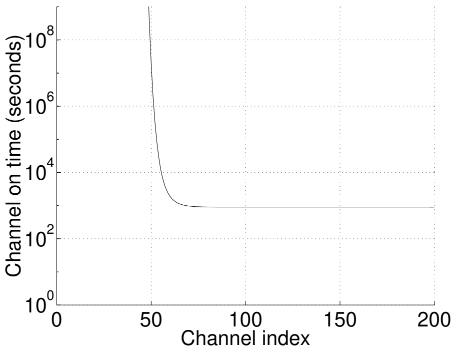

Figure 5 The mean on times $T_{i,\mathrm{on}}$ in seconds for each channel.

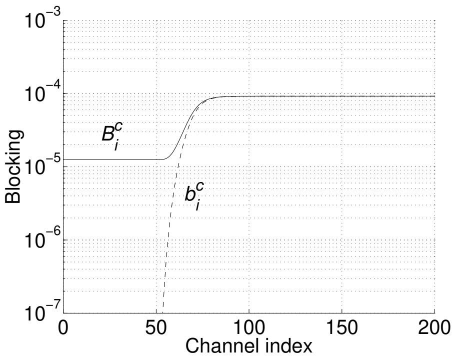

Figure 6 Channel B_i^c and call blockings b_i^c for each channel (B_i^c solid line, b_i^c dashed line).

are able to calculate the capacity needed for a link, the average channel on times, and blocking introduced by finite capacity.

The calculation started from a distribution of the users' preferences on multicast channels, from which we derived the link occupancy distribution in an infinite capacity system, and, further, the blocking probabilities in a finite system by mapping the problem to an equivalent generalized Engset system. Finally, an example was given with the geometric distribution of users' preferences.

In this paper, we have limited ourselves to the case of a single link. The case where blocking occurs on several links of a network is left for future work. Possibly methods such as Reduced Load Approximation, see for example (Ross 1995), can be used, but blockings on successive links may exhibit significant dependencies.

It is also likely that the users' actual preference distribution affects significantly the blocking introduced in the network. This part of the study would require statistics from a real-life trace, and is left for future work.

ACKNOWLEDGEMENTS

The authors wish to thank prof. Gunnar Karlsson, Petteri Koponen and Juhana Räsänen from the Helsinki University of Technology for their inspiring thoughts.

REFERENCES

Cohen, J.W. (1957) The generalized Engset formulae. *Philips Telecommunication Review* **18**(4), 158–170.

Diot, C., Dabbous, W., and Crowcroft, J. (1997) Multipoint communication: A survey of protocols, functions, and mechanisms. *IEEE Journal on Selected Areas in Communications* **15**(3), 277–290.

Giacomazzi, P., and Trecordi, V. (1995) A study of non blocking multicast switching networks. *IEEE Transactions on Communications* **43**(2–4), 1163–1168.

Hui, J. (1990) *Switching and traffic theory for integrated broadband networks*. Boston, Kluwer Academic Publishers.

Iversen, V.B. (1987) The exact evaluation of multi-service loss systems with access control. *Teleteknik, English Edition* **31**(2), 56–61.

Kaufman, J.S. (1981) Blocking in a shared resource environment. *IEEE Transactions on Communications* **com-29**(10), 1474–1481.

Kim, C.-K. (1996) Blocking probability of heterogenous traffic in a multirate multicast switch. *IEEE Journal on Selected Areas in Communications* **14**(2), 374–385.

Kleinrock, L. (1975) *Queueing systems; Volume 1: Theory*. New York, John Wiley & Sons.

Naoumov, V. (1995) Normal-type approximation for multi-service systems with trunk reservation. *Telecommunication Systems*, **4**, 113–118.

Roberts, J.W. (1981) A service system with heterogenous user requirements – application to multi-services telecommunications systems, in *Performance of data communication systems and their applications* (ed. G. Pujolle), Amsterdam, North-Holland, pp. 423–431.

Ross, K.W. (1995) *Multiservice Loss Models for Broadband Telecommunication Networks*. London, Springer Verlag.

Shacham, N., and Yokota, H. (1997) Admission control algorithms for multicast sessions with multiple streams. *IEEE Journal on Selected Areas in Communications* **15**(3), 557–566.