Telecommunication Systems 16:3,4, 467–481, 2001


# Blocking of Dynamic Multicast Connections

JOUNI KARVO and OLLI MARTIKAINEN {jouni.karvo;olli.martikainen}@hut.fi
*Helsinki University of Technology, Laboratory of Telecommunications Software and Multimedia, P.O. Box 5400, FIN-02015 HUT, Finland*

JORMA VIRTAMO and SAMULI AALTO {jorma.virtamo;samuli.aalto}@hut.fi
*Helsinki University of Technology, Laboratory of Telecommunications Technology, P.O. Box 3000, FIN-02015 HUT, Finland*

**Abstract.** Multicast connections have a bandwidth saving nature. This means that a multicast connection – in taking the form of a tree with streams merging at the nodes – requires much less capacity from the network links than a bunch of separate point-to-point connections providing the same connectivity. In this paper, we consider dynamic multicast connections that can be used to model, for example, TV or radio delivery on a telecommunications network, such as an ATM network with virtual circuits. We show how to calculate the blocking probabilities of requests to join such a tree. First, we consider the blocking probabilities occurring in a single link. The resulting model is able to capture heterogeneous capacity requirements for different multicast channels. Then we extend the results to a whole network using the reduced load approximation. The accuracy of the approximation method is studied by simulations.

**Keywords:** blocking, multicast, circuit-switched system, ATM
**AMS subject classification:** 60K25, 60K30, 68M10

## 1. Introduction

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 providing the same connectivity (figure 1). This has considerable effects on applications like TV or radio delivery by ATM or other telecommunications network.

TV or radio delivery by a switched network has several characteristics: first, the number of recipients is large. For this reason, the intensity with which users join and leave the tree is large and the corresponding protocols need to be carefully designed [2]. Traffic is also mostly unidirectional – only signalling is carried towards the service centre. There are typically several programmes available at the service centre, from which the subscribers choose.

Our purpose in this paper is to show how to calculate blocking probabilities of the users' requests for programmes in these kinds of networks. The model used consists of a tree-type distribution network, whose root is called the service centre and leaves are called users. The service centre offers the users a set of programmes delivered by mul-

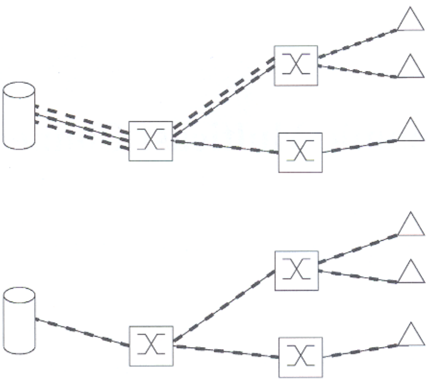

Figure 1. Point to point (top) vs. point to multipoint, or multicast connections (bottom). Data streams are presented with a thick broken line.

ticast channels. Each channel may have a different capacity requirement, allowing thus heterogeneous multicast traffic. The programmes run independently of their subscribers, who can join and leave the channel at any time. When a user chooses a programme, a new branch is created on the corresponding multicast tree. 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 on the tree extending from his 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.

The calculation of blocking probabilities of the users' requests to join multicast channels has two main steps: We start by considering a method for calculating blocking probabilities in a single link. Then we extend the results to calculate the end-to-end blocking probabilities for the whole network.

Traditional methods for calculating link blocking probabilities, such as the Kaufman–Roberts recursion [7,10], 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 our more dynamic environment, where the trees evolve with arriving and departing customers, these models are not adequate. Thus, new methods are needed.

A method for calculating the link blocking probabilities of multicast traffic in the case with infinite user population generating Poisson traffic was devised in [6]. In this study, the exact formulae for blocking probabilities are derived by mapping the problem to an equivalent generalized Engset system with unidentical users and generally distributed holding times. This model for the single link case is used in the present paper as the starting point for approximating the end-to-end blocking probabilities in the whole multicast network. These are calculated by applying the reduced load approximation,

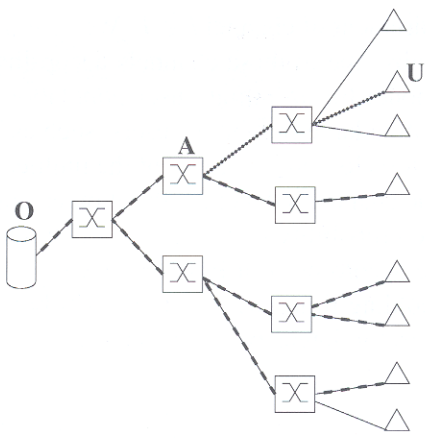


Figure 2. The dynamic multicast tree of a channel with a new branch created by a joining user U.

previously used for calculating the end-to-end blocking probabilities of point-to-point connections (see, e.g., [11]). The accuracy of the approximation method is studied by simulations.

In [2], a review on the work in the area of multicast traffic is given. Most work in the sense of blocking in ATM networks with multicasting has been done on blocking in multicasting capable switches, see, for example, [3,8]. In [12], a call admission control algorithm for real time multicast transmission is proposed. We use a similar setting and devise an algorithm for calculating call blocking probabilities.

The paper is organised as follows. In section 2 we present blocking calculations for a single link. Section 3 contains a brief description of the reduced load approximation. In section 4 we compare the results obtained by this method with those obtained by simulations. Section 5 gives a brief summary.

## 2. Blocking probabilities in a single link

In this section we consider a link in a system which has infinite capacity on all its other links, and where, consequently, blocking occurs only in the examined link. The model described in this section was first presented in [6]. In the first section, we set the mathematical model up including the assumptions needed later on. We consider a case where the link capacity is infinite, and 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 of the first subsection will be utilized in the second section, where we focus on the blocking problem of a link with finite capacity.

### *2.1. Link occupancy in an infinite system*

Consider a link in an infinite system. The multicast channel population is denoted by $I$, i.e., $I$ is the set of channels ('programmes') provided by the service centre. Let $c_i \in \mathbb{Z}_+$

denote the capacity requirement of channel $i \in I$. We assume that the users downstream of the considered link subscribe to these channels according to a Poisson process with intensity $\lambda$. 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, $\alpha_i$ 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 $i$,

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

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,

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

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

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,

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

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

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 [5], 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.

### 2.2. *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. Furthermore, we may say

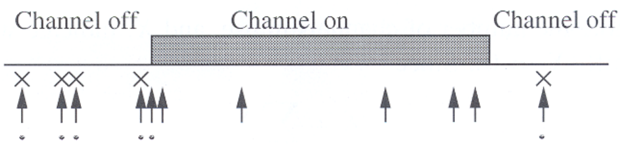


Figure 3. Call vs. channel blocking. Call attempts are represented by arrows, blocking events by crosses. The call attempts that are also attempts to turn a channel on are marked with dots. Call blocking in this trace is 5/11 and channel blocking 5/6.

that channel blocking probability equals call blocking probability conditioned to the fact that the channel is off. See figure 3. 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,

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

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

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

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 $\mathrm{M}/\mathrm{M}/m/m/K$ system with a finite user population, see [9]. 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 $\lambda_i^{-1}$. 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 $\pi_j$ 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 [1]. In the general case, it can be shown to follow from the insensitivity property of the product form probabilities of a multirate loss system [4]. 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.

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 $a_i$ ($b_i^c \approx B_i^c \mathrm{e}^{-a_i}$). For a channel with $a_i \ll 1$, the channel blocking and the call blocking seen by a user are approximately the same.

Since $b_i^c \leqslant B_i^c \leqslant B_i^t$, 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 approaches time blocking for channels with channel preferences $\alpha_i$ near zero.

So far we have considered only a single link. The result, equation (17), depends on the vector of traffic intensities offered to the link, **a**, vector of capacity requirements for channels, **c**, and link capacity $C$. In the sequel, we use the complete notation $b_i^c[\mathbf{a}, \mathbf{c}, C]$ to make these dependencies explicit. This result is exact, given the assumptions. In the next section, we utilise this exact result when calculating end-to-end blocking probabilities.

## 3. End-to-end blocking probabilities

The well known Reduced Load Approximation (RLA) (see, e.g., [11]) is used to calculate end-to-end blocking probabilities. The idea of the RLA is that traffic that is blocked in a link on its route does not affect the other links. The approximation yields fairly good results when no traffic class dominates and the total traffic intensity is not too high. Under these conditions, traffic in different links is nearly independent and Poissonian.

RLA consists of two alternating steps. First, assuming the call blocking probabilities for all channels and links known, we may calculate the traffic intensity for channel $i$ in link $j$. This is done by summing all traffic intensities for that channel from the leaves downstream from the link. These leaf traffic intensities are thinned with the call blocking associated with them in each link along the route (from the leaf to the root node), with the exception of the link we are calculating blocking for. Second, knowing the traffic intensities for each channel $i$ in link $j$, we may calculate call blocking for each channel in the link.

The process is iterated to give the approximate blocking probabilities. The iteration starts with zero blocking for each channel in each link.

To be more precise, let $R_u$ denote the route between the root node and the user population $u$ in a leaf of the distribution tree. Further, let $L_j^i$ denote the call blocking probability of channel $i$ in link $j$. We assume that the links behave independently, or that the blocking for each link is independent so that we may calculate the probability that a call is not blocked on the route by multiplying the corresponding probabilities of individual links. Let $a_{i,u}$ denote the traffic intensity offered by user population $u$ for channel $i$.

The blocking probabilities $L_j^i$ are calculated as stated in section 2,

$$L_j^i = b_i^c[\mathbf{r}_j, \mathbf{c}, C_j], \tag{21}$$

where the elements of the vector $\mathbf{r}_j$, $r_{j,i}$ represent the thinned traffic intensities for corresponding channels in the link $j$. They are calculated as follows:

$$r_{j,i} = \sum_{u \in U_j} a_{i,u} \prod_{k \in R_u - \{j\}} (1 - L_k^i) = \frac{\sum_{u \in U_j} a_{i,u} \prod_{k \in R_u} (1 - L_k^i)}{(1 - L_j^i)}, \tag{22}$$

where $U_j$ denotes the set of user populations downstream of link $j$.

The last form of equation (22) makes the calculation of the thinned traffic intensities more efficient in large networks. Traffic intensities are calculated for each leaf node and each channel by thinning the offered intensities with call blocking for all links along the route. The thinned traffic for a link and channel is then given by the sum of these thinned traffic intensities divided by the link's own thinning factor.

Equations (21) and (22) form a fixed point equation of the form $\mathbf{L} = \mathbf{T}(\mathbf{L})$. The solution, a vector of $L_j^i$, may be found by repeated substitution. Note, however, that the equation might have several solutions. (For the traditional RLA used for point-to-point calls, the solution is not always unique if capacity requirements of different traffic classes differ, see, e.g., [11].) After the values for $L_j^i$ have been found, the end-to-end blocking probabilities $B_u^i$ for user $u$ and channel $i$ are calculated as follows:

$$B_u^i = 1 - \prod_{k \in R_u} (1 - L_k^i), \tag{23}$$

which follows directly from the assumption of independency.

## 4. Simulation

In order to find out the accuracy of the reduced load approximation in this multicast context, several examples were studied by both calculations using RLA and by simulations. The comparisons were carried out for two different network settings and several traffic intensities. We restricted ourselves to the case where capacity requirements $c_i$ of different channels are identical, $c_i = 1$ for all $i$, except for network (b), where $c_i = 1$ if $i$ is odd, and $c_i = 2$ if $i$ is even.

The networks considered are presented in figures 4 and 5. All user populations u behave identically offering the network traffic with total intensity $a$. The total traffic intensity is distributed to different channels according to a truncated geometric preference distribution. The parameter of the distribution is chosen to be $p = 0.2$ and the number of channels $|I| = 30$. For an intuitive interpretation of the traffic load, we calculated the offered traffic load, $\mathrm{E}[Y]/C$, for every link on the routes and provide the maximum and minimum of these figures as "min avg" and "max avg" for each route in the tables. We examined only the blocking probability of the least used channel, which is a natural candidate for network dimensioning, since its call blocking probability is the highest. The channel viewing time is the same for all channels, $1/\mu = 1$.

For each run, we simulated $10^{10}$ calls per user population. Of these calls, there was an average of $10^{10} \cdot 0.2 \cdot 0.8^{29}/(1 - 0.8^{30}) \approx 3.1 \cdot 10^6$ calls for the least loaded channel. Even for systems as small as these, the simulation task becomes an excessive effort. One simulation took about 3–4 days to carry out in a Pentium II 233 MHz machine running Linux. No acceleration method was used, but clearly there is a need for such for making the simulation faster.

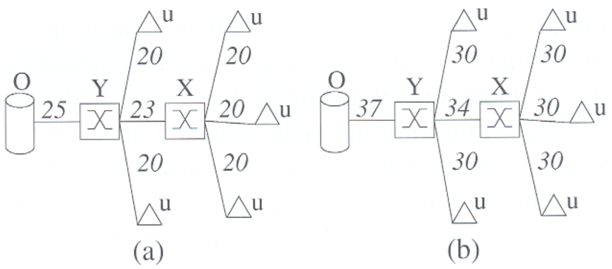


Figure 4. The example networks. Link capacities (channels per link) are shown in italics.

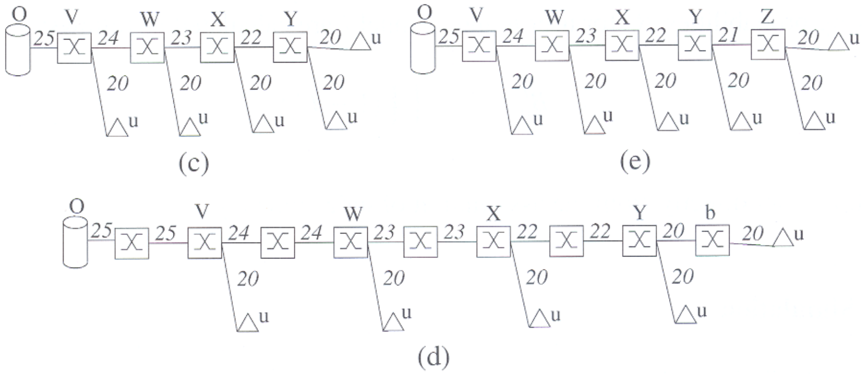


Figure 5. The example networks for link dependency studies. Link capacities (channels per link) are shown in italics.

### 4.1. *Tree-type network*

We simulated traffic in two tree-type networks, see figure 4. In these networks, the routes are short and the networks are nearly symmetric, thus one would expect the RLA be rather accurate.

The network (a) was offered traffic with the capacity requirement $c_i = 1$ for each channel. The network (b) was offered similar traffic, but the capacity requirement for channels with an even channel number was $c_i = 2$ and for odd channels $c_i = 1$.

The results for the two networks are shown in tables 1 and 2. Several intensity levels and routes are considered. For each intensity level and route, the tables show the 95% confidence interval for the simulated result, the result of RLA, and the error defined as the relative deviation of the calculated result from the simulated one.

The approximation is seen to be reasonable for networks (a) and (b). As with using RLA with unicast traffic, the accuracy of the approximation declines as the traffic increases. This was to be expected since, with low traffic intensity, blocking is rare, and the traffic in each link is closer to Poisson traffic. When traffic intensity increases, blocking starts to shape traffic, which reduces real blocking. However, the model for single link blocking does not take this into account and gives higher blocking probabilities.

Table 1
Simulation results for network (a).

| Intensity $a$ | Route | Link occupation min avg (%) | Link occupation max avg (%) | $b_c^i$ simulated | $b_c^i$ RLA | Error (%) |
|---|---|---|---|---|---|---|
| 50 | O-Y-X-u | 67 | 81 | (6.85e−3,7.11e−3) | 8.00e−3 | 15 |
| | O-Y-u | 67 | 81 | (3.55e−3,3.76e−3) | 4.06e−3 | 11 |
| 60 | O-Y-X-u | 71 | 84 | (1.98e−2,2.03e−2) | 2.38e−2 | 19 |
| | O-Y-u | 71 | 84 | (1.04e−2,1.08e−2) | 1.21e−2 | 14 |
| 75 | O-Y-X-u | 76 | 88 | (5.63e−2,5.71e−2) | 7.08e−2 | 25 |
| | O-Y-u | 76 | 88 | (3.05e−2,3.11e−2) | 3.72e−2 | 21 |
| 100 | O-Y-X-u | 82 | 93 | (1.48e−1,1.49e−1) | 1.97e−1 | 33 |
| | O-Y-u | 82 | 93 | (8.32e−2,8.42e−2) | 1.11e−1 | 33 |

Table 2
Simulation results for network (b) where for even channels capacity requirement $c_i = 2$ and for odd channels $c_i = 1$.

| Intensity $a$ | Route | Link occupation min avg (%) | Link occupation max avg (%) | ch # | $b_c^i$ simulated | $b_c^i$ RLA | Error (%) |
|---|---|---|---|---|---|---|---|
| 50 | O-Y-X-u | 66 | 82 | 29 | (8.60e−3,8.89e−3) | 1.03e−2 | 18 |
| | | | | 30 | (2.64e−2,2.69e−2) | 3.19e−2 | 20 |
| | O-Y-u | 66 | 82 | 29 | (4.56e−3,4.73e−3) | 5.34e−3 | 15 |
| | | | | 30 | (1.50e−2,1.54e−2) | 1.65e−2 | 9 |
| 60 | O-Y-X-u | 70 | 85 | 29 | (2.09e−2,2.13e−2) | 2.55e−2 | 21 |
| | | | | 30 | (6.02e−2,6.10e−2) | 7.44e−2 | 23 |
| | O-Y-u | 70 | 85 | 29 | (1.09e−2,1.12e−2) | 1.33e−2 | 20 |
| | | | | 30 | (3.43e−2,3.49e−2) | 3.92e−2 | 13 |
| 75 | O-Y-X-u | 75 | 87 | 29 | (4.96e−2,5.02e−2) | 6.27e−2 | 26 |
| | | | | 30 | (1.33e−1,1.34e−1) | 1.69e−1 | 27 |
| | O-Y-u | 75 | 87 | 29 | (2.57e−2,2.61e−2) | 3.31e−2 | 28 |
| | | | | 30 | (7.78e−2,7.88e−2) | 9.26e−2 | 18 |
| 100 | O-Y-X-u | 81 | 93 | 29 | (1.11e−1,1.12e−1) | 1.46e−1 | 31 |
| | | | | 30 | (2.71e−1,2.73e−1) | 3.55e−1 | 31 |
| | O-Y-u | 81 | 93 | 29 | (5.75e−2,5.81e−2) | 8.09e−2 | 40 |
| | | | | 30 | (1.63e−1,1.64e−1) | 2.12e−1 | 30 |

### 4.2. *Link dependency*

To investigate further validity of the link independency assumption, we examined another three network scenarios. These results are presented in this section.

The networks (c)–(e) represent a parking-lot configuration. They have a long route to which short branches are added. Network (d) is a variant of network (c) in which no user populations have been added, but part of the links are replaced by two links in series. Obviously, the true blocking probabilities do not change, but the RLA algorithm produces different results. Network (e) is a parking-lot network which has one node and one user population more than network (c).

Table 3
Simulation results for network (c).

| Intensity $a$ | Route | Link occupation min avg (%) | Link occupation max avg (%) | $b_c^i$ simulated | $b_c^i$ RLA | Error (%) |
|---|---|---|---|---|---|---|
| 50 | O-V-W-X-Y-u | 67 | 81 | (9.16e−3,9.45e−3) | 1.40e−2 | 51 |
| | O-V-W-X-u | 67 | 81 | (8.51e−3,8.78e−3) | 1.28e−2 | 48 |
| | O-V-W-u | 67 | 81 | (6.24e−3,6.48e−3) | 8.88e−3 | 40 |
| | O-V-u | 67 | 81 | (3.03e−3,3.20e−3) | 4.01e−3 | 29 |
| 60 | O-V-W-X-Y-u | 71 | 84 | (2.53e−2,2.58e−2) | 4.07e−2 | 59 |
| | O-V-W-X-u | 71 | 84 | (2.37e−2,2.42e−2) | 3.65e−2 | 53 |
| | O-V-W-u | 71 | 84 | (1.71e−2,1.76e−2) | 2.54e−2 | 47 |
| | O-V-u | 71 | 84 | (8.91e−3,9.21e−3) | 1.18e−2 | 30 |
| 75 | O-V-W-X-Y-u | 76 | 88 | (6.95e−2,7.04e−2) | 1.14e−1 | 63 |
| | O-V-W-X-u | 76 | 88 | (6.37e−2,6.45e−2) | 1.01e−1 | 58 |
| | O-V-W-u | 76 | 88 | (4.69e−2,4.76e−2) | 7.12e−2 | 51 |
| | O-V-u | 76 | 88 | (2.57e−2,2.62e−2) | 3.50e−2 | 35 |
| 100 | O-V-W-X-Y-u | 82 | 93 | (1.74e−1,1.75e−1) | 2.86e−1 | 64 |
| | O-V-W-X-u | 82 | 93 | (1.59e−1,1.60e−1) | 2.52e−1 | 58 |
| | O-V-W-u | 82 | 93 | (1.18e−1,1.19e−1) | 1.83e−1 | 55 |
| | O-V-u | 82 | 93 | (6.95e−2,7.03e−2) | 1.00e−1 | 43 |

Table 4
Simulation results for network (d).

| Intensity $a$ | Route | Link occupation min avg (%) | Link occupation max avg (%) | $b_c^i$ simulated | $b_c^i$ RLA | Error (%) |
|---|---|---|---|---|---|---|
| 50 | O-V-W-X-Y-b-u | 67 | 81 | (9.16e−3,9.45e−3) | 2.72e−2 | 192 |
| | O-V-W-X-Y-u | 67 | 81 | (9.16e−3,9.45e−3) | 2.70e−2 | 190 |
| | O-V-W-X-u | 67 | 81 | (8.51e−3,8.78e−3) | 2.45e−2 | 183 |
| | O-V-W-u | 67 | 81 | (6.24e−3,6.48e−3) | 1.70e−2 | 167 |
| | O-V-u | 67 | 81 | (3.03e−3,3.20e−3) | 7.60e−3 | 144 |
| 60 | O-V-W-X-Y-b-u | 71 | 84 | (2.53e−2,2.58e−2) | 7.50e−2 | 194 |
| | O-V-W-X-Y-u | 71 | 84 | (2.53e−2,2.58e−2) | 7.42e−2 | 190 |
| | O-V-W-X-u | 71 | 84 | (2.37e−2,2.42e−2) | 6.66e−2 | 178 |
| | O-V-W-u | 71 | 84 | (1.71e−2,1.76e−2) | 4.63e−2 | 167 |
| | O-V-u | 71 | 84 | (8.91e−3,9.21e−3) | 2.13e−2 | 135 |
| 75 | O-V-W-X-Y-b-u | 76 | 88 | (6.95e−2,7.04e−2) | 1.92e−1 | 174 |
| | O-V-W-X-Y-u | 76 | 88 | (6.95e−2,7.04e−2) | 1.88e−1 | 169 |
| | O-V-W-X-u | 76 | 88 | (6.37e−2,6.45e−2) | 1.67e−1 | 161 |
| | O-V-W-u | 76 | 88 | (4.69e−2,4.76e−2) | 1.18e−1 | 150 |
| | O-V-u | 76 | 88 | (2.57e−2,2.62e−2) | 5.71e−2 | 120 |
| 100 | O-V-W-X-Y-b-u | 82 | 93 | (1.74e−1,1.75e−1) | 4.14e−1 | 137 |
| | O-V-W-X-Y-u | 82 | 93 | (1.74e−1,1.75e−1) | 4.05e−1 | 132 |
| | O-V-W-X-u | 82 | 93 | (1.59e−1,1.60e−1) | 3.60e−1 | 126 |
| | O-V-W-u | 82 | 93 | (1.18e−1,1.19e−1) | 2.64e−1 | 123 |
| | O-V-u | 82 | 93 | (6.95e−2,7.03e−2) | 1.41e−1 | 102 |

Table 5
Simulation results for network (e).

| Intensity | Route | Link occupation | | $b_c^i$ | | Error |
|---|---|---|---|---|---|---|
| $a$ | | min avg (%) | max avg (%) | simulated | RLA | (%) |
| 50 | O-V-W-X-Y-Z-u | 67 | 84 | (3.83e−2,3.88e−2) | 6.67e−2 | 73 |
| | O-V-W-X-Y-u | 67 | 84 | (3.58e−2,3.64e−2) | 6.10e−2 | 69 |
| | O-V-W-X-u | 67 | 84 | (2.78e−2,2.83e−2) | 4.59e−2 | 64 |
| | O-V-W-u | 67 | 84 | (1.64e−2,1.69e−2) | 2.69e−2 | 62 |
| | O-V-u | 67 | 84 | (6.75e−3,7.00e−3) | 1.06e−2 | 54 |
| 60 | O-V-W-X-Y-Z-u | 71 | 87 | (8.37e−2,8.46e−2) | 1.45e−1 | 72 |
| | O-V-W-X-Y-u | 71 | 87 | (7.74e−2,7.83e−2) | 1.32e−1 | 70 |
| | O-V-W-X-u | 71 | 87 | (5.91e−2,5.99e−2) | 9.96e−2 | 67 |
| | O-V-W-u | 71 | 87 | (3.53e−2,3.60e−2) | 6.00e−2 | 68 |
| | O-V-u | 71 | 87 | (1.54e−2,1.58e−2) | 2.49e−2 | 60 |
| 75 | O-V-W-X-Y-Z-u | 76 | 91 | (1.74e−1,1.76e−1) | 2.91e−1 | 66 |
| | O-V-W-X-Y-u | 76 | 91 | (1.59e−1,1.60e−1) | 2.62e−1 | 64 |
| | O-V-W-X-u | 76 | 91 | (1.20e−1,1.21e−1) | 2.02e−1 | 68 |
| | O-V-W-u | 76 | 91 | (7.29e−2,7.39e−2) | 1.26e−1 | 72 |
| | O-V-u | 76 | 91 | (3.41e−2,3.48e−2) | 5.64e−2 | 64 |
| 100 | O-V-W-X-Y-Z-u | 82 | 95 | (3.29e−1,3.30e−1) | 5.12e−1 | 55 |
| | O-V-W-X-Y-u | 82 | 95 | (2.94e−1,2.96e−1) | 4.63e−1 | 57 |
| | O-V-W-X-u | 82 | 95 | (2.19e−1,2.21e−1) | 3.69e−1 | 68 |
| | O-V-W-u | 82 | 95 | (1.37e−1,1.39e−1) | 2.47e−1 | 79 |
| | O-V-u | 82 | 95 | (7.28e−2,7.36e−2) | 1.25e−1 | 71 |

The results of the simulations are shown in tables 3–5. The error of RLA in networks (c) and (e) is seen to grow when the length of the route becomes longer. This can be traced back to the fact that the traffic processes in the different links are not independent, since the traffic on the route O-V-W-X-Y is dominating. That is, the links O-V and V-W, for example, carry almost the same set of connections. Thus, the RLA's assumption of independence is not valid, introducing an error to the calculations.

Network (d) gives further evidence to this. For example, the two links between V and W, have the same capacity and exactly the same calls at all times. Thus, they both always introduce blocking at the same time. The links are definitely not independent. Here, the RLA gives really bad performance for all traffic loads. In this sense, the network considered is nearly the worst case for the approximation.

The results show that the RLA yields results that are in the best cases almost equal and in the worst cases of the same magnitude as the true blocking probabilities. Since the blocking probabilities are usually "steep" functions of link capacity, an order of magnitude in blocking probabilities does not correspond to an excessive amount of link capacity. The estimated call blocking probabilities are also greater than the real ones, so that the approximation is conservative. We thus conclude that the RLA yields accurate enough results for most practical needs.

## 5. Summary

In this paper, we have presented a method for calculating the end-to-end call blocking probabilities for a network carrying heterogeneous 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. First, a method for calculating exact blocking probabilities for a single link was reviewed. Then, the reduced load approximation for multicast traffic was presented. We studied the accuracy of the approximation by simulating several network scenarios and comparing the results to the calculated ones. We concluded that the reduced load approximation yields accurate enough approximations for most practical needs. However, if the multicast trees are long and do not have much diversity, the RLA gives overly pessimistic blocking probabilities.

We leave for the future work improving the accuracy of the approximation, as well as calculation of bounds for the blocking probabilities. Also, fast simulation methods need to be studied to allow accurate results to be achieved for complex networks.

## Acknowledgements

The authors are grateful to the anonymous reviewers for their several suggestions for improving the paper.

## References

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

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

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

[4] J. Hui, *Switching and Traffic Theory for Integrated Broadband Networks* (Kluwer Academic Publishers, Boston, 1990).

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

[6] J. Karvo, J. Virtamo, S. Aalto and O. Martikainen, Blocking of dynamic multicast connections in a single link, in: *Proc. of BC '98: The Future of Telecommunications*, Stuttgart, Germany (1998) pp. 473–483.

[7] J.S. Kaufman, Blocking in a shared resource environment, IEEE Transactions on Communications 29(10) (1981) 1474–1481.

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

[9] L. Kleinrock, *Queueing Systems; Volume 1: Theory* (Wiley, New York, 1975).

[10] J.W. Roberts, 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 (North-Holland, Amsterdam, 1981) pp. 423–431.
[11] K.W. Ross, *Multiservice Loss Models for Broadband Telecommunication Networks* (Springer-Verlag, London, 1995).
[12] N. Shacham and H. Yokota, Admission control algorithms for multicast sessions with multiple streams, IEEE Journal on Selected Areas in Communications 15(3) (1997) 557–566.