

An analysis of retransmission strategies for reliable multicast protocols

M. Schuba, P. Reichl
Informatik 4, Aachen University of Technology
52056 Aachen, Germany
email: marko|peter@i4.informatik.rwth-aachen.de

Abstract

In this paper we present an analytical comparison of retransmission strategies for three types of reliable multicast protocols. In the first protocol (called sender-originated) it is the task of the source to guarantee reliable delivery of data to all receivers. The other two approaches additionally allow either receivers or routers to retransmit packets (these protocols are termed receiver-originated and router-originated, resp.). Thus the load on the sender is relieved while at the same time transmission cost decreases since packets can be retransmitted with a limited scope. We determine the retransmission cost of all three protocol types based on discrete-time Markov chains. In contrast to other analytical models this approach does not require loss events to be independent for different receivers. Numerical results show that the cost for sender-originated protocols may become unacceptable even if loss probabilities within the network are small. In larger networks with widely spread groups the performance of receiver-originated protocols is also very limited. In such scenarios only router-originated protocols yield acceptable cost.

Keywords

Reliable multicast, performance analysis, Markov chain

1 INTRODUCTION

The widely spread introduction of multicast protocols (e.g. IP multicast (Deering, 1988), (Hermanns and Schuba, 1996)) has led to an increasing interest in groupware applications. An important feature required by such applications is the support of reliable data transfer to a group of receivers, for instance for realizing distributed simulation, updating software on a large number of computers, pushing web pages, distributing learning materials in a distance learning environment or news in a news group. To achieve reliability some error recovery mechanism for lost packets has to be implemented. In protocols designed for small (local area) multicast groups these mechanisms are usually realized at the sender, which is responsible for processing positive or negative acknowledgements (ACKs/NAKs) and for retransmitting packets. Examples for such “sender-originated” reliable multicast protocols are MTP (Armstrong et al., 1992) or AMTP (Heinrichs, 1994). However, with increasing size or geographic spread of the multicast group the performance of these protocols gets worse, and more scalable protocols are required. Only recently some reliable multicast protocols have been developed for that purpose. Besides the sender these protocols allow either dedicated receivers (e.g. in RMTP (Paul et al., 1997) or TMTP Yavatkar et al., 1995) or routers (e.g. in SRMT (Schuba, 1998)) to handle ACKs/NAKs and to retransmit packets for members in their local environment (we will call these protocols “receiver-originated” and “router-originated”, respectively). Thus the cost for retransmissions can be decreased and the ACK processing load on the sender is relieved. Moreover, by a hierarchical ordering of retransmitting nodes the work can be homogeneously distributed in the network.

In the past performance evaluations of reliable multicast protocols were usually based on measurements or simulations of a single protocol. There are only some exceptions to this rule, e.g. the analytical investigations in (Pingali et al., 1994) (which are restricted to different types of sender-originated strategies) or (Schuba, 1998). However, these analyses make simplifying assumptions, e.g. they assume loss probabilities to be identical and loss events to be mutually independent for all receivers. Our approach differs from previous work in several ways. First, we compare generic protocols with respect to their retransmission strategy, i.e. sender-originated, receiver-originated and router-originated protocols. Second, our analysis is based on homogeneous discrete-time Markov chains. Hence, loss events are allowed to be dependent for different receivers, and loss probabilities can be defined separately in the multicast tree and thus may be different for all destinations.

The remainder of the paper is structured as follows. In chapter 2 we motivate the demand for new analytical models by demonstrating in a simple example that the independence assumption of loss events may lead to significant errors in analytical results. Next we describe how Markov chains can be used for a more realistic modelling of retransmission strategies (chapter 3). We then apply this model for an example multicast tree and different retransmission strategies in chapter 4, 5 and 6, respectively. The numerical results of our analysis are presented in chapter 7. Finally, concluding remarks are given in chapter 8.

2 DEPENDENT VS. INDEPENDENT LOSS EVENTS

First, we want to demonstrate that investigations of reliable multicast protocols which are based on the simplifying assumption of independent loss events for all receivers in a multicast group, may lead to a significant overestimation of the expected number of retransmissions. This error is due to the tree-based transmission of multicast data, where packets are duplicated only at branching points of the delivery tree (see fig. 1). If a link in the tree has a positive packet loss probability and this link leads to more than one receiver, the packet loss events that happen on this link are obviously dependent for its receivers. Because a large number of links in a multicast tree transport packets to several receivers (bold links in fig. 1 left), the independence assumption may lead to serious errors in analytical results.

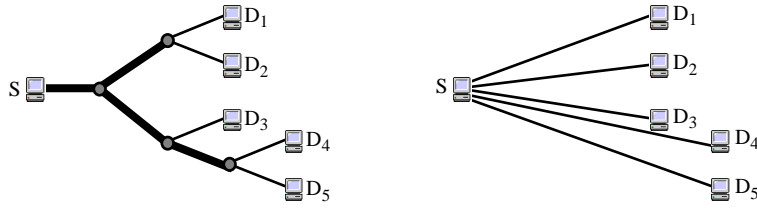


Figure 1 Real multicast tree vs. “tree” with independent loss events.

For a comparison of dependent and independent loss events let the random variable X denote the number of retransmissions until all receivers in a multicast tree have obtained a packet. Assuming independent loss events (cf. fig. 1 right) and a loss probability p for each of the D destinations the expected number of retransmissions $E[X]$ (see (Pingali et al., 1994) for details) is given by

$$E[X] = \left[\sum_{i=1}^D \binom{D}{i} (-1)^{i+1} \frac{1}{(1-p^i)} \right] - 1. \quad (1)$$

We will use this formula to calculate $E[X]$ for multicast groups with 5, 10 and 20 destinations.

Now imagine a multicast tree with only one unreliable link. If this link happens to be located between the source and the first branching of the tree (e.g. the first link in fig. 1 left), all receivers (independently of their number) will be affected by a lost packet on this link. Thus the packet loss events for all receivers are strongly dependent. The number of retransmissions in such a multicast tree is geometrically distributed with mean $E[X] = p / (1 - p)$ (what may also be computed using (1) with $D = 1$).

Fig. 2 compares the expected number of retransmissions for the four different scenarios. Note that the loss probability for each receiver is always the same.

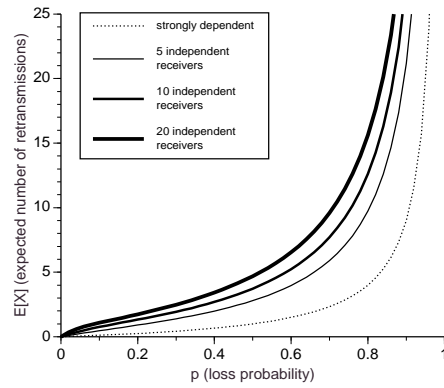


Figure 2 Independent vs. strongly dependent loss probabilities.

As can be seen clearly, the overestimation increases with the number of receivers, for which independent loss events are assumed. If we consider the worst (real) case, i.e. all loss events are dependent (dotted line), the values of $E[X]$ may be overestimated by more than factor 2 even for only 5 receivers. Thus, when assuming independent loss events, all results should be interpreted very carefully.

To overcome the overestimation of retransmissions in our analyses we will use a Markov chain model with only the following assumptions:

- All retransmissions are sent as multicast to the receivers.
- ACKs or NAKs are never lost.

3 MODELLING OF MULTICAST RETRANSMISSION STRATEGIES

We first need an appropriate cost metric for the comparison of sender-originated, receiver-originated and router-originated protocols. From our point of view a reliable multicast protocol should minimize transmission time as well as the required network resources. Since the time until all receivers successfully have received a data packet is closely related to the number of retransmissions we use retransmission cost as measure for our investigations. Let the random variable X_k denote the number of retransmissions of a node k . Then we define the retransmission cost produced by this node as

- the average number of times $E[X_k]$ a packet is retransmitted by node k until all its receivers have obtained the packet
multiplied by
- the number of multicast tree links c_k (cost parameter) involved in a retransmission of node k .

The overall retransmission cost R for a packet can be written as

$$R := \sum_{\text{all } k} E[X_k] \cdot c_k. \quad (2)$$

For example consider the multicast tree in fig 3. Obviously a retransmission from a node near the receivers yields less cost than a retransmission from a node farther away. For instance each retransmission multicasted by sender S to D_1 and D_2 results in additional retransmission cost of $c = 3$ (cost per link = 1) while a retransmission multicasted from router MR to D_1 and D_2 only leads to retransmission cost of $c = 2$.

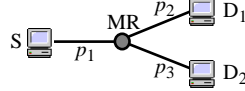


Figure 3 A simple multicast tree.

Now look once more at the multicast tree given in fig 3. Let the variables p_i ($i \in \{1, 2, 3\}$) denote the packet loss probabilities of the respective links. We first examine a sender-originated protocol. The probability distribution function of X (number of retransmissions) can be determined based on a homogeneous discrete-time Markov chain (DTMC) shown in fig. 4.

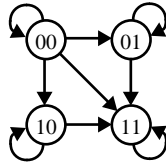


Figure 4 DTMC for the simple multicast tree.

The DTMC consists of four two-bit states d_1d_2 , $d_i \in \{0, 1\}$, $i \in \{1, 2\}$. $d_i = 1$ means that the transmission for destination D_i has been a success while $d_i = 0$ indicates a failure. The corresponding transition probabilities* can be written as matrix

$$P = \begin{pmatrix} p_1 + (1-p_1)p_2p_3 & (1-p_1)(1-p_2)p_3 & (1-p_1)p_2(1-p_3) & (1-p_1)(1-p_2)(1-p_3) \\ 0 & p_1 + (1-p_1)p_3 & 0 & (1-p_1)(1-p_3) \\ 0 & 0 & p_1 + (1-p_1)p_2 & (1-p_1)(1-p_2) \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

*. We will also use the binary notation for transition probabilities, e.g. $p_{01,11}$. The state order is given by the decimal value of the binary state number, e.g. the binary state 11 equals state number 3 (decimal).

The initial state distribution of the DTMC is given by $\pi(0) = (\pi_{00}(0), \pi_{01}(0), \pi_{10}(0), \pi_{11}(0)) = (p_{00,00}, p_{00,01}, p_{00,10}, p_{00,11})$, since one transmission is already finished before retransmissions begin, i.e. we just have to consider the transition probabilities originating from state 00. The probability to be in a certain state of the DTMC after n steps $\pi(n)$ can be determined simply by calculation of the n th power of P (see e.g. (Stewart, 1994)). We are only interested in being in state 11, which indicates that all receivers have obtained the packet. Hence, we get

$$\pi_{11}(n) = \sum_{\text{all } d_i d_j} p_{d_i d_j, 11}^{(n)} \cdot \pi_{11}(0) \quad (3)$$

where we take the n -step transition probabilities $p_{d_i d_j, 11}^{(n)}$ from matrix P^n . The probability f_n that exactly n retransmissions are required is given by $f_n = \pi_{11}(n) - \pi_{11}(n-1)$ for $n > 0$ and $f_n = \pi_{11}(0)$ for $n = 0$. The expected retransmission cost for a sender-originated protocol now can be written as

$$E[R] = E[X] \cdot c = c \cdot \sum_{n=0}^{\infty} f_n \cdot n. \quad (4)$$

4 ANALYSIS OF SENDER-ORIGINATED RETRANSMISSIONS

For the comparison of the three retransmission strategies we use the multicast tree with five receivers shown in fig. 5.

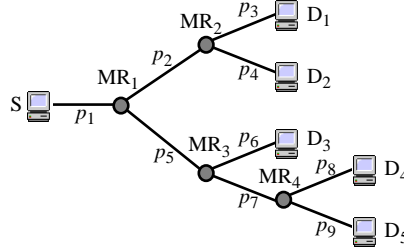


Figure 5 Multicast scenario to be analyzed.

We limit the group size to five because an analysis with i receivers results in 2^i states in the DTMC and 2^{2i} transition probabilities, i.e. 32 states and 1024 probabilities in our example. This multicast tree is simple but very useful for analyzing many different multicast scenarios, e.g. by changing the loss probabilities of the links. Thus local groups (small p_i) can be examined as well as groups with large geographic spread, i.e. large p_i values. Moreover, the example tree corresponds to typical structures of real multicast groups.

For example look at the map of the German MBONE (Multicast Backbone, fig. 6) in which all german IP multicast routers together with their logical connections are shown. The logical interconnection of our example tree (fig. 5) can be found very often in this map, e.g. if the sender is located in Aachen, two receivers are listening in Berlin, another receiver is located in Stuttgart and two more receivers are connected to the multicast router in Nürnberg.

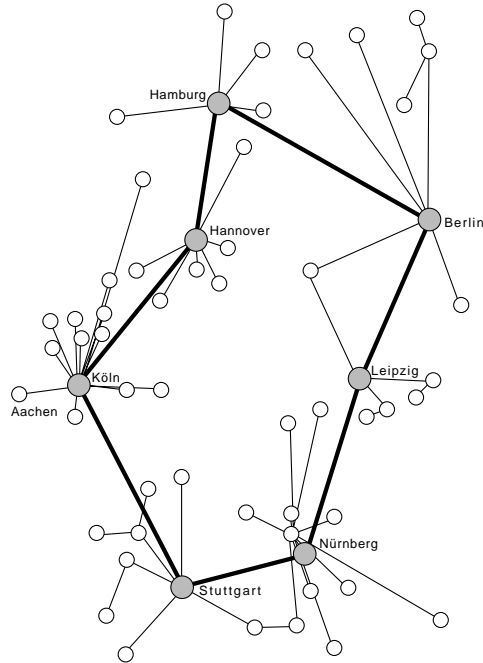


Figure 6 German MBONE (December 1997).

The analysis of the sender-originated strategy can be performed in the same way as in the example for two receivers, now numbering the states with bits $d_1d_2d_3d_4d_5$, $d_i \in \{0, 1\}$, $i \in \{1, \dots, 5\}$ where $d_i = 0$ (1) corresponds to the success (failure) of D_i . Because the corresponding matrix P is too large to be printed here we just give the following transition probabilities as example. All other probabilities look similarly.

$$\begin{aligned}
 p_{00000,00000} &= p_1 + (1 - p_1) \cdot [p_2 + (1 - p_2)p_3p_4] \cdot [p_5 + (1 - p_5)p_6(p_7 + (1 - p_7)p_8p_9)] \\
 p &= (1 - p)(1 - p)(1 - p)p[p + (1 - p)(p + (1 - p)p)]
 \end{aligned}$$

The expected overall retransmission cost can be directly derived from (4) with parameter c set to 9 because there are 9 links in the multicast tree originated at the source.

5 ANALYSIS OF RECEIVER-ORIGINATED RETRANSMISSIONS

For the receiver-originated approach we first have to divide the multicast group into subgroups. Let us assume that a receiver-originated protocol would use three (local) subgroups in our example (see fig. 7):

- Group₁ consisting of S, D₂ and D₃ (the source being responsible for retransmissions),
- Group₂ with D₁ and D₂ (D₂ being the dedicated receiver) and
- Group₃ consisting of D₃, D₄ and D₅ (D₃ being the dedicated receiver).

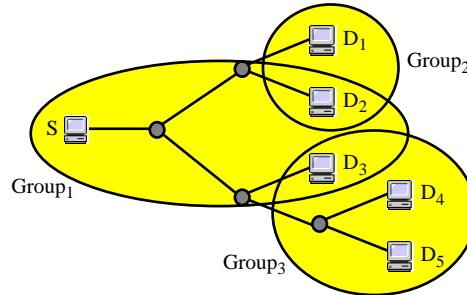


Figure 7 Local groups for receiver-originated retransmissions.

While source S globally multicasts packets to all receivers (but listens only to ACKs from D₂ and D₃), we assume that retransmissions in Group₁ and Group₂ are multicasted only to the members of the respective subgroup (local multicast).

To analyze this scenario let us first look at Group₁. The sender S retransmits a packet until D₂ and D₃ have correctly received it. This is just like a sender-originated protocol with two receivers. Thus the number of retransmissions can be computed according to (3). For the expected cost $E[R_{\text{Group}_1}]$ (defined in (4)) we have to set the cost parameter c to 9 because each retransmission is sent as global multicast.

For Group₂ the situation is different. Obviously there exist three different retransmission situations in which D₁ might receive a packet for the first time:

1. The packet is retransmitted by S only, i.e. D₂ (the dedicated receiver for D₁) has not yet received the packet.
2. S and D₂ retransmit the packet, i.e. D₂ has received the packet but D₃ has not.
3. Only D₂ retransmits this packet, i.e. both, D₂ and D₃, have received the packet.

Since all retransmissions of S (case 1 and 2) have already been taken into account for the cost of Group₁ we only have to consider the retransmissions from D₂ (case 2 and 3). Once more we can use a Markov chain (we call it DTMC₁) to determine the number of retransmissions from D₂ (see fig. 8 left). As before the bits in the states

indicate the reception status of the receivers D_1 , D_2 and D_3 , state 010 corresponding to case 2 and 011 to case 3, resp. State 1xx is the success state, i.e. there is no more retransmission required for D_1 , independently of D_2 and D_3 .

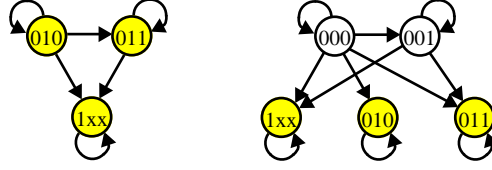


Figure 8 DTMC₁ (left) and DTMC₂ (right) for Group₂.

The (non-null) transition probabilities for DTMC₁ are given by

$$\begin{aligned}
 p_{010,010} &= (p_1 + (1-p_1) \cdot [p_2 + (1-p_2) \cdot p_3]) \cdot (p_5 + (1-p_5) \cdot p_6) \cdot (p_4 + (1-p_4) \cdot p_3), \\
 p_{010,011} &= (1-p_1) \cdot (p_2 + (1-p_2) \cdot p_3) \cdot (1-p_5) \cdot (1-p_6) \cdot (p_4 + (1-p_4) \cdot p_3), \\
 p_{010,1xx} &= (1-p_1) \cdot (1-p_2) \cdot (1-p_3) + (p_1 + (1-p_1) \cdot [p_2 + (1-p_2) \cdot p_3]) \cdot (1-p_4) \cdot (1-p_3), \\
 p_{011,011} &= p_4 + (1-p_4) \cdot p_3, \\
 p_{011,1xx} &= (1-p_4) \cdot (1-p_3), \\
 p_{1xx,1xx} &= 1.
 \end{aligned}$$

Note that $p_{010,010}$, $p_{010,011}$ and $p_{010,1xx}$ include retransmissions from S and D_1 . DTMC₁ might start in each of the states depending on the success/failures of the transmissions initiated by the sender before. In order to get the initial state distribution $\pi(0) = (\pi_{010}(0), \pi_{011}(0), \pi_{1xx}(0))$ we need Markov chain DTMC₂, (see fig. 8 right), modelling the possible transitions from the very start to the states of DTMC₁. The limiting distribution of DTMC₂ determines for each of the three states 1xx, 010 and 011 the probability that it is reached first, i.e. the initial state for DTMC₁.

Given the initial state distribution we can determine $\pi_{1xx}(n)$ for DTMC₁ similarly to (3) and the expected retransmission cost $E[R_{\text{Group2}}]$ according to (4) with $c = 2$ (due to local multicasting of packets).

To achieve the expected retransmission cost for Group₃ $E[R_{\text{Group3}}]$ we proceed similarly to Group₂. The respective Markov chains become a little bit larger since there are three receivers in Group₃. The overall retransmission cost for receiver-originated retransmissions are given by $E[R] = E[R_{\text{Group1}}] + E[R_{\text{Group2}}] + E[R_{\text{Group3}}]$.

6 ANALYSIS OF ROUTER-ORIGINATED RETRANSMISSIONS

The analysis of the router-originated retransmission strategy is very simple now because it can be viewed as a composition of several trees with sender-originated strategy. We distinguish between two scenarios. In the first, we call it Router_A, retransmissions are performed by sender S and by the multicast routers MR₂ and MR₃. In the second scenario (Router_B) all routers are involved in retransmissions.

For analysis of scenario Router_A we divide the multicast tree (see fig. 5) into three disjoint subtrees (rooted at S, MR₂ and MR₃, respectively) and analyze each subtree separately (similar to the sender-originated analysis). The cost parameter c for the subtrees are 3 (for the one rooted at S), 2 (for the MR₂-subtree) and 4 (for the MR₃-subtree). We get $E[R] = E[R_S] + E[R_{MR_2}] + E[R_{MR_3}]$ for the expected overall retransmission cost.

In the same way scenario Router_B results in expected retransmission cost of $E[R] = E[R_S] + E[R_{MR_1}] + E[R_{MR_2}] + E[R_{MR_3}] + E[R_{MR_4}]$ where the cost parameter c must be set to 1 for the sender-subtree and to 2 for all other subtrees.

7 NUMERICAL RESULTS

We first examine how the expected retransmission cost for sender-originated, receiver-originated and router-originated protocols varies with the loss probabilities $p_i := p$, i.e. with the geographic spread of the multicast group. The results are shown in fig. 9 (the right diagram is just an enlargement of the bottom left corner of the left one).

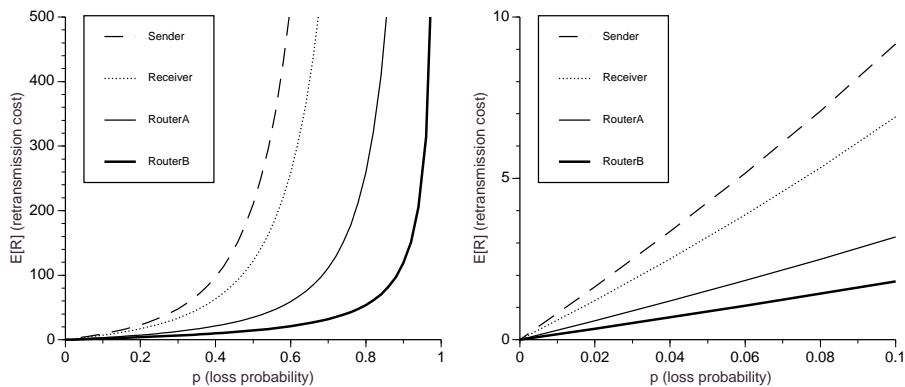


Figure 9 Retransmission cost when varying all p_i .

Obviously the cost of sender-originated protocols soon becomes unacceptable in our multicast tree. Even for small loss probabilities the retransmission cost is rather

high. The improvement achieved by the receiver-originated approach is only marginal. In comparison to this both router-originated scenarios perform well, especially scenario Router_B, where all routers are responsible for retransmissions. This approach results in acceptable cost even for loss probabilities higher than 50%.

However, in real networks the loss probabilities will not be the same for all links. In peak traffic times most links will be rather reliable while only a small number of links will have high loss probabilities at the same time. The results of a corresponding scenario are presented in fig.10. All loss probabilities (except one varying parameter) have been set to 0.1, a realistic value in a large network with high traffic.

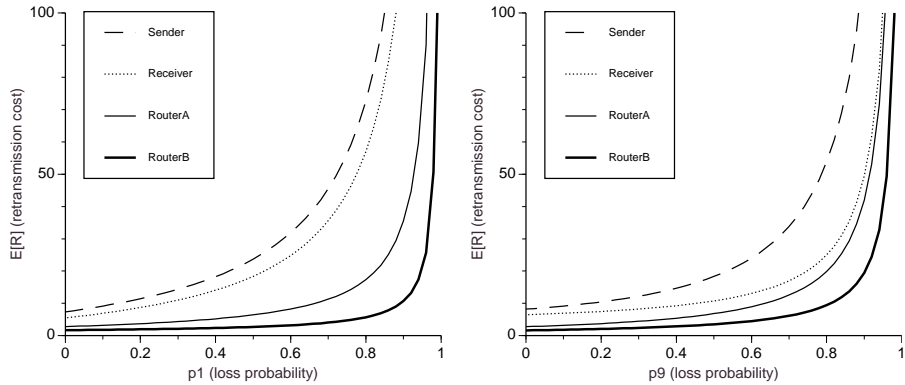


Figure 10 Retransmission cost if the loss probability varies on one link only left: $p_i = 0.1$ ($i = 2, \dots, 9$), p_1 varies, right: $p_i = 0.1$ ($i = 1, \dots, 8$), p_9 varies.

As can be seen an unreliable link near the source (fig. 10 left) has more impact on the cost for sender-originated and receiver-originated protocols than an unreliable link near a receiver (fig. 10 right). In contrast to this the cost of the router-originated protocol is hardly influenced by the location of the unreliable link since error recovery is handled locally. The performance is very well in both cases.

8 CONCLUSIONS

In this paper we have analytically examined different retransmission strategies for reliable multicast protocols. Particularly, we have presented a Markov chain model for calculating the retransmission cost of sender-originated, receiver-originated and router-originated protocols. An important feature of our analysis is that we allow a dependence between packet loss events for different receivers. Moreover, packet loss probabilities may be defined on a per link basis. Thus our method yields more realistic results than previous work in this context. Numerical results have shown that the cost for sender-originated protocols are only acceptable in the local area where loss probabilities are low. With growing size of the network and increasing

geographic spread of the groups (high loss probabilities) either receiver-originated or even better router-originated protocols have to be used instead.

9 REFERENCES

- Deering S. (1988) Multicast Routing in Internetworks and Extended LANs. Proceedings of *ACM SIGCOMM*.
- Hermanns, O. and Schuba, M. (1996) Performance Investigations of the IP Multicast Protocols. *Computer Networks and ISDN Systems* **28**, 429-39.
- Armstrong S., Freier A. and Marzullo K. (1992) Multicast Transport Protocol. *RFC 1301*.
- Heinrichs B. (1994) AMTP: Towards a High Performance and Configurable Multi-peer Transfer Service. *Architecture and Protocols for High-Speed Networks*. Danthine, Effelsberg, Spaniol (Eds.), Kluwer Academic Publishers.
- Paul S., Sabnani K. K., Lin J. C.-H. and Bhattacharyya S. (1997) Reliable Multicast Transport Protocol (RMTP). *IEEE Journal on Selected Areas in Communications*, **Vol. 15**, No. 3, 407-21.
- Yavatkar R., Griffioen J. and Sudan M. (1995) A Reliable Dissemination Protocol for Interactive Collaborative Applications. *ACM Multimedia '95*, 333-44.
- Schuba M. (1998) SRMT - A Scalable and Reliable Multicast Protocol. Proceedings *ICC'98*.
- Pingali S., Towsley D. and Kurose J. F. (1994) A Comparison of Sender-Initiated and Receiver-Initiated Reliable Multicast Protocols. *SIGMETRICS'94*, 221-30.
- Stewart W. J. (1994) *Introduction to the Numerical Solution of Markov Chains*. Princeton University Press

10 BIOGRAPHIES

Marko Schuba has studied computer science in Aachen, where he received his Diplom (the German equivalent to M.Sc.) in 1995. Since then he is with the Computer Science Department of Aachen University of Technology. His research interests include performance modelling and evaluation of computer networks, multicast network and transport protocols, and interworking between connectionless protocols and ATM. Currently, he is working towards his Ph.D.

Peter Reichl has studied mathematics, physics, philosophy and music in Munich and at Cambridge University, where he wrote his Diploma thesis under the supervision of F. P. Kelly. Since 1995 he is with the Computer Science Department of Aachen University of Technology where he is working towards his Ph.D. His current interests include traffic modelling, models for self-similarity and the modelling and evaluation of network protocols.