Performance Analysis of Reliable Multicast Mechanisms for Widely Spread Distributed Applications in the Internet

Marko Schuba
Informatik 4 (Communication and Distributed Systems)
Aachen University of Technology
52056 Aachen, Germany

Abstract This paper compares Internet packet transmission times of different reliable multicast protocols for widely spread distributed applications. The analytical model takes realistic Internet delays into account and calculates the mean transmission time and its variance for individual receivers in a multicast group. The results of an example scenario show that local error recovery techniques significantly reduce the transmission time of packets. Especially protocols which are based on local retransmissions by routers perform very well.

Keywords: Performance Analysis, Reliable Multicast, Distributed Applications, Internet

1 Introduction

The interest in reliable multicast protocols has steadily grown in the last years because a number of distributed applications like distributed simulation, replicated databases or distributed computing require reliable delivery of data to a group of receivers. In order to provide scalability, i.e. efficient data distribution to a large and/or widely spread multicast group in the Internet different protocols have been proposed. These protocols can be classified according to their retransmission strategy, i.e. the responsibility of nodes to perform error recovery. Firstly, there are sender-based protocols where only the sender retransmits lost packets (e.g. RAMP [1], or AMTP [2]). To become more scalable sender-based reliable multicast protocols usually concentrate on the reduction of the ACK implosion at the sender. Secondly, we have approaches with local recovery mechanisms, i.e. lost packets can be retransmitted locally by other nodes than the sender. Protocols using such strategies can be further subdivided into receiver-based approaches (e.g. LGC [3], SRM [4], or RMTP [5]) where a subgroup or all members of the multicast group may be involved in the retransmission process, and router-based protocols (also called server-based, e.g. LBRM [6], or SRMT [7]) which allow retransmissions by non-members of the group (routers, or servers co-located with routers), too.

Until now only a few analyses compare these different retransmission strategies. In [8] and [9] processing cost and bandwidth requirements of sender-based and receiver-based protocols are examined. Investiga-

1. This work was supported by Ericsson Eurolab
tions in [7], [10] and [11] consider all three protocol types but the analyses are again limited to processing cost and bandwidth usage. Only little attention has been paid to the actual transmission time so far ([12] being an exception for a sender-based protocol). The reason for this lack of transmission time analyses lies in the fact that from the application’s point of view reliability is much more important than transmission time. Nevertheless, transmission time plays an important role in distributed applications and a protocol which delivers data reliably and additionally in a fast way would obviously be a preferred solution.

In this paper we compare the Internet transmission time of packets for individual receivers in sender-based, receiver-based and router-based protocols with either ACK or NAK acknowledgement mechanisms. Our investigations are based on multicast trees with edges corresponding to typical Internet connections characterized by three parameters: average one-way delay, its variance, and packet loss probability. The parameter values are taken from [13]; the method for assessing the one-way delays is also described in [14].

The paper is structured as follows. In section 2 we give a short description of the different retransmission strategies used by reliable multicast protocols. Section 3 presents our model for analytical comparison of the different protocol types. An example multicast scenario and the resulting transmission times are described in section 4. Section 5 summarises the paper.

2 Retransmission strategies of reliable multicast protocols

In this section reliable multicast protocols are briefly classified according to their retransmission strategy.

2.1 Sender-based protocols

If a sender-based protocol detects packet loss, the respective packets are retransmitted as multicast by the sender to the whole group or are unicasted to individual “unsuccessful” members (if the number of them is below a given threshold). Examples for sender-based protocols are RAMP or AMTP.

2.2 Receiver-based protocols

Here, receivers of the same region form a local group. A designated receiver of each group is responsible for processing acknowledgements and starting retransmissions if one of the local group members detects a packet loss. The groups can be ordered hierarchically to make the approach more scalable. While the sender usually multicasts all packets globally (but listens only to acknowledgements of the designated receivers within its local group), retransmissions of designated receivers have only local scope, i.e. are limited to their local groups. Representatives of this class of protocols are LGC and RMTP.

Error recovery in SRM differs from the mechanisms of LGC and RMTP in so far that in SRM every packet loss is recovered by the nearest (successful) receiver, e.g. within a limited area of the multicast tree. Since the influence of this difference on our analysis performed later is only marginal we put SRM into the class of receiver-based protocols, too.

2.3 Router-based protocols

Router-based protocols extend the idea of receiver-based protocols by using a hierarchy of special routers (or servers co-located with routers) which store packets for possible retransmissions. Similar to receiver-based protocols this results in local groups
with a sending node (sender or router) and several receiving nodes (routers and/or receivers). After a packet loss, which is signalled to the local sending node, the respective packets are retransmitted locally within the group. Examples of such router-based protocols are LBRM (with two levels of hierarchy) and SRMT.

3 Modelling transmission times

In order to analyse transmission times of reliable multicast protocols for distributed applications in the Internet, we first need a realistic model for packet delays and loss probabilities in a multicast tree. Since such a multicast tree usually is just a composition of several point-to-point connections it is sufficient to model each edge of the multicast-tree as a separate Internet connection. We characterize an Internet connection by three parameters: average packet delay, packet delay variance and packet loss probability. Typical values for such parameters are described in [13] and shown in table 1. The parameters refer to Internet connections from a host in Aachen (Germany) to other hosts located in Germany (Aachen, Köln, Karlsruhe, Dresden), i.e. we have four national connections $C_1$ to $C_4$.

<table>
<thead>
<tr>
<th>Destination Host</th>
<th>Average Delay [ms]</th>
<th>Delay Variance</th>
<th>Packet Loss [%]</th>
<th>Connection Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aachen</td>
<td>3</td>
<td>3</td>
<td>0.1</td>
<td>$C_1$</td>
</tr>
<tr>
<td>Köln</td>
<td>6</td>
<td>6</td>
<td>0.2</td>
<td>$C_2$</td>
</tr>
<tr>
<td>Karlsruhe</td>
<td>19</td>
<td>5</td>
<td>0.8</td>
<td>$C_3$</td>
</tr>
<tr>
<td>Dresden</td>
<td>27</td>
<td>6</td>
<td>1.6</td>
<td>$C_4$</td>
</tr>
</tbody>
</table>

Table 1: Rounded parameters for unidirectional Internet connections

3.1 Multicast subtrees

The different protocols described in section 2 split the overall multicast tree into one or more subtrees, each with packets transmitted reliably from a sender to a number of receivers. The sender might be the original sender, a designated receiver or a router; a receiver is either a real group member (maybe a designated receiver) or a router. If we assume for the receiver-originated protocols that the sender transmits packets only to the (designated) receivers of its local group but not globally to any receiver of another designated receiver’s group, then each subtree may be analysed separately as if using a sender-based protocol. Thus, we first concentrate on the sender-based transmission time analysis of subtrees and compute the measures for the overall tree later (see section 3.4).

The path from a sender to a particular receiver $i$ in a subtree may consist of several edges, i.e. a number of Internet connections in our model. In order to analyse the packet transmission time from the sender to the receiver we have to calculate the parameters for the overall sender-receiver connection from the parameters given for the individual edges. Let the connection consist of $T$ edges with independent random variables $X_{1i}, ..., X_{Ti}$ for the delay and packet loss probabilities $0 \leq p_{1i}, ..., p_{Ti} < 1$. Then the overall delay $X_i$ and the overall packet loss probability $p_i$ can be derived very simple as sum $\sum X_{ji}$ or product $1 - \prod (1 - p_{ji})$ of the individual edge values.

3.2 Transmission time in a subtree with one receiver

Let us first look at the special case of a subtree with just one receiver. Delay and packet loss probability of the overall connection are given by the random variable $X$ and $p$, resp.
Assuming that the protocol uses positive acknowledgments for error detection a retransmission is started if a packet has not been acknowledged by the receiver after expiration of a timer $T_{Ack}$. We define the value of $T_{Ack}$ to be twice the round-trip time (RTT) of the connection. For simplicity reasons we assume all connections to be symmetric and use the doubled average RTT for the calculation of $T_{Ack}$, i.e. $T_{Ack} = 4 \cdot E[X]$. Because the number of required retransmissions $N$ is geometrically distributed, i.e.

$$P(N=n) = p^n \cdot (1-p) \quad \forall n \geq 0,$$  

we get the overall transmission time $Y$ as sum of the number of times $T_{Ack}$ has to expire multiplied by the value of $T_{Ack}$ and the delay of the last successful transmission, hence

$$Y = N \cdot T_{Ack} + X.$$  

The expected value and variance of the transmission time is then given by

$$E[Y] = E[N] \cdot T_{Ack} + E[X]$$

and (because $N$ and $X$ are independent)

$$V[Y] = V[N] \cdot T_{Ack}^2 + V[X].$$

### 3.3 Transmission time in a subtree with several receivers

If the number of receivers in the subtree is $D \geq 1$ with connection delays $X_1, ..., X_D$, then the sender has to adapt its retransmission speed to these different delay values. The sender must collect ACKs from all receivers, i.e. the timer $T_{Ack}$ has to be set with respect to the maximum RTT. Therefore, the value of $T_{Ack}$ is set to $T_{Ack} = 4 \cdot E_{max}$, where $E_{max} = \max_{1 \leq i \leq D} E[X_i]$.

Expected value and variance of the transmission time are calculated using (3) and (4).

### 3.4 Transmission time for a sequence of subtrees

In the case of receiver-based and router-based protocols packets for some receivers have to traverse several subtrees until they reach their destination. The expected value and the variance of the overall transmission time for such receivers can be calculated by just summing up the respective values of the intermediate receivers on the way to the final destination.

### 4 Example scenario and numerical results

The example scenario to be analysed describes a “national” multicast transmission, e.g. within a single country. The multicast group consists of ten members $D_1, ..., D_{10}$ and the respective multicast tree is presented in fig. 2.

![Figure 2: Example scenario](image-url)
for retransmissions. For the sender-based protocol the situation is clear. In the case of the receiver-based protocol we define six designated receivers: $D_1$ and $D_6$ for the first hierarchy level, $D_2$, $D_4$, $D_7$, and $D_9$ for the second. In the router-based protocol all routers shown in the tree (in reality there might be more in between) perform error-recovery. The resulting logical structure of the multicast trees for the different protocols is shown in fig. 3. Each arrow in a tree corresponds to a connection on which retransmissions are performed.

Fig. 4a and b show the expected value and the variance of the transmission times of the receivers with the shortest (node $D_1$) and longest (node $D_{10}$) average distance to the sender. All results are plotted depending on the packet loss probability $p$.

In the case of the sender-based protocol the average transmission time to receiver $D_1$ increases rather fast with the loss probability. The main reason for this is the setting of timer $T_{Ack}$ which has a significant influence on the performance of the protocols. The more receivers a sender is responsible for, the higher the probability that the value of $T_{Ack}$ and thus the transmission time is large. For the receiver-based protocol the results are better because the timer for $D_1$ is only increased because of receiver $D_6$. The best transmission time is achieved by the router-based protocol, because of its short retransmission paths.

For receiver $D_{10}$ the situation is slightly different. Although the receiver-based approach is able to benefit from the hierarchical structure and thus from small timer values, its overall performance is even worse than the sender-based protocol’s. This is due to the fact that packets get to receiver $D_{10}$ in a roundabout way (via $D_6$ and $D_9$). The behaviour of the router-based protocol is much better. Even for a packet loss probability of 5% the transmission time grows only by 15 ms.

The variance of the transmission time is very similar for both receivers; only the absolute values differ. The positive influence of local recovery on the variance is rather obvious, here. Especially the transmission time of the router-based approach turns out to have very small fluctuations.

5 Conclusions

In this paper we have presented a new analytical model for the evaluation of sender-based, receiver-based and router-based reliable multicast protocols in terms of average transmission time and transmission time variance. After modelling a realistic Internet

![Figure 3: Logical tree structure for different protocol types](image)
scenario (e.g. of a distributed application) we have demonstrated the transmission time reduction that can be gained by local recovery techniques for protocols using positive acknowledgment mechanisms. The router-based approach has shown to perform best with respect to average transmission time and variance. In contrast to this the transmission time of receiver-based protocols has turned out to be worse than expected, because the improvement achieved in comparison to the sender-based approach is only marginal (if there is an improvement at all). Taking earlier results concerning bandwidth requirements into account (e.g. [7] or [10]), router-based protocols seem to be a promising approach for reliable multicast in widely spread distributed applications.

6 References


