Delay Bounds and Scalability for Overlay Multicast[⋆]

György Dán and Viktória Fodor

ACCESS Linnaeus Centre, School of Electrical Engineering
KTH, Royal Institute of Technology
Stockholm, Sweden
{gyuri,vfodor}@ee.kth.se

Abstract. A large number of peer-to-peer streaming systems has been proposed and deployed in recent years. Yet, there is no clear understanding of how these systems scale and how multi-path and multihop transmission, properties of all recent systems, affect the quality experienced by the peers. In this paper we present an analytical study that considers the relationship between delay and loss for general overlays: we study the trade-off between the playback delay and the probability of missing a packet and we derive bounds on the scalability of the systems. We use an exact model of push-based overlays to show that the bounds hold under diverse conditions: in the presence of errors, under node churn, and when using forward error correction and various retransmission schemes.

Keywords: Overlay multicast, Scalability, Delay, Large-deviation theory.

1 Introduction

Overlay multicast is promising for distributing streaming data simultaneously to a large population of users. The architectures proposed for overlay multicast (a.k.a. peer-to-peer streaming) generally fall into one of two categories: push-based or pull-based. Solutions of both categories utilize multi-path transmission. Multi-path transmission offers two advantages. First, disturbances on an overlay path lead to graceful quality degradation in the nodes. Second, the output bandwidth of the peers can be utilized more efficiently.

Push-based overlays follow the traditional approach of IP multicast: nodes are organized into multiple transmission trees and relay the data within the trees. The streaming data is divided into packets and packets are transmitted at round-robin through the transmission trees, providing path diversity for subsequent packets in this way. The transmission trees are constructed at the beginning of the streaming session and are maintained throughout the session by a centralized or a distributed protocol. Node churn leads to the disconnection of the trees and hence to data loss, which is one of the main deficiencies of push-based overlays.

[⋆] This work was in part supported by the Swedish Foundation for Strategic Research through the project Winternet.

A. Das et al. (Eds.): NETWORKING 2008, LNCS 4982, pp. 227–239, 2008.

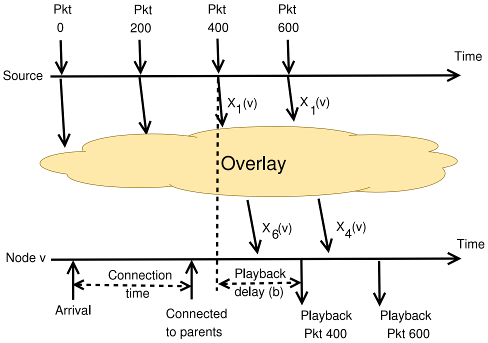

Fig. 1. The playback delay and the time needed to connect to the overlay

The pull method (also called swarming) follows the approach of batch peer-to-peer content distribution: nodes know about a subset of all nodes (their neighbors); they both receive data from and forward data to their neighbors. There is no global structure maintained, hence the scheduling of data transmissions is determined locally. Pull-based overlays are resilient to node churn as forwarding decisions are taken based on the actual neighborhood information, but their efficiency depends on the scheduling algorithm.

Several works deal with the management of push-based overlays ([1,2] and references therein) and with scheduling algorithms for pull-based overlays ([3,4] and references therein). There are also numerous proposals on how to improve the robustness of the overlays to errors using coding techniques such as forward error correction (FEC), multiple description coding (MDC) and network coding [5]. The evaluation of the proposed solutions is mostly based on simulations and small scale measurements; the analytical modeling of overlay multicast has not received much attention.

There are a number of commercial deployments of overlay multicast, e.g. [6,7]. Commercial systems often serve hundreds of thousands of peers simultaneously [8], yet little is known how they would behave if the number of concurrent users increased to its tenfold. We argue that there is a need for an analytical understanding of the performance of large systems in order to be able to design systems that can provide predictable and controllable quality under a wide range of operating conditions.

The most important difference between overlay multicast systems and peer-to-peer content distribution, such as Bittorrent, is the delay aspect: data should be delivered to the nodes before their playout deadline. The probability that data arrive before their playout deadline depends on the playback delay b: the lag between the time of the generation of a packet at the source and the time of the playback at the peers, as shown in Fig. 1. The necessary playback delay for providing good streaming quality may depend on many factors: the overlay's architecture and size, which determine the nodes' distances from the source; the per-hop delay distribution, the packet loss probability between the nodes and the error control solutions used; and the frequency of node departures and the time needed to reconnect to the overlay in the case of push-based overlays.

In this paper we consider two questions related to the playback delay. First, how fast does the probability of missing a packet decrease as a function of the playback delay. Second, how fast should the playback delay be increased to maintain the probability of missing a packet unchanged as the overlay's size increases. We give bounds on the decrease of the packet missing probability and the necessary increase of the playback delay under general assumptions. Our results facilitate the choice of benchmarking metrics for the performance evaluation of overlay multicast systems.

The rest of the paper is organized as follows. Section 2 gives an overview of the related work. Section 3 presents bounds on the playback delay and the scalability of the overlays based on the foundations of large deviation theory. Section 4 discusses the delay bounds and the performance of the overlays based on an exact mathematical model of overlay multicast systems, and we conclude our work in Section 5.

2 Related Modeling Work

The trade-off between the available resources and the number of nodes that can join the overlay was studied for overlay multicast systems utilizing a single transmission tree in [9]. The first models that describe the data distribution performance of multi-tree-based overlay multicast were proposed in [10,11] and showed that these systems exhibit a phase-transition when using FEC. The effect of the forwarding capacity on multi-tree-based overlays was investigated in [12] using a queuing theoretic approach, and in [13] based on a fluid model. The delay characteristics of a pull-based overlay were investigated in [4], and the authors showed an exponential relationship between the playback delay and the packet missing probability. The analytical results presented there are limited to a specific packet forwarding algorithm and to complete graphs. In [14] we presented an analytical model of push-based overlays and used the model to identify the primary sources of delay in overlay multicast. We are however not aware of analytical results neither on the scalability of overlay multicast architectures in terms of delay, nor on the effects of the playback delay on the performance.

3 Delay Bounds

We model the overlay as a directed graph $G = (V, E)$ with $N = |V|$ vertices. The set of vertices and edges can change over time due to node churn and due to the overlay management. We chose to omit the time dimension in our notation in order to ease understanding. Let us denote by s the source of the multicast, and by T_i the spanning tree rooted at the source, through which the copies of packet i reach the nodes in V.

In a push-based overlay with τ trees the T_i are predetermined by the overlay maintenance entity and $\cup T_i = E$. In a pull-based overlay the T_i are a result of local decisions taken in the nodes, such that all edges $(u, v) \in T_i$ are chosen from E. E is maintained by the overlay maintenance entity. Let us denote by the random variable $L_i(v)$ the length of the simple path from s to v in T_i, and by the random variable $D_i(v)$ the time it takes for packet i to reach node v from s in T_i. We assume that every packet reaches every node after some finite amount of time, i.e., $\lim_{b\to\infty} P(D_i(v) \leq b) = 1$. Both push-based overlays with retransmissions and pull-based overlays (e.g., using the policy described in [4]) can fulfill this requirement.

$D_i(v)$ is the sum of $L_i(v) \leq N - 1$ per-hop delays. We denote the per-hop delays by the non-negative random variables $X_h(v)$, i.e., $D_i(v) = \sum_{h=1}^{L_i(v)} X_h(v)$. For example, in Fig. 1, $L_{400}(v) = 6$ and $L_{600}(v) = 4$, i.e., the packets reach node v on different overlay paths. The distribution of the $X_h(v)$ depends on many factors, e.g., the data distribution model (time spent for coordination between nodes), the probability of losses (due to churn and network congestion), the nodes' upload capacities, and the distance from the source (many proposed architectures place nodes with large upload capacities close to the source). The probability that node v with playback delay b misses an arbitrary packet i is $P(D_i(v) > b)$.

3.1 Playback Delay in Stationary State

First, we consider an overlay in which N is a stationary process, and consequently $L_i(v)$ is a stationary process as well. $L_i(v)$ has a finite support and the per-hop delays $X_h(v)$ follow distributions with finite moment generating functions (m.g.f), i.e., a light-tailed distribution. We are not aware of measurements that would show heavy-tailed end-to-end delay distributions, but theoretical work has shown that heavy-tailed distributions can arise in the presence of self-similar traffic [15]. We argue that the per-hop delays can be modeled with a distribution with finite m.g.f. even in this case, if the back-off scheme in use is sub-exponential, e.g., polynomial. Large delays trigger retransmission requests that cut the heavy tail of the distribution, and the resulting distribution will be geometric-like, which has a finite m.g.f. We leave the case of exponential back-off schemes, i.e., heavy-tailed per-hop-delays, to be subject of future work. Given the per-hop-delays with finite moment generating functions, the sum of the per-hop delays has finite m.g.f. even if the per-hop delays are positively correlated, as shown by the following lemma.

Lemma 1. *Given non-negative random variables X_h ($h = 1 \ldots n, n > 0$) with marginal p.d.f $f_h(x)$ and joint p.d.f $f(x_1, \ldots, x_n)$ such that $E[e^{\theta X_h}] < \infty$, for $S_n = \sum_{h=1}^{n} X_h$ we have $E[e^{\theta S_n}] < \infty$.*

Proof. For independent r.v.s the proof is trivial, $E[e^{\theta S_n}] = \Pi_{h=1}^{n} E[e^{\theta X_h}]$. For correlated r.v.s, we prove the lemma for $n = 2$, induction can be used for $n > 2$. Let us order X_1 and X_2 such that $\int_0^x f_1(t)dt \leq \int_0^x f_2(t)dt$ for $\forall\ x > x_0$ ($x_0 > 0$). Let us denote by X_2^{**} a random variable that is distributed as X_2 but is in perfect positive dependence with X_1 (see [16] for a definition), that is $x_2 = g(x_1)$ for some function g. $E[e^{\theta(X_1+X_2)}]$ is a convex, monotonically increasing function, hence [16]

$$E[e^{\theta S_2}] = E[e^{\theta(X_1+X_2)}] \leq E[e^{\theta(X_1+X_2^{**})}]. \tag{1}$$

Since there exists $\theta' > 0$ such that $E[e^{\theta' X_1}] < \infty$

$$E[e^{\theta S_2}] = \int_0^\infty \int_0^\infty e^{(x_1+x_2)\theta} f(x_1,x_2) dx_1 dx_2 \leq \int_0^\infty e^{(x_1+g(x_1))\theta} f_1(x) dx_1 \tag{2}$$

$$\leq a(x_0,\theta) + \int_{x_0}^\infty e^{2x_1\theta} f_1(x) dx_1 < \infty, \tag{3}$$

where (2) holds because of (1), (3) holds because $g(x) \leq x$ for $x > x_0$ due to the ordering of X_1 and X_2, and (3) holds for every $0 < \theta \leq \theta'/2$. □

Based on Lemma 1 and on results from large deviation theory [17] we can prove the following theorem.

Theorem 1. *The decrease of the probability that an arbitrary node with playback delay b misses an arbitrary packet in an overlay with N nodes is asymptotically at least exponential in b if the per-hop delays have finite m.g.f.*

Proof. We use the law of total probability to express the probability of missing a packet

$$P(D_i(v) \geq b) = \sum_{l=1}^{N-1} P(D_i(v) \geq b | L_i(v) = l) P(L_i(v) = l), \tag{4}$$

and in the following we show that the decrease of $P(D_i(v) \geq b | L_i(v) = l)$ is asymptotically at least exponential in b for any l.

According to Chernoff's bound for the average A_n of n i.i.d random variables X (note that it is not the sum, but the average of the r.v.s.)

$$P(A_n \geq x) \leq e^{-nI(x)}, \tag{5}$$

where $I(x)$ is the rate function given by

$$I(x) = max_{\theta>0} \theta x - ln(M_X(\theta)),$$

and $M_X(\theta) = E[e^{X\theta}]$ is the moment generating function of X. Fig. 2 shows the rate functions for two distributions with different parameters. Chernoff's bound holds for any distribution for which $M_X(\theta) < \infty$, and for $x > E[X]$. The rate function $I(x)$ is convex for scalar random variables, is monotonically increasing on $(E[X], \infty)$ and $I(E[X]) = 0$ [17]. For non-negative r.v.s X with $E[X] \geq 0$, the derivative $\frac{\partial I(x)}{\partial x}|_{x_0} \geq I(x_0)/x_0$, so that we have

$$I(ax) \geq I(x) + (ax - x)I(x)/x = aI(x) \qquad \text{for} \quad \forall a > 1 \tag{6}$$

and hence

$$P(A_n \geq ax) \leq e^{-nI(ax)} \leq e^{-naI(x)}. \tag{7}$$

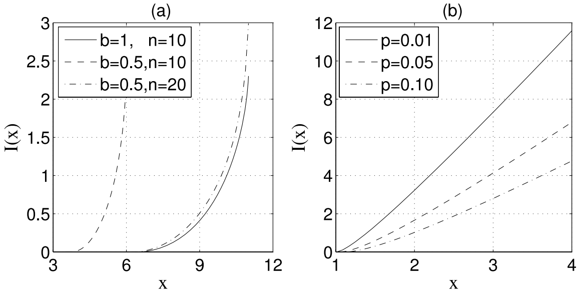

Fig. 2. Rate function for two distributions (a) $X = a+yb$ where $a = 1$ and y has discrete uniform distribution on $[1, n]$ and (b) geometric distribution with failure probability p

We apply (7) for the r.v. $D_i(v)$ conditioned on $L_i(v) = l$ with $n = 1$

$$P(D_i(v) \geq ab|L_i(v) = l) \leq e^{-I_l(ab)} \leq e^{-aI_l(b)}, \tag{8}$$

where $I_l(b)$ is the rate function conditioned on $L_i(v) = l$. Eq. (8) holds for any l, which together with (4) proves the theorem. □

The result is independent of the distribution $P(L_i(v) = l)$, and holds whenever there is enough forwarding capacity in the overlay. It is also independent of the number of packets in the stream and does not make any assumption on the graph's connectivity or the distribution scheme, in particular, it does not assume a complete graph. The simulation results presented in [4] support our analytical result for pull-based overlays, and we show results later that support the theorem for push-based overlays.

3.2 Scalability

The question we address here is how the probability of missing a packet changes as the overlay's size increases. The change of the overlay's size affects the data distribution through the distribution of the $L_i(v)$, hence we focus on the question how the increase of the path length $L_i(v)$ affects the probability of missing a packet. This way we decouple the problem of scaling in terms of delay from the problem of overlay maintenance, e.g., neighbor selection: the neighbor selection algorithm, random or optimized according to some metric, influences the scaling of the path lengths $L_i(v)$. Given the scaling of the path lengths we can evaluate the scaling in terms of delay. A trivial lower bound on the scaling in terms of delay is given by the increase of $E[D_i(v)]$, which is proportional to $E[L_i(v)]$. In the following we show that the upper bound is proportional to $L_i(v)$ as well.

We consider the case when the $X_h(v)$ are i.i.d. random variables with $M(\theta) < \infty$. We note that there are no asymptotic results available for correlated r.v.s, but we conjecture that the following theorem holds for correlated and for non identically distributed r.v.s as long as $E[X_h(v)]$ is bounded from above, and leave the proof to be subject of future work.

Theorem 2. *If $L_i(v) \sim O(logN)$ then the increase of the playback delay b needed to maintain the probability of missing an arbitrary packet unchanged is at most asymptotically logarithmic in N.*

Proof. We prove the theorem by showing that the upper bound of the playback delay increases logarithmically. We look for a $d \geq 0$ such that for $a \geq 0$

$$P(D_i \geq b | L_i(v) = l) = P(D_i \geq b + d | L_i(v) = l + a). \tag{9}$$

We use Chernoff's bound to express the upper bounds of the probabilities

$$e^{-lI(\frac{b}{l})} = e^{-(l+a)I(\frac{b+d}{l+a})}. \tag{10}$$

We omit the base and rearrange the exponents to get

$$l\left[I(\frac{b}{l}) - I(\frac{b+d}{l+a})\right] = aI(\frac{b+d}{l+a}). \tag{11}$$

The right hand side of (11) is always positive. As the rate function is convex and monotonically increasing on $(E[X_h(v)], \infty)$ we have the condition

$$\frac{b+d}{l+a} \leq \frac{b}{l}, \tag{12}$$

which can be rearranged to

$$d \leq \frac{ba}{l}. \tag{13}$$

If $L_i(v) = O(logN)$ then as the overlay grows from N_1 to N_2 we have $a \sim log(N_2/N_1)$ and this proves the theorem. □

In general, (12) shows that it is sufficient to increase the playback delay at the same pace as the depth of the spanning trees grows in order to maintain the probability of missing a packet unchanged. E.g., if the nodes' distances from the source grow as a linear function of the overlay's size, then the playback delay should be increased in direct proportion to the growth of the overlay to keep the packet missing probability constant. Consequently, in order to show that an overlay scales as $O(logN)$ in terms of playback delay, it is enough to show that $L_i(v) = O(logN)$. Even though this result is in accordance with one's expectations, it is not straightforward. Unfortunately, the converse of the theorem cannot be proved: *there is no upper bound on the increase of the packet missing probability for constant playback delay as the overlay's size grows*, because the increase depends on the shape of the rate function.

These asymptotic results indicate that *the exponential decrease* of the packet missing probability as a function of the playback delay *is not a good measure of the efficiency of a scheduling scheme* in overlay multicast: all scheduling schemes that manage to distribute the data to all nodes have this property under the assumption that the per-hop-delays have finite m.g.f. The scalability with respect to the overlay's size is however a good measure: Theorem 2 shows that in a scalable overlay the playback delay does not have to be increased faster than the logarithm of the increase of the overlay's size to keep the packet missing probability constant. While the exponential decrease was shown via simulations for a specific scheduling algorithm in [4], we are not aware of any scheduling algorithm for pull-based systems with analytical results on its scalability.

4 Numerical Results

In the following we present numerical results obtained using an exact model of multi-tree-based overlays presented and validated in [14], and show that the bounds derived using the large deviation approach hold in the presence of various retransmission schemes and FEC.

4.1 System Description

We denote the number of trees in the overlay by τ. We assume the existence of a tree maintenance entity (centralized [18] or decentralized [19,20]) that finds suitable predecessors for arriving nodes and for nodes that lose their predecessors due to node churn or preemption [1,21]. We denote by $\mathcal{L}_m(v)$ the distance of node v from the source in tree m, and say that node v is in level l of tree m if $\mathcal{L}_m(v) = l$. Packet i is distributed in tree $m = (i \ mod \ \tau) + 1$. To simplify the notation, we introduce the notion of stripe, and say that packet i belongs to stripe m if it is distributed in tree m.

We consider two forms of error control: forward error correction and retransmissions. When forward error correction (FEC) is used, the source adds c redundant packets to every k packets, resulting in a block length of $n = k + c$. We denote this FEC scheme by FEC(n,k). Once a node receives at least k packets of a block of n packets, it may recover the remaining c packets, and forwards the reconstructed packets if necessary. Block based FEC can be used to implement PET and the MDC scheme considered in [18], where different blocks (layers) of data are protected with different FEC codes: the probability of reception in the different layers depends on the strength of the FEC codes protecting them.

We consider three retransmission schemes. A node that detects a packet loss in stripe m requests the retransmission of the packet by one of these three strategies:

(RF) from its predecessor in tree m. If the loss is due to node churn then the node will have to wait until a new predecessor is found.

(RB) from another node that is forwarding packets in tree m in the same level as the node's actual predecessor. This scheme assumes that every node maintains a list of backup nodes, but we do not model the overhead of maintaining such a list.

(RS) from a predecessor in another tree. A predecessor in another tree is likely to be far away from the source in tree m, hence the retransmission might take longer than using a backup list.

For the RF and the RB strategies, the level of node v in the tree in which packet i should be distributed ($\mathcal{L}_{i \; mod \; \tau}(v)$) is the same as the distance of v in the spanning tree T_i through which the copies of packet i reach the nodes ($L_i(v)$), i.e., $L_i(v) = \mathcal{L}_{i \; mod \; \tau}(v)$. For the RS strategy $L_i(v) \geq \mathcal{L}_{i \; mod \; \tau}(v)$, i.e., due to the retransmissions the spanning tree T_i can be deeper than the trees maintained by the overlay maintenance entity.

4.2 System Parameters

We model the propagation delay D_p by a normal distribution truncated at 180 ms with mean and variance ($E[D_p] = 67$ ms, $\sigma_{D_p} = 21$ ms) extracted from a transit-stub network of 10^4 nodes generated with the GT-ITM topology generator [22]. We consider the streaming of a $B = 400$ kbps data stream, and the capacity of the source node's output link is $C(s) = 100$ Mbps. The outdegree of the source, $\mathcal{O}_s$, is set to 50 throughout the paper for easy comparison and to ensure that the overlay is feasible for all considered values of the number of trees [12], though the particular value of $\mathcal{O}_s$ does not affect the validity of our conclusions. The packet size is 1410 bytes. The distribution of the nodes' output capacities (C_v) and outdegrees ($\mathcal{O}_v$) is as in [4] and is shown in Table 1. Since the effect of the input capacity of the nodes is small on the results [14], we consider 10 Mbps for all nodes.

Table 1. Distribution of node output capacities and outdegrees

Ratio	15%	25%	40%	20%
C_v	10 Mbps	1 Mbps	384 kbps	128 kbps
$\mathcal{O}_v$	2.5τ	2τ	0.75τ	0.25τ

The inter-arrival times of nodes joining the overlay are exponentially distributed, this assumption is supported by several measurement studies, e.g., [23]. The session holding times M follow the log-normal distribution, the mean holding time is $E[M] = 306\,\mathrm{s}\,(\mu = 4.93, \sigma = 1.26)$ [23]. The nodes are prioritized according to their outdegrees as proposed in [21], hence large contributors are closer to the source in the trees in which they forward data and reconnect faster to the trees.

4.3 Packet Losses

First we evaluate the packet missing probability as a function of the playback delay for the case of packet losses. Fig 3 shows results for the RF scheme with two different packet loss detection times. We denote by $t_{dld} = 0$ the case when the loss of a packet is detected at the instant when it should have arrived (if it had not been lost), i.e., an ideal loss detection algorithm. We denote by $t_{dld} = 1$ when a retransmission is requested 1 s after the packet or the retransmission request has been sent out. The figure shows results for $N = 10^4$ nodes organized in $\tau = 4$ trees. In the absence of losses ($p = 0$) the decrease of the packet missing probability is faster than exponential. This is predicted by Fig. 2 (a), as the rate function for the discrete uniform distribution grows faster than linear. The

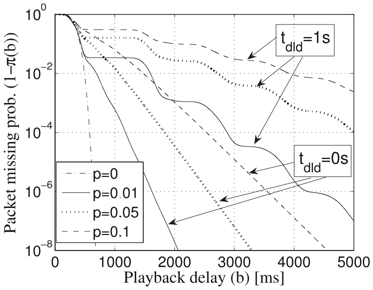

Fig. 3. Packet missing prob. vs. playback delay for different packet loss probabilities, and packet loss detection times, $\tau = 4$, $N = 10^4$

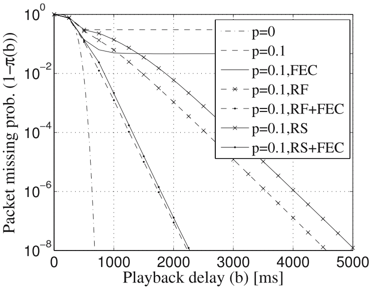

Fig. 4. Packet missing prob. vs. playback delay for different retransmission schemes, $\tau = 4$, $N = 10^4$

rate function of the geometric distribution is however close to linear (Fig. 2 (b)), hence we expect that in the presence of losses the decrease of the packet missing probability is not much faster than exponential. This is supported by the curves that show results for $p > 0$. The slope of the curves is related to the slope of the rate function of the per-hop delay distribution, the steeper the rate function, the faster the decrease of the packet missing probability. Though for $t_{dld} = 1$ the loss detection time is big compared to the per-hop delays and hence the packet missing probability decreases almost in a stepwise manner, we still observe the exponential decay. The curves for different loss probabilities and loss detection times show similar properties, they only differ in the slopes of the curves, so in the following we show results for $p = 0.1$ and $t_{dld} = 0$.

Fig. 4 shows results without losses and with losses for the RF and the RS retransmission schemes. The $N = 10^4$ nodes are organized in $\tau = 4$ trees, and FEC(4,3) is used when indicated. We observe that in the presence of losses the exponential decay does not hold when retransmissions are not used. In this case the analysis of the asymptotic behavior presented in [11] with respect to N,p and the FEC code applies to $lim_{b\to\infty}\pi(b)$: the system converges to the asymptotically stable fixed point of the discrete dynamic system shown in [11], and $lim_{b\to\infty}\pi(b) < 1$. Consequently, the assumptions of Theorem 1 are not fulfilled, because all nodes do not receive all data. When using retransmissions, the decay is exponential, as shown by the results for both the RF and the RS retransmission schemes, with and without FEC. FEC decreases the necessary playback delay to achieve a certain packet missing probability, but the exponential decay still holds: in the presence of FEC $D_i(v)$ is the minimum of two random variables, the time until the packet would be received through the corresponding parent and the time to FEC recovery (which is the k^{th} order statistic of $n-1$ random variables with finite m.g.f.), and hence $D_i(v)$ has finite m.g.f. Consequently, an alternative of increasing the playback delay in order to achieve a certain packet missing probability is to introduce FEC. Nevertheless, the ratio of FEC redundancy has to be adjusted dynamically based on feedback from the nodes. We observe a small difference between the results obtained with the two retransmission schemes. Using the RS scheme, retransmission of a packet in stripe m becomes possible only once nodes that do not forward data in tree m receive the packet. These nodes are in the last level of tree m, and hence, we observe a slow decay of the packet missing probability close to the point where the curves with and without retransmissions separate. Surprisingly, the difference between the results obtained with the RF and the RS schemes is small, especially when FEC is used, in which case the decrease of the decay close to the point where the curves with and without retransmissions separate is significantly smaller as well.

Fig 5 shows results for the RF retransmission scheme for different overlay sizes for $\tau = 4$ trees. Both in the presence of losses and in the absence of losses it is enough to increase the playback delay logarithmically in order to maintain the packet missing probability constant. Surprisingly however, the smaller the playback delay needed to achieve a certain packet missing probability for a given overlay size, the more sensitive is the overlay to the increase of the number of nodes. For $p = 0$ the packet missing probability increases by orders of magnitude if the overlay's size increases by a factor of ten, for $p = 0.1$ the increase is significantly smaller. Consequently, even if one could achieve a low packet missing probability with a small playback delay, the playback delay should be overdimensioned to ensure that the packet missing probability does not become too high if the overlay suddenly grows.

4.4 Node Churn

In the following we show analytical results for the case of node churn. For the reconnection times (Ξ) and the disconnection times (Ω) we use values similar to the measured data presented in [1]: $E[\Xi] = 5$ s and $E[\Omega] = 200$ s in the tree in which a node forwards data, and $E[\Xi] = 30$ s and $E[\Omega] = 100$ s in the trees in which it does not. (Nodes are disconnected with a higher probability in the trees in which they do not forward data.) In lack of a measured distribution we model the reconnection time Ξ with a normal distribution $N(E[\Xi], E[\Xi]/3)$. Based on these values the loss probability experienced by a node in a tree in which it forwards data ($p = 0.024$) and in which it does not forward data ($p = 0.1968$) can be calculated as described in [14].

The distribution of the retransmission times depends on the retransmission scheme used. For the *RB* and *RS* schemes we use the retransmission times as discussed in [14]. For the *RF* scheme retransmission occurs once the new predecessor is found, hence the retransmission time is the sum of the forward recurrence time of a renewal process with inter-renewal time Ξ (see [24]) and the retransmission time of the *RB* scheme. For all three schemes, retransmissions are asked from nodes that are present in the overlay and consequently $\lim_{b\to\infty} P(D_i(v) \leq b) = 1$.

Figure 6 shows results for the case of node churn and the three retransmission schemes. As expected, the RB scheme, which involves tremendous control overhead, performs best. Surprisingly however, the RS scheme performs nearly as good as the RB scheme, both without and with FEC. This is because under churn nodes experience more frequent losses in the trees in which they do not forward data, i.e., far from the source: when these losses occur, data is already

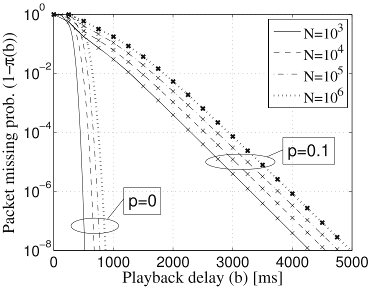

Fig. 5. Packet missing prob. vs. playback delay for different overlay sizes

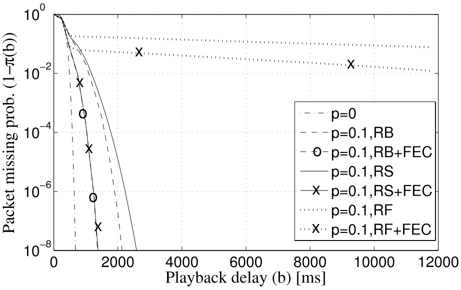

Fig. 6. Packet missing prob. vs. playback delay for $N = 10^4$, node churn and three retransmission schemes

available in large parts of the overlay, hence the additional delay introduced by the RS scheme is small. The RF scheme, due to the large retransmission delays, performs almost as bad as if there were no retransmissions at all. Nevertheless, we observe the exponential decay with a very slow decay rate. The bad performance of the RF scheme suggests that resilience to node churn in a push-based overlay requires retransmission schemes that abandon the rigid structure of the trees, and converge towards pull-based architectures, e.g., the RS and RB schemes.

5 Conclusion

In this paper we presented analytical results that show that in overlay multicast systems the packet missing probability decreases at least exponentially as a function of the playback delay under very general conditions. Consequently, the exponential decrease of the packet missing probability as a function of the playback delay is not a good measure of the efficiency of a scheduling scheme in overlay multicast: all scheduling schemes that manage to distribute the data to all nodes have this property. The scalability with respect to the overlay's size is however a good measure: the playback delay should not have to be increased faster than the logarithm of the overlay's size to keep the packet missing probability constant. We used an exact model of push-based overlays to show that the exponential decrease of the packet missing probability holds using various retransmission schemes and FEC. It will be subject of future work to design a provably scalable pull-based scheduling algorithm based on the analytical results on scalability. To the best of our knowledge, our work is the first to present a continuous time analytical model of the effects of the playback delay on overlay multicast.

References

1. Sung, Y.-W., Bishop, M., Rao, S.: Enabling contribution awareness in an overlay broadcasting system. In: Proc. of ACM SIGCOMM, pp. 411–422 (2006)
2. Liao, X., Jin, H., Liu, Y., Ni, L.M., Deng, D.: Anysee: Scalable live streaming service based on inter-overlay optimization. In: Proc. of IEEE INFOCOM (April 2006)

3. Magharei, N., Rejaie, R.: PRIME: Peer-to-peer Receiver drIven MEsh-based streaming. In: Proc. of IEEE INFOCOM (May 2007)
4. Massoulie, L., Twigg, A., Gkantsidis, C., Rodriguez, P.: Randomized decentralized broadcasting algorithms. In: Proc. of IEEE INFOCOM (2007)
5. Fodor, V., Dán, G.: Resilience in live peer-to-peer streaming. IEEE Communications Magazine 45(6) (June 2007)
6. 'PPLive," (June 2007), http://www.pplive.com/
7. "OctoShape," (June 2007), http://www.octoshape.com/
8. Hei, X., Liang, C., Liang, J., Liu, Y., Ross, K.W.: A measurement study of a large-scale P2P IPTV system. IEEE Trans. Multimedia 9(8), 1672–1687 (2007)
9. Small, T., Liang, B., Li, B.: Scaling laws and tradeoffs in peer-to-peer live multimedia streaming. ACM Multimedia (October 2006)
10. Dán, G., Fodor, V., Karlsson, G.: On the stability of end-point-based multimedia streaming. In: Proc. of IFIP Networking, May 2006, pp. 678–690 (2006)
11. Dán, G., Fodor, V., Chatzidrossos, I.: Streaming performance in multiple-tree-based overlays. In: Proc. of IFIP Networking, May 2007, pp. 617–627 (2007)
12. Dán, G., Fodor, V., Chatzidrossos, I.: On the performance of multiple-tree-based peer-to-peer live streaming. In: Proc. of IEEE INFOCOM (May 2007)
13. Kumar, R., Liu, Y., Ross, K.W.: Stochastic fluid theory for P2P streaming systems. In: Proc. of IEEE INFOCOM (May 2007)
14. Dán, G., Fodor, V.: An analytical study of low delay multi-tree-based overlay multicast. In: Proc. of ACM P2P-TV (August 2007)
15. Lelarge, M., Liu, Z., Xia, C.H.: Asyumptotic tail distribution of end-to-end delay in networks of queues with self-similar cross traffic. In: Proc. of IEEE INFOCOM (March 2004)
16. Denuit, M., Genest, C., Marceau, É.: Stochastic bounds of sums of dependent risks. Insurance: Mathematics and Economics 25(1), 85–104 (1999)
17. Schwartz, A., Weiss, A.: Large Deviations for Performance Evaluation: Queues, communication and computing. Chapman and Hall, Boca Raton (1995)
18. Padmanabhan, V.N., Wang, H.J., Chou, P.A.: Resilient peer-to-peer streaming. In: Proc. of IEEE ICNP, pp. 16–27 (2003)
19. Castro, M., Druschel, P., Kermarrec, A.-M., Nandi, A., Rowstron, A., Singh, A.: SplitStream: High-bandwidth multicast in a cooperative environment. In: Proc. of ACM SOSP (2003)
20. Setton, E., Noh, J., Girod, B.: Rate-distortion optimized video peer-to-peer multicast streaming. In: Proc. of ACM APPMS, pp. 39–48 (2005)
21. Bishop, M., Rao, S., Sripanidkulchai, K.: Considering priority in overlay multicast protocols under heterogeneous environments. In: Proc. of IEEE INFOCOM (April 2006)
22. Zegura, E.W., Calvert, K., Bhattacharjee, S.: How to model an internetwork. In: Proc. of IEEE INFOCOM, March 1996, pp. 594–602 (1996)
23. Veloso, E., Almeida, V., Meira, W., Bestavros, A., Jin, S.: A hierarchical characterization of a live streaming media workload. In: Proc. of ACM IMC 2002, pp. 117–130 (2002)
24. Feller, W.: An Introduction to Probability Theory and its Applications. John Wiley and Sons, Chichester (1966)