The Utility of Feedback in Layered Multicast Congestion Control

Sergey Gorinsky   Harrick Vin

Technical Report TR2001-04
Department of Computer Sciences, University of Texas at Austin
Taylor Hall 2.124, Austin, Texas 78712-1188, USA
{*gorinsky, vin*}@*cs.utexas.edu*

February 22, 2001

Abstract—**Layered multicast is a common approach for dissemination of audio and video in heterogeneous network environments. Layered multicast schemes can be classified into two categories – feedback-based and feedback-free – depending on whether or not the scheme delivers feedback to the sender of the multicast session. Advocates of feedback-based schemes claim that the feedback is necessary to match the heterogeneous receiver capabilities efficiently. Supporters of feedback-free schemes believe that feedback introduces significant complexity and that a moderate amount of additional layers can balance any benefit the feedback provides. Surprisingly, there has been no systematic evaluation of these claims. This paper compares feedback-based and feedback-free schemes quantitatively with respect to their abilities to align the provided service to the capabilities of the heterogeneous receivers. We believe that such an evaluation supplies valuable insights and guidelines to the designers of future multicast congestion control protocols.**

1 Introduction

Layered multicast has been suggested as a solution for real-time dissemination of audio and video to heterogeneous receivers. In a layered scheme, the sender encodes media content into a stack of cumulative layers. The capability of a receiver determines which layers it can receive.

Layered multicast schemes can be classified into two categories – feedback-based and feedback-free. Feedback-based schemes discover the receiver capabilities and communicate them to the sender. Based on this feedback, the sender adjusts the layer transmission rates to improve their alignment with the receiver capabilities. Examples of feedback-based schemes include SAMM [8] and SIM [2]. Feedback-free schemes deliver no feedback to the sender: the sender transmits the layers at predetermined constant rates; the receivers indicate to the network their desire to add or drop a layer, and, in response, routers modify their multicast routing tables. RLM [4], RLC [7], and FLID [1] are examples of feedback-free schemes.

While researchers have dedicated substantial efforts to the design of specific schemes, it is not established which approach – feedback-free or feedback-based – is preferable. Advocates of feedback-based schemes claim that feedback is necessary to provide efficient operation in heterogeneous environments. Their opponents believe that feedback-based schemes are inherently more complex and that a small amount of additional layers can offset any benefit the feedback provides. Yet, there has been no systematic evaluation of these claims.

This paper compares feedback-based and feedback-free schemes quantitatively with respect to their abilities to align the provided service to the capabilities of the heterogeneous receivers. Our findings indicate tangible incentives for designing light-weight feedback-based schemes.

2 Model

We consider a multicast session with C different receiver *capabilities* where the capability of a receiver is the maximum fair rate at which the receiver can receive data from the sender. Similarly to earlier studies of layered multicast [5], we represent these capabilities c_i (where $i = 0, \ldots, C-1$) with positive real numbers. We use n_i to denote the number of receivers with capability c_i. Thus, the number n of receivers in the multicast session is given by

$$n = \sum_{i=0}^{C-1} n_i. \tag{1}$$

The sender multicasts the content using up to T cumulative layers. Let t_k (such that $k = 0, \ldots, T-1$) denote the cumulative transmission rate for layer k. We assume that layer 0 is the base layer of the hierarchical data encoding and that, for $k = 1, \ldots, T-1$, layer k refers to the k-th enhancement layer of the encoding. That is, we have $0 < t_0 < t_1 < \ldots < t_{T-1}$.

The key difference between feedback-based and feedback-free schemes is how much information about the receiver capabilities is provided to the sender. Unlike a feedback-based scheme, a feedback-free scheme does not notify the sender about the actual capabilities of the receivers. Thus, to select transmission rates for the feedback-free scheme, the sender can rely only on a priori estimates of the capabilities. These estimates can be obtained from network statistics as well as from the transmission requirements of the content. We model the estimates of the capabilities as a *possible range* $[l,\ vl]$ where l is the *lowest possible capability* ($l > 0$) and v is the *estimated heterogeneity* ($v \geq 1$). We assume that the actual capabilities lie within this possible range, i.e., $l \leq c_i \leq vl$ for each $i = 0, \ldots, C-1$. Note that the size $l(v-1)$ of the possible range can be substantially larger than the span of actual capabilities: e.g., when all the receivers share the same bottleneck link, their capabilities are identical.

2.1 Metrics

For a receiver with capability c_i, we define a *receiver satisfaction* s_i as:

$$s_i = \frac{\max_{t_k \leq c_i}\{0, t_k\}}{c_i}; \quad 0 \leq s_i \leq 1. \tag{2}$$

Thus, to evaluate the satisfaction of a receiver with a scheme, we consider only those layers that do not create congestion. For example, if each layer adds 1 Mbps to the total transmission rate, and the capability of a receiver is 1.25 Mbps, then the receiver can obtain (without causing congestion) only the base layer, and this gives the receiver a satisfaction of $1/1.25 = 80\%$. Since the receiver cannot obtain the enhancement layers in their entirety, they are not considered. In this respect, our metric is similar to the "goodput" measure used in [8] to represent the quality of layered video. If the receiver can not obtain even the base layer, then $s_i = 0$; when the transmission rates match the receiver capability exactly, $s_i = 1$. While we claim no special wisdom in modeling the satisfaction of a receiver, our index is consistent with earlier approaches and, in fact, is an instantiation of more generic metrics such as the inter-receiver fairness [3] and quality of the received signal [6].

To quantify the overall utility of the scheme for the session, we define a *session satisfaction* S as the average of the receiver satisfaction indices of all the receivers in the session:

$$S = \sum_{i=0}^{C-1} \frac{n_i}{n} s_i. \tag{3}$$

Since feedback refines the estimates of the capabilities, it is reasonable if a feedback-based scheme provides a higher session satisfaction than a feedback-free scheme with the same number of layers. The key question is how significantly do the provided session satisfactions differ. Because the units of satisfaction are somewhat arbitrary, a good way to assess the significance of the difference is to measure how many additional layers a feedback-free scheme may need to provide a comparable session satisfaction. We formally define this additional amount of layers as a *layer overhead* d:

$$d = \min_{S_0(T+g) \geq f S_1(T)} \{g\} \tag{4}$$

where $S_0(T+g)$ is the session satisfaction delivered by the feedback-free scheme with $T+g$ layers, $S_1(T)$ denotes the session satisfaction given by the feedback-based scheme with T layers, and f is a *satisfaction similarity* characterizing the closeness of the session satisfactions ($0 < f \leq 1$).

2.2 Compared Schemes

Shacham [6] designed a dynamic programming algorithm that, given C, T, c_i, and n_i, computes an optimal scheme with respect to the session satisfaction. We refer to this scheme as an *Optimal Layering* (OL) scheme and use it as the (best possible) representative of feedback-based schemes.

In feedback-free schemes, the sender knows only the possible range $[l, vl]$ of receiver capabilities and selects the transmission rates to cover it. We examine two feedback-free schemes suggested in the literature:

- *Additive Layering* (AL) scheme, where each enhancement layer increases the cumulative transmission rate by additive $a = \frac{(v-1)l}{T}$:

$$t_k = l + ak = \left(1 + \frac{k}{T}(v-1)\right)l; \tag{5}$$

- *Multiplicative Layering* (ML) scheme, where enhancement layers raise the cumulative transmission rate multiplicatively by factor $m = v^{\frac{1}{T}}$:

$$t_k = l \cdot m^k = l \cdot v^{\frac{k}{T}}. \tag{6}$$

3 Experiments

3.1 Methodology

We pick the values of c_i randomly, under the assumption of uniform distribution, from an interval within the possible range $[l,\ vl]$. We call this interval the *actual span* and characterize it with two parameters, *size h* and *shift* z, which take their values between 0 and 1 (see Figure 1 for examples):

- *Actual span size* h refers to the percentage of the possible range the actual span covers; $h = 1$ when the actual span coincides with the possible range $[l,\ vl]$; if $h = 0$, all the receivers have the same capability;

- *Actual span shift* z specifies the location of actual span within the possible range $[l,\ vl]$. Formally, we define

$z = \frac{x}{(1-h)(v-1)l}$ where x measures the gap between the lowest possible capability l and the actual span while $(1-h)(v-1)l$ is the maximum value of this gap. For instance, $z = 0$ when the lower border of the actual span coincides with l; if $z = 0.5$, the actual span is in the middle of the possible range; when $z = 1$, the upper border of the actual span coincides with the highest possible capability vl.

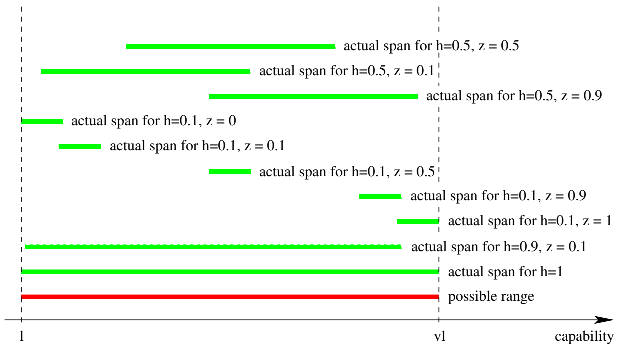


Figure 1: Characterizing the actual capabilities.

We pick the values of n_i randomly, under the assumption of uniform distribution, from interval [1, p] where p is the maximum number of receivers with the same capability.

The default parameter settings in our experiments are as follows: $T = 5$ (up to 5 layers), $C = 50$ (50 different capabilities), $p = 399$ (the number of receivers with a particular capability is picked randomly from interval [1,399]; thus, the expected number of receivers is $\frac{p+1}{2}C = 10000$), $l = 1$, $v = 100$ (the possible range is [1, 100], this can correspond to the range of video rates from 60 Kbps to 6 Mbps), $z = 0.5$, $h = 0.5$ (the actual span is in the middle of the possible range and covers half of it), and $f = 0.99$ (we measure how many additional layers a feedback-free scheme needs to provide a session satisfaction that is 99% of the one given by the OL scheme).

When we vary a parameter, we consider a large number – 100 in most of the experiments – of its settings distributed uniformly throughout the examined range. For each considered setting, we generate 100 (different due to the randomness in our experimental setup) session configurations and compute the session satisfactions provided by the OL, ML, and AL schemes as well as the layer overheads for the ML and AL schemes. We present the results graphically as lines connecting the points that correspond to the averages, over all the generated configurations, of the computed values.

3.2 Results

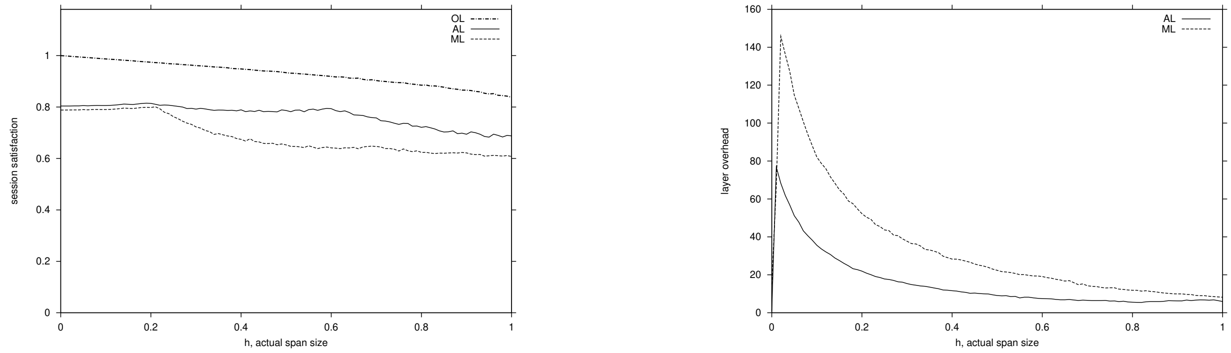
As Figure 2 shows, the feedback-based OL scheme consistently provides a higher average satisfaction than the feedback-free AL and ML schemes. Figure 2a demonstrates that when the actual span of the capabilities is much smaller than their possible range (i.e., for low values of h), OL provides an almost perfect satisfaction while the feedback-free schemes need on the order of 100 additional layers to supply a comparable level of efficiency. As the actual span size h increases, the efficiency of OL decreases, and the layer overheads of AL and ML converge to about 10 layers. With increase of the actual span shift z (see Figure 2b), the satisfaction given by OL improves, and the performance balance between the feedback-free schemes flips: since ML, in comparison to AL, places its layers closer to the lowest possible capability, ML outperforms AL if the value of z is low (i.e., when the actual span is close to l); for larger values of z, AL gives a better satisfaction than ML.

In Figure 2c, we vary C, the number of different capabilities, while keeping the expected number of receivers close to 10000. When the number of different capabilities is at most the number of layers, OL yields the 100% satisfaction. For larger values of C, the satisfaction given by OL declines slightly, and the satisfactions provided by AL and ML remain on lower but relatively constant levels. When the numbers of layers and different capabilities are of the same order, the layer overhead for the feedback-free schemes is large but decreases as the number of capabilities grows. In contrast, we observed that n, the number of receivers, makes virtually no impact on the performance of the OL, AL, and ML schemes.

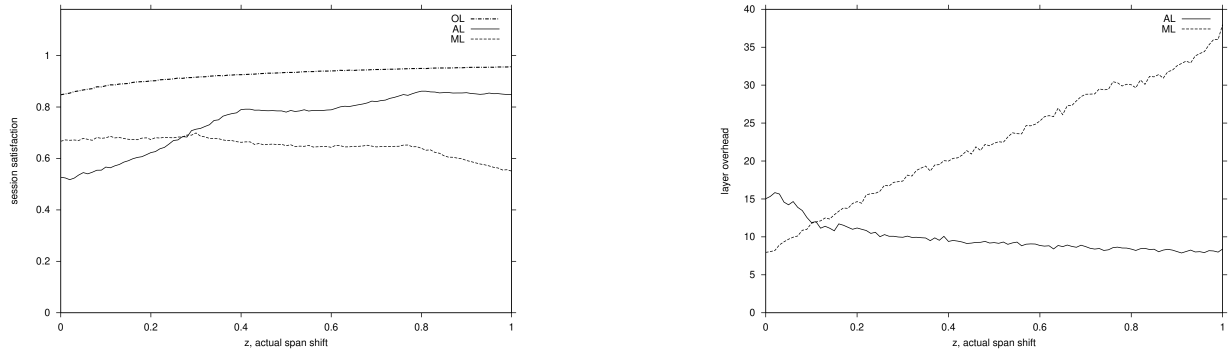
If the capabilities are the same and known a priori, all the schemes provide the 100% satisfaction. Figure 2d shows that when the estimated heterogeneity v (and, due to the fixed value of h, the actual heterogeneity too) increases even slightly, OL provides a considerably higher satisfaction than the AL and ML schemes that need at least 9 additional layers to supply similar satisfactions. We observed that the lowest possible capability l, unlike the estimated heterogeneity v, does not affect the performance of the OL, AL, and ML schemes; this indicates that the units of capability measurements are irrelevant.

Figure 3 shows that as T, the number of layers, grows, ML and AL fail to reach the satisfaction provided by OL. Moreover, as T increases, they incur greater layer overheads to provide comparable satisfactions. Another interesting observation is that less than 10 layers enable OL to bring the satisfaction closely to 100%.

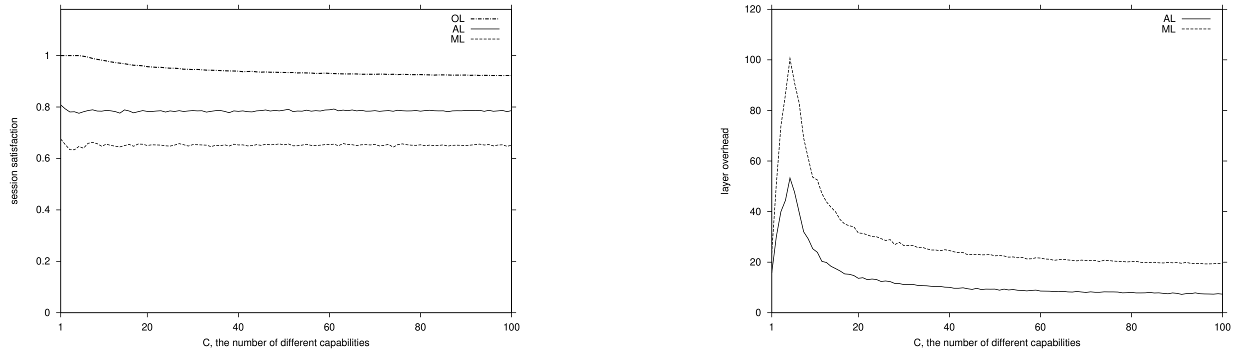
Figure 4 studies the dependence of the layer overheads on the satisfaction similarity f. To reach 85% of the satisfaction provided by the OL scheme, AL and ML need about 1 and 4 additional layers respectively. As we increase the satisfaction similarity f, the layer overheads grow. To provide the same satisfaction as OL, the AL and ML schemes require about 11 and 26 additional layers respectively.



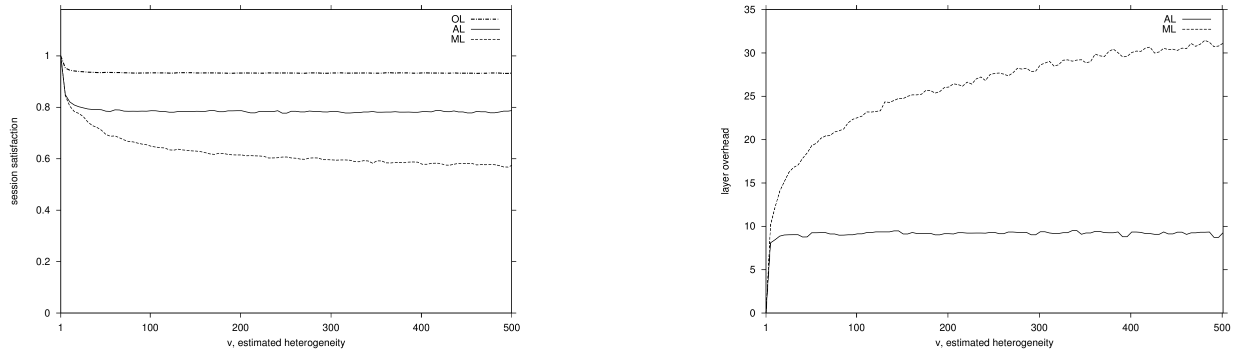
(a) $T = 5$, $C = 50$, $p = 399$, $l = 1$, $v = 100$, $z = 0.5$, and $f = 0.99$.



(b) $T = 5$, $C = 50$, $p = 399$, $l = 1$, $v = 100$, $h = 0.5$, and $f = 0.99$.



(c) $T = 5$, $p = \lfloor \frac{20000}{C} \rfloor - 1$, $l = 1$, $v = 100$, $z = 0.5$, $h = 0.5$, and $f = 0.99$.



(d) $T = 5$, $C = 50$, $p = 399$, $l = 1$, $z = 0.5$, $h = 0.5$, and $f = 0.99$.

Figure 2: The average session satisfactions and layer overheads.

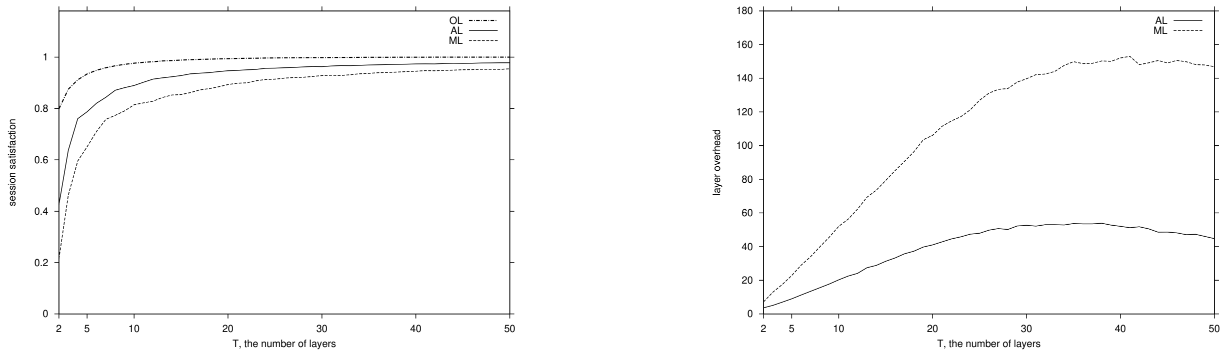


Figure 3: The dependence on the number of layers for $C = 50$, $p = 399$, $l = 1$, $v = 100$, $z = 0.5$, $h = 0.5$, and $f = 0.99$.

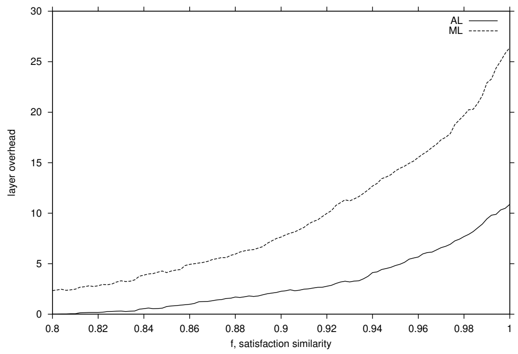


Figure 4: The impact of the satisfaction similarity for $T = 5$, $C = 50$, $p = 399$, $l = 1$, $v = 100$, $z = 0.5$, and $h = 0.5$.

4 Discussion

We observed that OL needs a small number (in single digits) of layers to provide an almost perfect satisfaction while AL and ML require many (up to the order of 100 in the scenarios when the number of different capabilities is small or when the capabilities are homogeneous in comparison to the a priori estimates) additional layers to provide a comparable satisfaction. The larger number of layers means a higher encoding complexity at the sender, larger decoding complexity in the receivers, and additional state and processing overhead in the routers (since a feedback-free scheme uses a separate multicast group per layer). Besides, increasing the number of layers usually reduces the compression ratio, and this increases the transmission rates required to provide the same perceived quality. Furthermore, feedback-free schemes need even more additional layers to be responsive in dynamic environments [1].

Feedback-based schemes have their own share of extra complexity associated with measuring and communicating the capabilities to the sender. In particular, to resolve the problem of feedback implosion in large sessions, many feedback-based schemes use routers or designated servers to aggregate feedback. This infrastructure upgrade represents a detriment to the deployment of such schemes. Also, the examined OL is the best possible feedback-based scheme since it assumes that the sender knows all the capabilities exactly. The precise measurement of receiver capabilities is difficult to realize in a heterogeneous multicast session. The common method of probing for capacities (increase the transmission if there is no congestion, decrease the transmission when congestion is detected) does not reveal all the capabilities if their number exceeds the number of layers. In addition, since the sender needs to adjust dynamically the transmission rates of the layers, the encoding complexity may increase.

While the exact costs of additional layering or feedback-based adjustment depend on the specifics of the coding and networking algorithms, our findings indicate tangible incentives for designing light-weight feedback-based schemes. Layered multicast can greatly benefit even from limited feedback: if the sender learns the actual span of the capabilities (a much easier task than discovering all the capabilities), it needs a much smaller number of layers to provide a high satisfaction than when it knows only their possible range. Thus, a successful solution can be a hybrid of the feedback-free and feedback-based paradigms: feedback informs – possibly on a coarse timescale – the sender about the lowest and highest capabilities in the session, and the sender uses this range to set – in the feedback-free fashion – the transmission rates of the layers.

References

[1] J. Byers, M. Frumin, G. Horn, M. Luby, M. Mitzenmacher, A. Roetter, and W. Shaver. FLID-DL: Congestion Control for Layered Multicast. In *Proceedings NGC 2000*, November 2000.

[2] S. Gorinsky, K.K. Ramakrishnan, and H. Vin. Addressing Heterogeneity and Scalability in Layered Multicast Congestion Control. Technical Report TR2000-31, Department of Computer Sciences, The University of Texas at Austin, November 2000.

[3] T. Jiang, E. W. Zegura, and M. H. Ammar. Inter-Receiver Fairness: A Novel Performance Measure for Multicast ABR Sessions. In *Proceedings ACM SIGMETRICS'98*, June 1998.

[4] S. McCanne, V. Jacobson, and M. Vetterli. Receiver-driven Layered Multicast. In *Proceedings ACM SIGCOMM'96*, August 1996.

[5] D. Rubenstein, J. Kurose, and D. Towsley. The Impact of Multicast Layering on Network Fairness. In *Proceedings ACM SIGCOMM'99*, September 1999.

[6] N. Shacham. Multipoint Communication by Hierarchically Encoded Data. In *Proceedings IEEE INFOCOM'92*, May 1992.

[7] L. Vicisano, L. Rizzo, and J. Crowcroft. TCP-like Congestion Control for Layered Multicast Data Transfer. In *Proceedings IEEE INFOCOM'98*, March 1998.

[8] B. Vickers, C. Albuquerque, and T. Suda. Source-Adaptive Multi-Layered Multicast Algorithms for Real-Time Video Distribution. *IEEE/ACM Transactions on Networking*, December 2000.