

Achieving Inter-Receiver Fairness Utilizing Layered Media Multicast Control (LMMC)

Homayoun Yousefi'zadeh Hamid Jafarkhani

Department of Electrical Engineering and Computer Science
University of California, Irvine
[hyousefi, hamidj]@uci.edu

Abstract—In this paper, we present an analytical solution to the problem of rate allocation and receiver partitioning in layered media systems. The framework of our proposed protocol Layered Media Multicast Control (LMMC) determines the layer rates and the corresponding partitioning of the receivers maximizing a mathematically well-behaved approximation of the so-called max-min fairness metric.

Index Terms—Layered Media, Multicast IP Networks, Heterogeneity, Optimality, Fairness, Rate Allocation, Receiver Partitioning.

I. INTRODUCTION

Transmitting real-time compressed digital media over multicast IP networks has been the subject of heavy research in the recent years as surveyed by Li et al. in [6] and the references cited therein. Layered media streaming over multicast IP networks allows a multimedia source for transmitting its layered bitstream over different multicast groups. Intended receivers then join at least the base layer group and as many enhancement layer groups as their capacities allow. The idea of layered media streaming over multicast IP networks was first proposed by Deering et al. [3] in the context of multicast routing and further enhanced by McCanne et al. [8] in the context of RLM protocol, Amir et al. [1] in the context of SCUBA protocol, and Li et al. [7] in the context of rate control aspect of LVMR protocol. We provide a comprehensive review of literature articles in [14].

The main objective of this paper is to provide an analytical framework for the partitioning strategy and rate allocation of distributing multimedia content over multicast IP networks. The objective is achieved in the context of Layered Media Multicast Control (LMMC) protocol. In this study, we do not address the issues of error control, inter-session fairness, and congestion control. We refer the interested reader to our work of [15] and [12] for a treatment of those subject. An outline of the paper follows. In Section II, we formulate the problem of receiver partitioning and rate allocation by means of max-

min fairness extrapolation. In Section III, we utilize a two-phase approach for solving the problem of Section II. In the first phase, we analytically solve the optimal rate allocation problem assuming a given partitioning. In the second phase, we use the allocated rates of the first phase to obtain a near-optimal partitioning strategy. In Section IV, we introduce an iterative approach relying on the solution of Section III to reach a near-optimal solution. Section V focuses on performance evaluation and includes simulation results along with practical considerations. Finally, Section VI contains concluding remarks.

II. FORMULATION OF THE PROBLEM BY MEANS OF FAIRNESS EXTRAPOLATION

In this section, we formulate the general rate allocation problem of layered media sessions in a manner similar to the formulation of [4] and [13]. Consider a multicast media session with a partitioning of the receivers into K groups. For a media session with N receivers and K groups, a set $P = \{G_1 | \dots | G_K\}$ is called a partitioning of the receiver set $R = \{1, \dots, N\}$ if P is a decomposition of the set R into a family of disjoint sets. The term group rate is used to denote the aggregate receiving rate of a receiver in the group while the term layer rate is used to denote the transmission rate to a specific layer. For an ordered partitioning of the receivers into K groups with ordered group rates of g_1, g_2, \dots, g_K such that $g_1 \leq g_2 \leq \dots \leq g_K$, the layer rates of a layered media session are calculated in the form of

$$g_1, g_2 - g_1, g_3 - g_2, \dots, g_K - g_{K-1} \quad (1)$$

A receiver in group k subscribes to layers 1 through k receiving an aggregate rate of g_k . Reference [14] includes a treatment of our formulation in the case of replicated media sessions.

The optimization problem is formulated by means of defining a per receiver max-min fairness utility with the objective of maximizing the session utility defined as the

sum of receiver utilities over the layered media session. Each receiver is assumed to have an isolated multi-rate max-min fair rate of r_i as described in [10]. This is the reception rate of the receiver and is typically determined by a network bottleneck link from the source to the receiver or the receiver itself. Assuming $r_1 \leq r_2 \leq \dots \leq r_N$, denoting the loss tolerance of receiver i by L_i , and considering the max-min fairness utility of [5] defined as

$$F(r_i, g_k) = \frac{\min(r_i, g_k)}{\max(r_i, g_k)} \quad (2)$$

the rate allocation optimization problem can be formulated as

$$\begin{aligned} \max_{g_1, \dots, g_K} IRF_T &= \max_{g_1, \dots, g_K} \sum_{k=1}^K IRF_k \\ &= \max_{g_1, \dots, g_K} \sum_{k=1}^K \sum_{i \in G_k} \frac{\min(r_i, g_k)}{\max(r_i, g_k)} \end{aligned} \quad (3)$$

$$\text{Subject To: } g_k \leq \frac{r_i}{1 - L_i} \quad (4)$$

for the optimal receivers partitioning of $P^* = \{G_1^* | G_2^* | \dots | G_K^*\}$ where $i \in G_k$ and $k = 1, \dots, K$. We note that the solution to the problem leads to the identification of the optimal group rates $g_1^*, g_2^*, \dots, g_K^*$. The objective of both heuristics of [4] and the dynamic programming algorithm of [13] which was derived based on the proposition of [9] is to determine the optimal partitioning and the optimal layer rate allocations such that the utility function of (3) is maximized while Constraint (4) is satisfied. Considering the fact that the max-min fairness utility function of (3) is not continuously differentiable, we introduce an extrapolation technique to replace the utility function of (3) with a mathematically well-behaved function over the set of real numbers. By mathematically well-behaved, we mean a continuously differentiable function with no real poles. We select a function $E(r_i, g_k)$ in the form of

$$E(r_i, g_k) = \frac{(2+a)r_i g_k}{g_k^2 + ar_i g_k + r_i^2} \quad (5)$$

and note that not only $E(r_i, g_k)$ is well-behaved for parameter a satisfying the boundary condition $-2 < a < 2$, but it satisfies the boundary and maximum conditions of function $F(r_i, g_k)$. Figure (1) shows generic sample plots of $F(r, g)$ and $E(r, g)$ versus g for a fixed r . Next, we utilize least square error estimation technique to identify the optimal choice of parameter a within the interval of interest $[0, \frac{r_i}{1-L_i}]$ such that $E(r_i, g_k)$ is the closest estimate of $F(r_i, g_k)$. As the result of solving the least square prob-

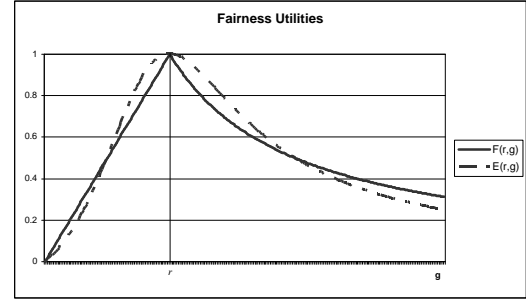


Fig. 1. Plots of $F(r, g)$ and $E(r, g)$ versus g for a fixed r .

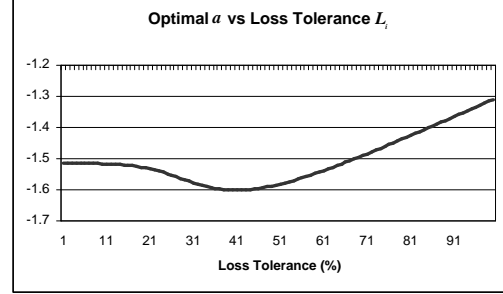


Fig. 2. The plot of optimal a versus loss tolerance L_i .

lem of

$$\begin{aligned} \min_a \{ & \int_0^{r_i} [\frac{(2+a)r_i g_k}{g_k^2 + ar_i g_k + r_i^2} - \frac{g_k}{r_i}]^2 dg_k \\ & + \int_{r_i}^{\frac{r_i}{1-L_i}} [\frac{(2+a)r_i g_k}{g_k^2 + ar_i g_k + r_i^2} - \frac{r_i}{g_k}]^2 dg_k \} \end{aligned} \quad (6)$$

we have observed that the optimal value of parameter a is only a function of loss tolerance L_i . Figure (2) displays the solution for the optimal value of parameter a in Equation (6) for different values of L_i in the interval of interest. We now formulate an alternative rate allocation problem as

$$\begin{aligned} \max_{g_1, \dots, g_K} IRFA_T &= \max_{g_1, \dots, g_K} \sum_{k=1}^K \sum_{i \in G_k} \frac{(2+a)r_i g_k}{g_k^2 + ar_i g_k + r_i^2} \\ \text{Subject To: } g_k &\leq \text{BWL}_k \quad k = 1, \dots, K \end{aligned} \quad (7)$$

where BWL_k is defined as $\text{BWL}_k \equiv \min_{i \in G_k} \frac{r_i}{1-L_i}$.

III. LMMC TWO-PHASE APPROACH TO RATE ALLOCATION AND PARTITIONING

A. Phase 1: Rate Allocation Solution

In this section, we provide an analytical solution to the optimal rate allocation problem formulated by Equation (7) and Constraint (8) that can be applied to both layered media and replicated media sessions. Considering the formulation of the problem, we note that it can be decom-

posed to K independent optimization problems as

$$\begin{aligned} \max_{g_k} IRFA_k &= \max_{g_k} \sum_{i \in G_k} \frac{(2+a)r_i g_k}{g_k^2 + ar_i g_k + r_i^2} \quad (9) \\ \text{Subject To: } &g_k \leq \text{BWL}_k \end{aligned}$$

where $k = 1, \dots, K$. The set of optimization problems of Equation (9) can then be solved by finding the roots of the equations

$$\frac{\partial IRFA_k}{\partial g_k} = \sum_{i \in G_k} \frac{(2+a)r_i(r_i^2 - g_k^2)}{(g_k^2 + ar_i g_k + r_i^2)^2} = 0 \quad (10)$$

and extracting the first local maximum satisfying Constraint (8).

The overall time complexity of solving for the optimum point of equation set (10) is $\mathcal{O}(KN \log N)$. The time complexity of the algorithm is by far better than $\mathcal{O}(N^2)$ the complexity of the dynamic programming algorithms of [13] and [9]. In [14], we investigate the impacts of facing some of the source and the receiver limitation scenarios when solving LMMC optimal control problem. Due to lack of space, we skip the discussion here.

B. Phase 2: Partitioning Strategy

In [4], a heuristic approach to partition the receivers of a layered media session is proposed without introducing a formal algorithm. In addition, the dynamic programming algorithm of [13] provides an optimal receiver partitioning strategy for a media session while computing the optimal layer rates. In this section, we introduce a near-optimal partitioning strategy with a time complexity of $\mathcal{O}(N)$ and show that our partitioning strategy maximizes the session utility for a set of given group rates. It is important to note that the result of Theorem (1) of [13], assures that the order of the resulting partitioning of this section are preserved.

Considering the general objective of maximizing the session utility of Equation (7), we observe that an optimal receiver partitioning strategy has to assign the receiver with isolated rate r_i to the group G_k with group rate g_k if

$$E(r_i, g_k) \geq E(r_i, g_l) \quad l \in \{1, \dots, K\} \quad (11)$$

We now translate the latter observation to a simple group assignment mechanism. Given the group rates $\{g_1, \dots, g_K\}$, we first plot the family of functions $E(r_i, g_k)$ versus r_i with parameters g_k where $k = 1, \dots, K$. Fig. 3 shows the sample plots for $K = 3$. Next, we find the intersection points of every two functions with consecutive group rates g_k and g_{k+1} .

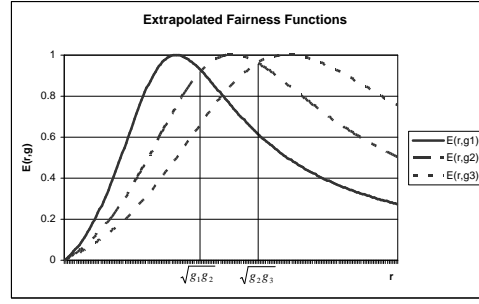


Fig. 3. Sample plots of $E(r_i, g_k)$ versus r_i for three values of g_k .

The values of r_i at the intersection points are obtained by finding the roots of $E(r_i, g_k) = E(r_i, g_{k+1})$ for the variable r_i . The solution to the equation can be expressed in the form of $r_i = \sqrt{g_k g_{k+1}}$ after a bit of algebraic manipulation. We now pay special attention to the key characteristic of the intersection points of the curves that we refer to as partitioning thresholds.

Theorem 3.1: *The value of the receiver utility as defined in Equation (5) is maximized for the choice of the group rate g_k for $k > 1$ and $k < K$ over the set of given group rates $\{g_1, \dots, g_K\}$ if $\sqrt{g_{k-1}g_k} < r_i \leq \sqrt{g_k g_{k+1}}$. The receiver utility is maximized for the choice of the group rate g_1 if $r_i \leq \sqrt{g_1 g_2}$ and for the choice of the group rate g_K if $r_i > \sqrt{g_{K-1}g_K}$.*

A proof of Theorem 3.1 is provided in [14]. We realize that Theorem 3.1 provides the best overall repartitioning strategy for an unconstrained problem. Considering the constraint set of (8), we notice that moving a receiver from group $k-1$ to group k can potentially introduce a new constraint for group k . If the new constraint is far from the existing optimal group rate g_k^* , it can cause a reduction in the sum of the utilities of groups $k-1$ and k after repartitioning. We use the following heuristic rule to address the issue. A receiver is allowed to move from group $k-1$ to group k if **one** of the following conditions holds

$$\left(\frac{r_i}{1 - L_i} \geq g_k^* \right) \parallel \left(C_2 g_k^* < \frac{r_i}{1 - L_i} < g_k^* \right) \quad (12)$$

In practice, we have observed that setting $C_2 \in [0.5, 0.9]$ yields good results for different values of receivers' loss tolerance. The same approach can be used to avoid a similar problem when moving a receiver from group k to $k-1$. Reference [14] includes another statistical heuristic rule.

LMMC near-optimal partitioning strategy then reorders the receivers such that each receiver is moved to a group maximizing its individual utility according to Theorem 3.1 and condition (12). Note that we use the term near-

optimal to mathematically consider the effect of solving the constrained problem besides the fact that LMMC partitioning strategy is optimal in the case of solving an unconstrained problem. Such an algorithm introduces a time complexity order of $\mathcal{O}(KN)$.

IV. LMMC NEAR-OPTIMAL ITERATIVE SOLUTION

In this section, we introduce an iterative approach that can reach a near-optimal solution considering the fact that the solution to our two-phase optimal problem is sub-optimal due to the impact of our phasing approach. A near-optimal solution can be achieved by iteratively applying the partitioning results of the second phase to the first phase and solving the optimal rate allocation problem again with the alternative partitioning strategy. The optimal layer rates of the first phase can then be applied to the near-optimal partitioning strategy of the second phase to partition the receivers according to the new set of rates. In what follows we propose the formal iterative algorithm of LMMC.

LMMC Iterative Rate Allocation-Partitioning Algorithm:

- Step 1: Start from an initial ordered partitioning of the receivers by uniformly distributing the receivers among the existing groups. In addition, set the initial iteration number $j = 0$ and the maximum number of iterations j_{max} .
- Step 2: Calculate the optimal group rates $g^* = \{g_1^*, \dots, g_K^*\}$ and the resulting session utility $IRFA_T$ by numerically solving the system of equations (10) for the local maximum satisfying condition (8). Save the previously calculated $IRFA_T$ in variable q_1 and the currently calculated $IRFA_T$ in variable q_2 .
- Step 3: If $\frac{|q_1 - q_2|}{q_1} < \delta$ or $j > j_{max}$ STOP.
- Step 4: *for* $(k = 2 \text{ to } K)$ {
 - Calculate the partitioning threshold $\sqrt{g_{k-1}g_k}$.
 - Repartition groups $k - 1$ and k . For every receiver belonging to groups $k - 1$ or k and isolated rate r_i , assign the receiver to group k if $r_i > \sqrt{g_{k-1}g_k}$ and condition (12) hold. Otherwise, assign the receiver to group $k - 1$.
 - Calculate the new optimal sending rate of group k according to the new partitioning.
- $\}$ /* *for* $(k = 2 \text{ to } K)$ */
- Step 5: Go back to Step 2.

We note that the time complexity of our iterative algorithm is $\mathcal{O}(IKN \log N)$ where I indicates the number of iterations. We have observed in our experiments that the value of I varies in the range of one to

twenty iterations. Comparing the overall complexity of LMMC algorithm with that of the dynamic programming algorithm of [13] $\mathcal{O}(N^3)$, LMMC algorithm achieves a much lower complexity. It is important to note that the implementation of the algorithm is fairly practical aside from its relative theoretical complexity. The following theorem guarantees the convergence of “LMMC Iterative Rate Allocation-Partitioning Algorithm”.

Theorem 4.1: *The convergence of “LMMC Iterative Rate Allocation-Partitioning Algorithm” mentioned in this section is guaranteed.*

A proof of Theorem 4.1 is provided in [14]. Intuitively, LMMC algorithm is employing steepest descent optimal strategy and is guaranteed to reach a near-optimal point if such a point exists. As shown by our numerical results, proper choice of the parameters in (12) leads to a fast convergence to a local optimal solution.

V. PERFORMANCE ANALYSIS

In this section, we compare the numerical LMMC partitioning and rate allocation results to those of the dynamic programming algorithm of [13]. The metrics of interest in our comparisons are the session utility and the experiment runtime as an indication of the time complexity. We note that the space complexity of the LMMC algorithm in our implementation is $\mathcal{O}(N)$ where as the space complexity of DP algorithm proposed in [13] is $\mathcal{O}(N^2)$.

In our simulations, we utilize tri-, quad-, and pent-modal distributions to generate receiver isolated rates. We select the means of distributions from the set of $\{128Kbps, 1Mbps, 10Mbps, 100Mbps, 1Gbps\}$ to properly represent ISDN, Cable/DSL, low-speed LAN, high-speed LAN, and Gigabit LAN users. For each distribution, we also set the standard deviation of the distribution at 20% of the mean value yielding disjoint successive distributions with greater than 99.7% certainty.

Fig. 4 and Fig. 5 compare sample results of LMMC algorithm with those of the DP algorithm of [13]. In each experiment, we have considered the same loss tolerance for all of the receivers of the session. The first pair of curves compare session utilities of the two techniques utilizing the fairness function of Equation (5) for LMMC and the fairness function of Equation (2) for DP. A review of the sample results of the figures shows a difference of less than 10% between the raw session utilities of the LMMC and the DP algorithms. Considering the fairness approximation as well as the deployment of a gradient based method by LMMC, the experiments indicate the robustness of LMMC iterative algorithm. The second pair

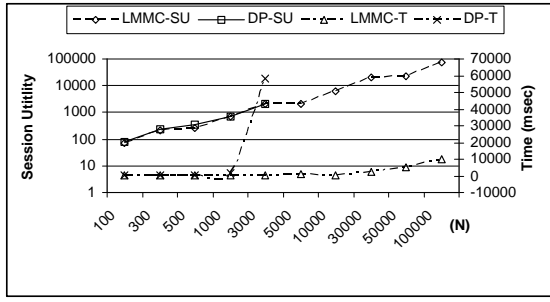


Fig. 4. Session Utility (SU) and Time (T) comparison of LMMC and DP versus number of receivers (N) for 3 categories of receiver isolated rates, $K=3$, and loss tolerance of 10%.

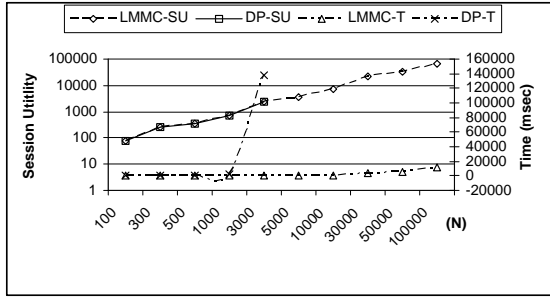


Fig. 5. Session Utility (SU) and Time (T) comparison of LMMC and DP versus number of receivers (N) for 5 categories of receiver isolated rates, $K=4$, and loss tolerance of 20%.

of curves display the runtime of the experiments. In this area, a review of the results reveals the great performance advantage of LMMC over DP. We note that the nonlinear increase in the runtime of the DP algorithm indicates the threshold at which the so-called curse of dimensionality prevents the DP algorithm from further scaling. We argue that although the specific numbers of our experiments are related to the capabilities of our host server, the same qualitative behavior is observed in general. We provide the results of our other experiments in [14].

We note that utilizing the lightweight feedback schemes of [11], we can further improve the efficiency of our approach by focusing on the discovery of the span of receivers isolated rates rather than all of the isolated rates. Such schemes are attractive in scenarios for which the span of isolated rates is much smaller than the bandwidth range useful for multicast content. Another set of such scenarios are the ones for which the number of layers and the number of different isolated rate ranges are roughly the same. In [14], we address the issue of feedback implosion when reporting the isolated rates and loss tolerance of the receivers by employing some of the well-known feedback suppression and aggregation literature techniques. The same article also includes a discussion of what the best strategy is for selecting the number of layers.

VI. CONCLUSION

In this article, we analytically solved the problem of rate allocation and partitioning for layered media systems within the context of Layered Media Multicast Control (LMMC) protocol. For a given session bandwidth, we aimed at providing a two-phase iterative solution converging to a near-optimal solution. Taking into consideration the scalability of LMMC techniques, we showed that the techniques could be effectively adopted in different size point-to-multipoint groups as well as different speeds of load change in the network. Finally, we evaluated the performance of LMMC solution and illustrated its applicability in realistic network topologies through the use of simulations.

REFERENCES

- [1] E. Amir, S. McCanne, R. Katz, "Receiver-Driven Bandwidth Adaptation for Light-Weight Sessions," *ACM Trans. on Multimedia*, November 1997.
- [2] S. Cheung, M. H. Ammar, X. Li, "On the Use of Destination Set Grouping to Improve Fairness in Multicast Video Distribution," *In Proc. of IEEE INFOCOM*, March 1996.
- [3] S.E. Deering, D.R. Cheriton, "Multicast Routing in Datagram Internetworks and Extended LANs," *ACM Trans. on Computer Systems*, May 1990.
- [4] T. Jiang, M. H. Ammar, E. W. Zegura, "On the Use of Destination Set Grouping to Improve Inter-Receiver Fairness for Multicast ABR Sessions," *In Proc. of IEEE INFOCOM*, March 2000.
- [5] T. Jiang, M. H. Ammar, E. W. Zegura, "Inter-Receiver Fairness: A Novel Performance Measure for Multicast ABR Sessions," *In Proc. of ACM SIGMETRICS*, June 1998.
- [6] X. Li, M. Ammar, S. Paul, "Video Multicast over the Internet," *IEEE Network Magazine*, April 1999.
- [7] X. Li, S. Paul, M. H. Ammar, "Layered Video Multicast with Retransmissions (LVMR): Evaluation of Hierarchical Rate Control," *In Proc. of IEEE INFOCOM*, 1998.
- [8] S. McCanne, V. Jacobson, M. Vetterli, "Receiver Driven Layered Multicast," *In Proc. of ACM SIGCOMM*, September 1996.
- [9] N. Shacham, "Multipoint Communication by Hierarchically Encoded Data," *In Proc. of IEEE INFOCOM*, May 1992.
- [10] D. Rubenstein, J. Kurose, D. Towsley, "The Impact of Multicast Layering on Network Fairness," *In Proc. of ACM SIGCOMM*, September 1999.
- [11] S. Gorinsky, H. Vin, "The Utility of Feedback in Layered Multicast Congestion Control," *In Proc. of NOSSDAV* 2001.
- [12] L. Vicisano, L. Rizzo, J. Crowcroft, "TCP-Like Congestion Control for Layered Multicast Data Transfer," *In Proc. of IEEE INFOCOM*, March 1999.
- [13] Y. R. Yang, M. S. Kim, S. S. Lam, "Optimal Partitioning of Multicast Receivers," *In Proc. of the 8th International Conference on Network Protocols*, November 2000.
- [14] H. Yousefi'zadeh, H. Jafarkhani, A. Habibi "Layered Media Multicast Control (LMMC): Rate Allocation and Partitioning," Submitted for Publication, 2002. Available at <http://www.ece.uci.edu/~hyousefi/publ/lmmcRAP.pdf>
- [15] H. Yousefi'zadeh, H. Jafarkhani, "Real-Time Error Recovery Utilizing Layered Media Multicast Control (LMMC)," *In Proceedings of IEEE ICC*, 2003.