Reducing Delivery Delay in HRM Tree

Sang-Seon Byun and Chuck Yoo

Department of Computer Science and Engineering, Korea University,
An-am dong, Seong-buk gu, Seoul, Korea
{ssbyun, hxy}@os.korea.ac.kr

Abstract. In hierarchical reliable multicast schemes, the number of repair proxies and their locations influence the delivery delay. Low delivery delay is essential for the transmission of real time media. In this paper, we propose a method to decide optimal locations of repair proxies that minimizes the mean delivery delay of all receivers in heterogeneous network using a dynamic programming approach. The evaluation results of our optimal proposal in a simulation topology show that the mean delivery delay of all receivers can be reduced by about 10ms in network size of 1000 nodes. Our method can be used by network providers in order to reduce delivery delay in their HRM network.

1 Introduction

Many hierarchical reliable multicast (HRM) protocols deploy repair proxies that perform local recovery and feedback consolidation. Repair proxy can be set up as an exclusive server [1][2][3] or can be designated among adequate receivers [4][5]. The performance of HRM is evaluated by 1) mean delivery delay (or expected delivery delay), 2) bandwidth overhead due to local recovery and feedback consolidation, and 3) interreceiver fairness [6][7][23][27]. Delivery delay is the time that is required to successfully transmit a packet from the sender to a receiver. Mean delivery delay is a packet delivery latency between the sender and a receiver with reflection of link heterogeneity and locations of repair proxies. Inter-receiver fairness is measured to estimate the diversity of each receiver's delivery delay. All these three metrics are affected by the locations of repair proxies. For example, if a proxy is adjacent with a receiver that is susceptible to packet losses and if the proxy is robust with packet losses, the recovery and feedback traffics are limited to the proxy's domain. Addition-ally, packet recovery time can be reduced, and therefore better mean delivery delay and inter-receiver fairness can be obtained.

Related with placement of proxies, many researches have focused on minimizing bandwidth overhead caused by recovery and feedback traffics [8][9] [10][13] and load balancing among proxies [9][11]. Some researches have not considered optimal placement of proxies [12][13]. Also, they assume that every link has uniform delivery delay and loss rate. Reference [14] suggests a method to localize proxies to minimize web distribution time using dynamic programming formulation. However, in order to reduce combinatorial complexity, the available location of a proxy is limited to some area of the web distribution tree.

M. Gavrilova et al. (Eds.): ICCSA 2006, LNCS 3981, pp. 1189–1198, 2006.

Low delivery delay is an important issue of real time applications like multi-conference system [25], interactive distributed simulations, distributed games or delivery of MPEG I-frames [22]. Besides time constraint, low delays are vital for providing high throughput with a window-based sending scheme [24], and the low delay is essential to increase responsiveness in web service. Additionally, heterogeneous loss rate and delivery delay have big impacts on the placement of proxies.

In this paper, we propose a scheme that determines optimal locations of repair proxies to minimize mean delivery delay using dynamic programming formulation.

The rest of paper is organized as follows. In section 2, we explain the HRM model used in this paper and describe mean delivery delay model. In section 3, we present our dynamic programming formulation for optimal placement of proxies. In section 4, comparison by numerical evaluation is given. Finally, concluding remarks and future work are presented in section 5.

2 HRM and Delivery Delay Model

In this section, we describe HRM and expected delivery delay model. In order to determine locations of proxies to minimize mean delivery delay, both HRM and mean delivery delay model deal with link heterogeneity and locations of proxies.

2.1 HRM Model

It is assumed that the HRM model in this paper has the following characteristics (Fig. 1).

- The root of a multicast tree is the unique source, all leaves are receivers, and all intermediate nodes can have proxy [1][2][3].
- The topology of control tree is identical to that of its underlying multicast tree (IP multicast tree), and loss probabilities and delivery delays at the links of the control tree are given. Reference [16], [17] and [18] describe a way of establishing a control tree that is identical to its underlying multicast tree and how to collect link loss statistics. Especially, reference [26] proposes a method to determine the per link loss rates in the logical multicast tree. Due to dynamic nature of memberships and network environments, it is hard to acquire the exact link statistics quickly in real world. However, using our proposed method, the locations of proxies can be obtained in a few tens of seconds in the network size of 1000 nodes. Finally, even if topology and loss rates are not known in real life, our method may still be useful for comparison and assessment purpose.
- The control tree is partitioned into subtrees that form a hierarchy rooted at the source. All nodes in a subtree are combined into a subgroup, and each subgroup has a proxy located at the root of its subtree. The source always has a proxy by default. These features are deployed in [1], [2], [3], [5], [8] and [19].
- A proxy multicasts the original data to its own subgroup. Each receiver sends feedback (NACK) to its proxy when a packet loss is detected, and the proxy retransmits the lost packet to the whole subgroup. Neither flow control nor

congestion control is considered. All feedback packets are delivered via an out-of-band channel, so all feedback packets are delivered safely to proxies.
- Feedbacks and transmissions/retransmissions are limited only between a proxy and the receivers of its subgroup and they do not reach receivers/proxies of any other subgroup. For this purpose, a new multicast address per subgroup is assigned [2], or TTL (Time to live) may be used to scope subgroup [4]. Additionally subcasting and TTL scoping can be used simultaneously [5].

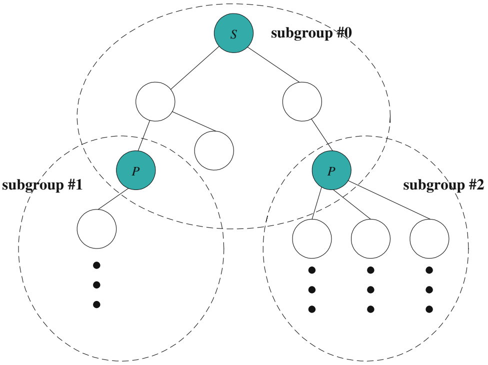


Fig. 1. HRM model. We consider a single-source multicast tree where the root is the unique source. The tree is divided into subgroups and feedbacks and transmissions/retransmissions are limited only between a proxy and the receivers of its subgroup.

2.2 Mean Delivery Delay Model

In this paper, we use a mean delivery delay model in Fig. 2, which reflects heterogeneity and locations of proxies. No queuing and transmission delay is considered in this model and neither flow control nor congestion control is considered. It is assumed that packet losses at each link are independent [12]. A summary of the used notations is given in Table 1.

$E(N_{a,b})$ can be written as follows, where t is inter-packet delay(gap):

$$E(N_{a,b}) = \frac{1 - \prod\limits_{i \in \psi(a,b)} (1 - p_i)}{\prod\limits_{i \in \psi(a,b)} (1 - p_i)} \times t + L_{a,b} \tag{1}$$

$E(D_{S,R(w)})$ can be written in two cases as follows:

1) In case $\pi(S, R(w)) = S$ (S is unique proxy between S and $R(w)$):

$$\begin{aligned} E(D_{S,R(w)}) = &\left[\prod_{i \in \psi(S,R(w))} (1 - p_i) \right] L_{S,R(w)} + \\ &\left[1 - \prod_{i \in \psi(S,R(w))} (1 - p_i) \right] \left(E(N_{S,R(w)}) + E(D_{S,R(w)}) \right) \end{aligned} \tag{2}$$

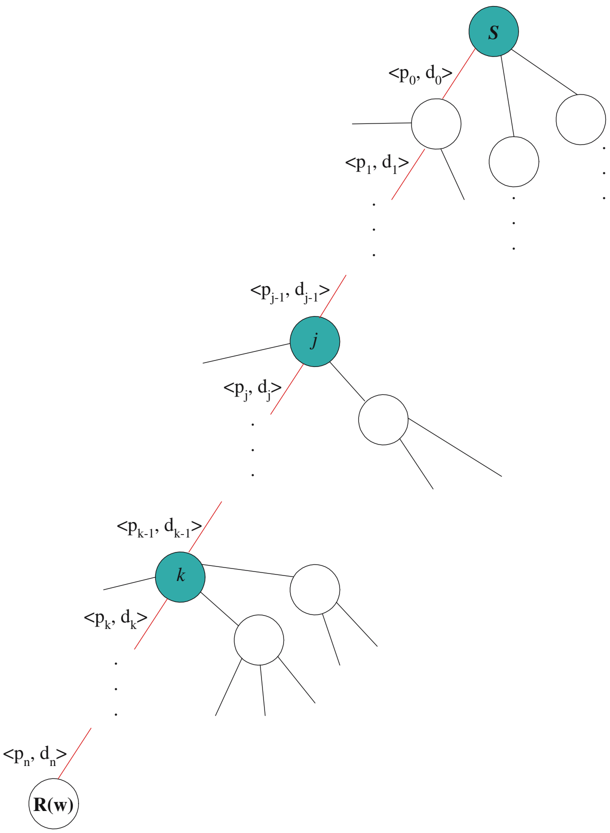


Fig. 2. Mean delivery delay model reflecting heterogeneity and locations of proxies in which $< p, d >$ on each link means $< lossrate, deliverydelay >$

Table 1. Notations for expected delivery model

Value	**Description**
$L_{a,b}$	Summation of delivery delays on all links between node a and node b
$E(N_{a,b})$	When node a is a proxy of node b, $E(N_{a,b})$ is the expected inter-arrival time of two consecutive packets in node b (their sequence numbers must be consecutive)
$\pi(a, b)$	Set of all proxies located on the path between node a and node b
$\theta(a, b)$	Number of elements in the set $\pi(a, b)$
$E(D_{S,R(w)})$	Mean delivery delay from sender node S and receiver node $R(w)$
$\psi(a, b)$	Set of all links between node a and node b

By eliminating $E(D_{S,R(w)})$ at the right side of (2), we obtain

$$E(D_{S,R(w)}) = L_{S,R(w)} + \frac{1 - \prod\limits_{i \in \psi(S,R(w))} (1-p_i)}{\prod\limits_{i \in \psi(S,R(w))} (1-p_i)} E(N_{S,R(w)}). \tag{3}$$

2) In case $\theta(S, R(w)) \geq 2$ and $node(j) \in \pi(S, R(w))$:

$$\begin{aligned} E(D_{S,R(w)}) = & \left[\prod_{i \in \psi(S,node(j))} (1-p_i) \right] \left(L_{S,node(j)} + E(D_{node(j),R(w)}) \right) \\ + & \left[1 - \prod_{i \in \psi(S,node(j))} (1-p_i) \right] \left(E(N_{S,node(j)}) + E(D_{S,R(w)}) \right) \end{aligned} \tag{4}$$

$$E(D_{S,R(w)}) = L_{S,node(j)} + \frac{1 - \prod\limits_{i \in \psi(S,node(j))} (1-p_i)}{\prod\limits_{i \in \psi(S,node(j))} (1-p_i)} E(N_{S,node(j)}) + E(D_{node(j),R(w)}) \tag{5}$$

By definition of $E(D_{S,R(w)})$, we obtain

$$L_{S,node(j)} + \frac{1 - \prod\limits_{i \in \psi(S,node(j))} (1-p_i)}{\prod\limits_{i \in \psi(S,node(j))} (1-p_i)} E(N_{S,node(j)}) \equiv E(D_{S,node(j)}), \tag{6}$$

so $E(D_{S,R(w)})$ can be written as

$$E(D_{S,R(w)}) = E(D_{S,node(j)}) + E(D_{node(j),R(w)}), \tag{7}$$

and if $node(k) \in \pi(node(j), R(w))$, we obtain

$$E(D_{node(j),R(w)}) = E(D_{node(j),node(k)}) + E(D_{node(k),R(w)}) \tag{8}$$

Thus by setting $\pi(S, R) = S, proxy_0, proxy_1, , proxy_z$, we obtain an recursive form as follows:

$$\begin{aligned} & E(D_{S,R}) = E(D_{S,proxy_0}) + E(D_{proxy_0,proxy_1}) + \cdots \\ & + E(D_{proxy_{z-1},proxy_z}) + E(D_{proxy_z,R}) \end{aligned} \tag{9}$$

So using this form, we can compute mean delivery delays of all receivers if a set of proxies is known. This mean delivery delay model is also used as an analytic model to evaluate our proxy placement method.

3 Optimal Placement of Proxies

As the network size grows, the ways of selecting proxies increase dramatically. So, finding optimal locations of proxies among these numerous ways becomes a combinatorial problem with large computational cost. In this paper, we deploy bottom-up dynamic programming formulation [15] to alleviate computational cost. When selecting m proxies in the network size of k nodes, the computation of minimizing the mean delivery delay can be performed using $O(k^2m)$ space.

3.1 The Dynamic Programming Formulation

First, a general HRM tree is converted into a binary tree by import of dummy nodes. Fig. 3 shows an example of binary tree conversion. For each node u having children u_L and u_R, for each θ, $0 \leq \theta \leq \theta_{max}$, where θ_{max} is the maximum number of nodes on that proxy can be placed, we can formulate the quantity $D(u, \theta, v)$ in four different cases with the additional notations in Table 2.

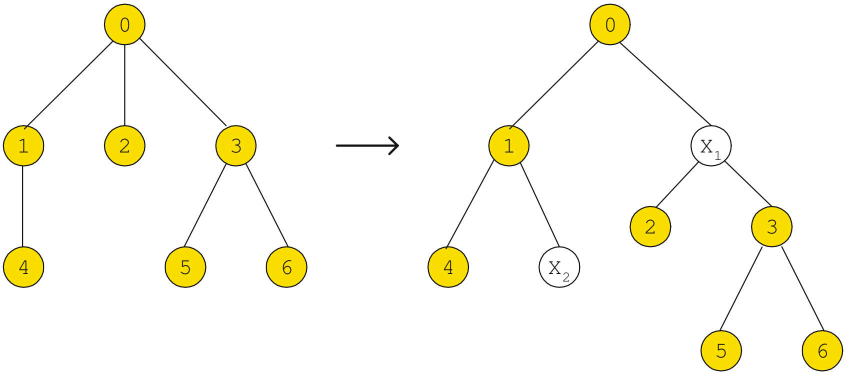


Fig. 3. Conversion of general tree into its binary form with import of dummy nodes X_1 and X_2

Table 2. Notations for the dynamic programming formulation

Value	Description
$\mathbf{T}_u$	Set of nodes placed in subtree rooted at node u
$\mathbf{R}_u$	Set of receivers placed in subtree rooted at node u
$n(\mathbf{R}_u)$	Size of $\mathbf{R}_u$
$D(u, \theta, v)$	Minimum total mean delivery delay of $\mathbf{R}_u$ from node v when θ proxies are placed in $\mathbf{T}_u$. Node v is a immediate proxy of node u.

1) If u is a leaf node:
We need not locate proxy at u. Thus θ is always 0.

$$D(u, \theta, v) = E(D_{v,u}) \tag{10}$$

2) If u is the sender,
Node u is a proxy itself by default, and it has no next-up proxy.

$$D(u, \theta, -) = \min_{0 \leq \theta' \leq \theta - 1} [D(u_L, \theta', u) + D(u_R, \theta - \theta' - 1, u)] \tag{11}$$

3) If u is an intermediate node, and we put a proxy at node u:

$$D(u, \theta, v) = \min_{0 \leq \theta' \leq \theta - 1} [D(u_L, \theta', u) + D(u_R, \theta - \theta' - 1, u) + E(D_{v,u}) \times n(\mathbf{R}_u)] \tag{12}$$

4) If u is an intermediate node, and no proxy is located at node u:

$$D(u, \theta, v) = \min_{0 \leq \theta' \leq \theta} [D(u_L, \theta', v) + D(u_R, \theta - \theta', v)] \tag{13}$$

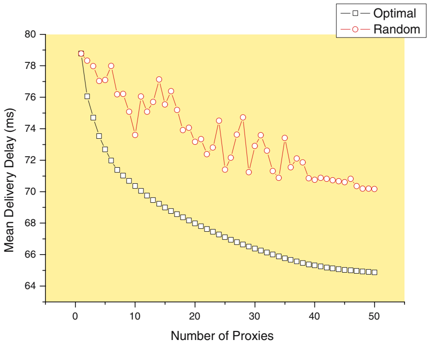
At each node, for each $0 \leq \theta' \leq \theta$ and its feasible next up-level proxy v, we have to check all possible partitions of θ' to the left and right subtrees. Therefore, if the size of a tree is n, the overall time complexity is bounded by $O(n^2\theta^2)$.

3.2 Configuration of Proxy Set

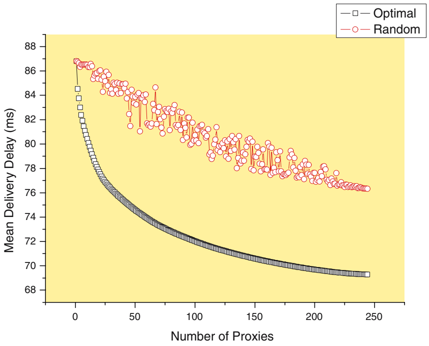
Our main purpose of this dynamic programming is to configure a set of proxies that minimize the total mean delivery delay of all receivers. A proxy set can be configured during the computation process of the matrix D. In this manner, if $n(\mathbf{T}_u) = n$, we need an additional space of $\theta^2 n$. However, if configuring a

```
Proxyset Proxy_Set(node u, size S) {

    if (S == 1) {
         Pu = Pu(S=1);
         return Pu;
    }

    if (SuR + SuL < S) // if u has a proxy
         Pu = Proxy_Set(uR, SuR) + Proxy_Set(uL, SuL) + {u};
    else // if u has no proxy
         Pu = Proxy_Set(uR, SuR) + Proxy_Set(uL, SuL);

    return Pu;

}
```

Fig. 4. Pseudo code for configuring proxy set $\mathbf{P}_u$. *Proxy_Set*() is called recursively.



(a) n(**T**)=200, n(**R**)=72, max[n(**P**)]=50



(b) n(**T**)=1000, n(**R**)=331, max[n(**P**)]=244

Fig. 5. Mean delivery delay of all receivers with respect to the number of proxies

proxy set is preceded by the completion of computing D, it can be done with an additional space of $3\theta n$. The configuring process can be implemented easily using the recurrence as in Fig. 4. Each node u, as $D(u, \theta, -)$ are computed, stores $\theta_{u_R} = n(\mathbf{P}_{u_R})$, $\theta_{u_L} = n(\mathbf{P}_{u_L})$, and $\mathbf{P}_{u(\theta=1)}$.

4 Numerical Evaluations

The performance of our proposal and the random placement method are compared. Simulation topologies are generated by ToGenD[20]. Delivery delays $(10 \sim 40\text{ms})$ and loss rates of all links $(0.0001 \sim 0.1)$ are assigned heterogeneously. The multicast delivery tree is constituted using Dijikstra Algorithm [21] in order to minimize the total delivery delay of all source-receiver pairs. We assume that inter-packet gap is 25ms. Fig 5 shows the arithmetic average of mean delivery delays of all receivers with respect to the number of proxies. As expected, our proposal for the placement of proxies yields a lower mean delivery delay than the random placement method regardless of the number of proxies. The computation is completed in a few tens of seconds on our Pentium IV 3.0GHz machine.

5 Conclusions

In this paper, we propose a scheme to configure repair proxies that can minimize mean delivery delay in heterogeneous network environments if proxy size is limited to some value. We describe mean delivery delay model to reflect heterogeneity and locations of proxies, and apply dynamic programming in order to configure an optimal proxy set in reasonable time. Also, we use the statistical method to determine the time of proxy relocation. Through numerical evaluations, the performance of our proposal is compared with that of a method that places the proxies randomly in the network.

Acknowledgement

This research was supported by the MIC(Ministry of Information and Communication), Korea, under the ITRC(Information Technology Research Center) support program supervised by the IITA(Institute of Infor-mation Technology Assessment) and a Korea University Grant.

References

1. Y. Chawathe, S. McCanne, and E. Brewer : RMX: Reliable Multicast in Heterogeneous Networks, IEEE INFOCOM '00
2. M. Hoffman, : A Generic Concept of Large-Scale Multicast, International Zurich Seminar on Digital Communications, Springer, Feb. 1996, vol. LNCS 1044, pp. 95-106.
3. H. Holbrook, S. Singhal, and D. Cheriton, : Log-Based Receiver-Reliable Multicast for Distributed Interactive Simulation, ACM SIGCOMM '95.
4. S. Floyd, V. Jacobson, C.-G. Liu, S. McCanne, and L. Zhang, : A Reliable Multicast Framework for Light-weight Sessions and Application Level Framing, IEEE/ACM Trans. Networking, vol. 5, pp. 784-803, Dec. 1997.
5. J. Lin, and S.Paul, : RMTP: A Reliable Multicast Transport Protocol, IEEE INFOCOM '98.
6. J. Nonnenmacher, M. Lacher, M. Jung, E. Biersack, and G. Carle, : How bad is Reliable Multicast without Local Recovery?, IEEE INFOCOM '98.
7. C. Maihofer, and K. Rothermel, : A Delay Analysis of Tree-based Reliable Multicast Proto-cols, Computer Communications and Networks 2001.
8. A. Markpoulou, and F. Tobagi, : Hierarchical Reliable Multicast: performance analysis and placement of proxies, ACM SIGCOMM '00.
9. H. Lin, and K. Yang, : Placement of Repair Servers to Support Server-based Reliable Multicast, ICC 2001.
10. Z. Wan, M. Kadoch, and A. Elhakeem, : Performance Evaluation of Tree-based Reliable Multicast, ICCCN 2003.

11. O. Daescu, R. Jothi, B. Raghavachari, and K. Sarac, : Optimal Placement of NAK-Suppressing Agents for Reliable Multicast: A Partial Deployment Case, ACM SAC 2004.
12. S. Ratnasamy, and S. McCanne, : Scaling End-to-End Multicast Transports with a Topo-logically-Sensitive Group Formation Protocol, ICNP '99.
13. B. Li, F. Chen, and L. Yin, : Server Replication and Its Placement for Reliable Multicast, ICCCN 2000.
14. B. Li, M. Golin, G. Italiano, X. Deng, and K. Sohraby, : On the Optimal Placement of Web Proxies in the Internet, INFOCOM '99.
15. A. Tamir, : An O(pn2) algorithm for the p-median and related problems on tree graphs, Oper. Res. Lett. vol. 19, pp. 59-64, 1996.
16. B. Levine, S. Paul, and J. Garcia-Luna-Aceves, : Organizing Multicast Receivers Deter-ministically by Packet-Loss Correlation, ACM International Conference Multimedia, 1998.
17. S. Casner, and A. Thyagarajan, : mtrace(8): Tool to Print Multicast Path form a Source to a Receiver, UNIX manual command.
18. T. Speakman, J. Crowcroft, J. Gemmell, D. Farinacci, S. Lin, D. Leshchiner, M. Luby, T. Montgomery, L. Rizzo, A. Tweedly, N. Bhaskar, R. Edmonstone, R. Sumanasekera, and L. Vicisano, : PGM Reliable Transport Protocol Specification, IETF, RFC 3208, Dec. 2001.
19. D. Li, and D.R. Cheriton, : OTERS (On-Tree Efficient Recovery using Subcasting): A Reliable Multicast Protocol, ICNP '98.
20. O. Mokryn, : ToGenD - A Notre-Dame Based Topology Generator, http://www.eng.tau.ac.il/ osnaty/togend.html.
21. S. Deryfus, : An Appraisal of Some Shortest-Path Algorithms, Operations Research, 17(3), 1969.
22. S. Pejhan, M. Schwartz, and D. Anastassiou, : Error Control Using Retransmission Schemes in Multicast Transport Protocols for Real Time Media, IEEE/ACM Transactions on Networking, vol. 4, no. 3, pp. 413-427, 1996.
23. T. Nunome, and S. Tasaka, : An Application-level QoS Comparisons of Inter-destination Synchronization Schemes for Continuous Media Multicasting, IEEE GLOBECOM '03.
24. M.Allman, D. Glover, and L. Sanchez, : Enhancing TCP over Satellite Channels Using Standard Mechanisms, RFC 2488, 1999.
25. T. Turletti, : The INRIA Videoconferencing System (IVS), ConneXions-The Interopera-bility Report Journal, vol. 8, no. 10, pp. 20-24, 1994.
26. R. Caceres, N. G. Duffield, J. Horowitz, and D. F. Towsley, : Multicast-Based Inference of Network-Internal Loss Characteristics, IEEE Trans. on Information Theory, vol. 45, no. 7, 1999.
27. S. Byun, and C. Yoo, : Placement of Repair Proxies to Improve Inter-receiver Delivery Delay Fairness in Hierarchical Reliable Multicast Networks", AICT 2005.