ADAPTIVE CONTENTS DELIVERY PROTOCOL FOR VARIABLE-SPEED MULTICAST TREES

Hideaki TETSUHASHI[†], *Takumi MIYOSHI*[††,†], *and Yoshiaki TANAKA*[†,†††]

[†] Global Information and Telecommunication Institute, Waseda University
1-3-10 Nishi-Waseda, Shinjuku-ku, Tokyo, 169–0051 Japan

[††] Department of Electronic Information Systems, Shibaura Institute of Technology
307 Fukasaku, Minuma-ku, Saitama-shi, Saitama, 337–8570 Japan

[†††] Advanced Research Institute for Science and Engineering, Waseda University
17 Kikuicho, Shinjuku-ku, Tokyo, 162–0044 Japan

ABSTRACT

Multicast communication is a remarkable technology that can substantially decrease network traffic and can save network resources and transmission costs. Yet exactly for this reason, users of a highly heterogenous network cannot use multicast. This problem in present multicast is solved by setting the transmission speed to the lowest among the available capacities of links in the multicast tree. Some side effects of this is the increase of data transmission time, and, therefore efficiency and real-time-based performance suffer. This paper proposes an adaptive content delivery system over multicast trees and discusses its implementation. Considering some limitations of real network, this paper proposes the use of link capacity measurement. Adaptiveness is based on the measurement and can trigger temporary storage of data in case the discrepancy is found among downstream links originating from the same node.

1. INTRODUCTION

With the rapid progress in development of computer networks, contents delivery services such as electronic publishing, software distribution, and multimedia delivery are highly demanded. Some of push-type information delivery services are already provided over the Internet. The capacity increase and price reduction of storage devices in personal computers, and broadband access lines of home users, enable to download large-size contents to store on hard-disks for later use. In the near future, the services provided over the network will be more diversified, the broadband portion of the network will grow and services will be provided over large distances due to increased scalability.

Particularly, broadband multimedia contents delivery services will increase traffic load greatly. Almost all of these services are transmitted from a server to many client receivers. Multicasting is one of the remarkable technologies that can provide such point-to-multipoint communication effectively [1, 2]. The advantage of multicast communication is that only one packet is transmitted to an intermediate node, where the information is copied and then transmitted to two or more destinations. The multicast connection can substantially decrease the traffic in the network and thus save network resources and transmission costs. If multicasting is used for contents delivery systems, however, the transmission speed must be set to the slowest one among the available capacities of links on multicast tree for all client terminals to receive the contents simultaneously. This results that in reality, multicast users receive data incompletely because of the packet loss at routers in the heterogeneous networks like the Internet. This problem often results in termination of data transfers when discrepancy among links' speed gets too large [3].

The more heterogeneous the network becomes, the more problems it has with using multicast. One good example would be the Internet, where the multicast is not widely used exactly due to the reason of extremely high-level heterogeneity. Multicast is, nevertheless, very commonly used in the MBone networks, the high-speed networks where the router and link parameters in the network topology are fairly similar and, therefore, multicast offers higher effectiveness in network operation.

This paper proposes to introduce changes into multicast routing protocol so that multicast could be used in heterogeneous environment by adaptive contents delivery system that does not have stringent real-time constraints. The proposed changes allow the multicast protocol to select either multicasting or store-and-forward transmission mode based on downstream line parameters. From the viewpoint of the data transmission time, computer simulation proves the efficiency of this system in earlier work in this area [4] .

The rest of this paper is organized as follows: Section 2 explains the proposed contents delivery system that uses both multicasting and buffering modes. Section 3 discusses the implementation of the proposed system. Section 4 contains conclusion and future work.

0-7803-8560-8/04/$20.00©2004IEEE

2. ADAPTIVE CONTENTS DELIVERY SYSTEM USING MULTICASTING AND BUFFERING

2.1. Principle

The multicast mechanism can substantially decrease traffic in the network and save network resources. Presently, when a multicast is used for contents delivery, the transmission speed must be set to the slowest one among the available capacities of links on multicast tree for all users to receive the contents simultaneously. Figure 1 (a) represents the case when a multicast connection is used for the contents delivery. The transmission speeds of all nodes are set to the value of the slowlest link in the tree to make sure that all multicast nodes in the tree are properly synchronized. Although such a configuration is very rare in MBone or high-speed network cores, it becomes a heavy problem in heterogeneous networks, and, therefore, in the Internet.

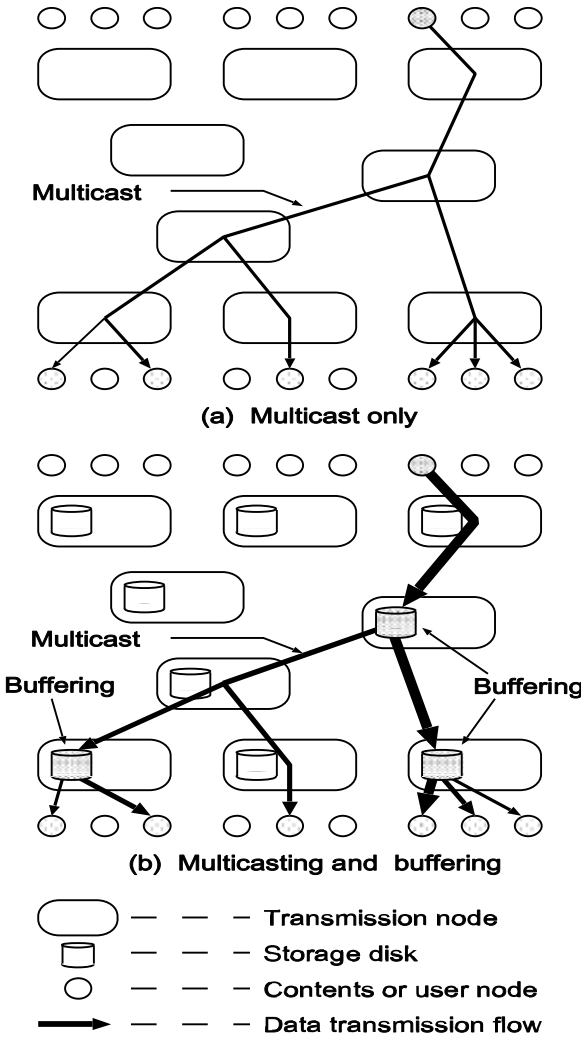

Fig. 1. Contents delivery systems.

The primary advantage of the proposed system is that high-speed users in the network are not forced to low-level transmission speeds to synchronize with slow nodes. Figure 1 (b) represents the general structure of multicast trees in the proposed system. On the contrary, high-speed users will be subjects to transmission at the speeds they are capable of, or, to be more correct, as the maximum speed of the immediate parent. This means that nodes that lie on the high-speed path from the source will obtain the data at the speed of the path. Slow links will put a limit on the speeds of the downstream tree. However, in the proposed system, they have no influence to the speed of the total system, or high-speed links.

The secondary advantage of the system is even more important when applied to multimedia contents. The proposed system attempts to decrease packet loss whenever possible. In the multicast world today, packet loss is one of the most important problems, given the fact that there is never a state of the network when the load in it is smoothly distributed. There are always overutilized and underutilized links, which cause data loss somewhere down the multicast tree.

2.2. Architecture

In the architecture of the proposed contents delivery system, routers or switching systems (to be called *transmission nodes*) have large-size storage disks in them as shown in Figure 1 (b). This design enables the transmission nodes with storage disks to serve as contents delivery servers for further re-transmissions: The transmission nodes can thus terminate the multicast connections of the data transmissions, and can resend them again to the child nodes. We believe that it should be a selectable option for network providers as to whether nodes would have storage disks. This paper assumes the scenario when all transmission nodes have the storage devices.

Detailed architecture of the proposed system is presented in Figure 2. It contains both the representation of currently used PIM_SM protocols, which uses IGMP protocol to create multicast trees by exchanging "hello" messages among immediate members to multicast.

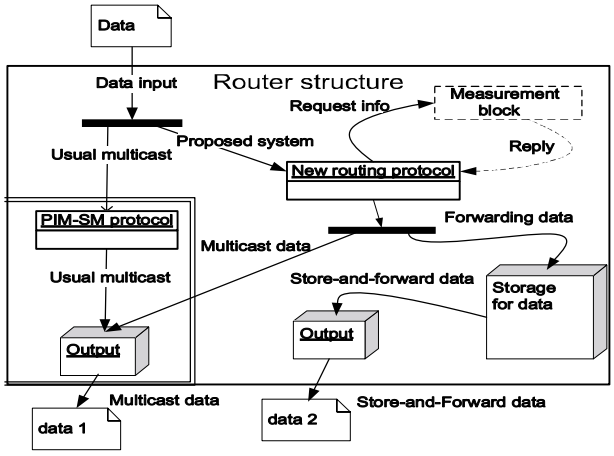

Fig. 2. Architecture of proposed multicast system.

Separately from currently existing PIM_SM protocol module, Figure 2 also contains the proposed sys-

tem. PIM_SM and the proposed system have common functionality. It is the part of serving multicast connections with no or very little discrepancies in down link speeds. In this case, the proposed system behaves the same as conventional multicast, i.e. the received packet is transmitted without storing. The operation is different with the discrepancies in down link speeds getting above the threshold. In that case, the data is parked as store-and-forward data and is stored in the storage device for later retransmission in one of the future connections.

The decision on whether to send the data in the multicast or store-and-forward mode is made based on the information obtained from measurement block. This block is responsible for performing measurements on the links to immediate neighbors in the multicast tree. Measurements can be performed once to define the state and make all future decisions based on this initial data. They can also be performed regularly with the purpose of refreshing the state of neighboring links to be able to make more reasonable decisions about transmission mode.

Below is the list of activities performed by a single multicast node using the proposed system:

- To accept transmission requests,
- To decide the data delivery routes (routing),
- To control the transport connection between multicast-enabled nodes,
- To multicast the data, and
- To store, copy, and re-transmit the data.

Figure 3 contains the sequence diagram of the proposed system. The figure represents the actions that are performed among parent node and two child nodes in the multicast tree. The parent node makes the decision about the transmission mode, and therefore, it also contains storage device to display communication between the storage device and decision-making module in a single router.

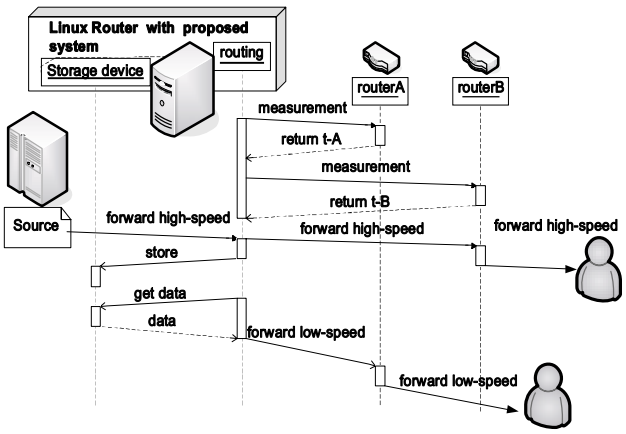

Fig. 3. Algorithm sequence diagram of proposed system.

2.3. To Select Multicasting or Buffering

When the node chooses multicasting or buffering modes, it is based on the information received from measurement block. If the differences among these available bandwidth values are small, the multicast transmission will work effectively. If the differences are large, on the contrary, it will be better to store the data and to transfer it using new other connections. The calculation is performed based on the simple Eq. (1).

Let the available bandwidth value on $link_i$ be B_i, when B denotes the bandwidth of the uplink from the present node and is the largest possible speed at which data can possibly be transmitted. The selection index S is defined as follows:

$$S = \frac{\max\limits_{i \in E} B_i - \min\limits_{i \in E} B_i}{B} \quad (1)$$

If the selection index S is within the following range, the buffering mode is selected.

$$L_{\text{low}} \leq S \leq L_{\text{high}}, \quad (2)$$

where lower L_{low} and higher L_{high} limits define the working space of the system. Too low values of L_{low} would result in very low effectiveness, as the data that could be sent in multicast mode would be stored and will result in delays for storing and reading back from the storage device. Too high values, on the other hand, will cause the storage device to use too much space, and may cause memory overflows. The upper limit plays the role of the admission mechanism in such a way, that the storage and transmission of the stored data will be suspended in case the memory limitation are violated.

3. IMPLEMENTATION CONSIDERATION

3.1. Router-Based Implementation Constraints

Presently, routers have certain resource constrains that are caused by their fairly simple structure as well as by the small memory size. IOS(Cisco Internetworking Operating System) in Cisco routers for example amount only to 2MByte of total size [8]. Besides, there is an approach that advocates keeping the simplicity of routers as it is, since routers perform large volume of work and require to be simple. That is why storage disks are never used in current switching equipments, and instead, flash memory devices are used.

The proposed system, in fact, does not require large memory size to operate. For this reason, the upper limit on storage probability in Eq. (2), which has direct relation to the storage size. As mentioned above, to keep connection in case of very large difference among available bandwidths of downstream links, the nodes with very low speeds will be cut from the multicast tree. The storage mode will keep its operation within the working range between the lower and higher limitations imposed on T_h.

3.2. Available Bandwidth Measurement

We can find many tools in today's Internet that offer measurements on one-hop path, as well as on multi-hop path. The simplicity and, therefore, the implementation validity is defined by the fact that the only measurement type required by the system is one-hop available bandwidth measurement between a node and the neighbor node on the multicast tree.

Such measurements are already implemented in some routers manufactured by Cisco as well as other manufacturers such as Juniper and others. The purpose of the measurement module in such routers is to obtain the information about link condition between the routers and other routers in the network by probing the link status with specifically created packets with specific probing pattern. The proposed system could make use of such module directly, as no special measurements are required, and the information from measurement module can be used as it is.

Apart from router-based measurements, separated measurements are also possible by using special probes. The probes can be set at certain points in the network where they will be able to probe effectively and define the available bandwidth over many paths in the network. As a few routers can be on the same path, they can share the path conditions provided by the probe. In this case, the measurement would be performed separately, and the status information would be uploaded to the router by using SNMP protocol.

3.3. Multimedia Contents Delivery Patterns

It is true that some multimedia contents in the Internet have real-time requirements. Some examples may be video and audio streaming, with users viewing the data as it is being received through the network. Some others, though, may store the stream locally before viewing it. However, both categories of users would not like the loss of the data due to the constraints in the conventional multicast.

The proposed system helps both real-time and non-real-time categories of users by offering high-speed transmission to real-time users on high-links, and lower, yet still reliable transmissions to users in low-speed areas of the network.

Another reason for high packet loss in the conventional multicast is congestion in the network that causes sudden changes in the available bandwidth on certain links in the network. That finally results in the packet loss, and, therefore, severed transmission from that point in the tree. The proposed system solves this problem, as the changing point which are grasped in the available bandwidth by the measurements. Thus, congestion conditions with the proposed system would result not in the packet loss, but in the increase of the storage memory size, where the data that cannot be transmitted at the time will be stored for later transmissions.

4. CONCLUSION

This paper discussed the problems that the conventional multicast suffers in the present days. Those problems are due to the lack of the regard to the possible differences in link speeds among the downstream nodes in multicast tree built over heterogeneous network, good example of which is the Internet itself. This problem with multicast causes its very limited use mainly in MBone-type networks, where the multicast core consists of the similar routers and has smoothly distributed load. This, however, is not the case with the Internet, where not only the load is unevenly distributed, the link speeds can vary greatly regardless of territorial nearness. The paper proposed the solution by introducing additional multicast mode—store-and-forward. This mode is turned on when the discrepancy among link speeds from the same intermediate node is very large. In this case, the data is temporarily stored to be retransmitted in later sessions with the nodes on the slow links. Even considering the fact that we impose memory size limitations, the effectiveness of the system should not suffer, unless the available bandwidth discrepancy among downstream links is extremely large.

In the future work, we are planning the implementation of the proposed system based on Linux router. Finally, when the implementation is finished and tested, we are planning to offer the software product for the use by Internet community.

5. REFERENCES

[1] S. Deering, "Host extensions for IP multicasting," *RFC 1112*, IETF, Aug. 1989.

[2] C. Diot, W. Dabbous and J. Crowsroft, " Multipoint communication: a survey of protocols, functions, and mechanisms, "*IEEE J. Select. Areas Commun*, vol. 15, no. 3, pp. 277–290, April 1997.

[3] J. R. Cooperstock and S. Kotsopoulos, "Why use a fishing line when you have a net? An adaptive multicast data distribution protocol, " *1996 USENIX Tech.Conf.*, pp. 343–352, Jan. 1996.

[4] T. Maeda, Y. Tanaka and H. Tominaga, "Adaptive contents delivery system using multicast and buffering(in Japanese), "*Tech. Rep. of IEICE*, SSE99-11, April 1999.

[5] B. M. Waxman, " Routing of multipoint connection, "*IEEE J. Select. Areas Commun.*, vol. 6, pp. 1617–1622, Dec. 1988.

[6] W. E. Leland, M.S. Taqqu, W. Willinger and D. V. Wilson, "On the self-similar nature of ethernet traffic, "*ACM SIGCOMM'93*, pp. 183–193, Sept. 1993.

[7] T. Miyoshi and Y. Tanaka, " Adaptive contents delivery system using multicasting and buffering in best effort networks, "*4th Asia-Pacific Symposium on Information and Telecommunication Technologies (APSITT 2001)*, pp. 101-105, Nov. 2001.

[8] Cisco Systems, Inc. http://www.cisco.com/, May 2004.