****

ATM Forum Technical Committee ATMF 95-1438

****

Title: Reliable Multicast Service Needs Cell-Level FEC Scheme

****

Abstract:

A number of protocol architectures for the provision of scalable reliable multicast services over the Internet have been proposed. These protocol architectures could also be applied for ATM networks. However, since ATM networks are based on switches rather than on broadcast-based subnetworks (e.g., Ethernet and FDDI), some architectural modification is needed, while a modification the end-station's protocol is not forced.

As shown in this memo, in order to provide efficient and scalable reliable multicast over an ATM cloud, we need an AAL-level FEC scheme to provide sufficiently small packet error ratio.

****

Source :

Hiroshi Esaki (Toshiba Corp.)
801 Schapiro Research Building,
c/o CTR, Columbia University,
530 West, 120th Street,
New York, New York, 10027, USA.
Tel:(212)-854-2365, Fax:(212)-316-9068
E-mail: hiroshi@ctr.columbia.edu

Georg Carle
Institute of Telematics, University of Karlsruhe
Zirkel 2
D-76128 Karlsruhe, Germany
Tel:+49-721-608-4027, Fax:+49-721-38809
E-mail: carle@telematik.informatik.uni-karlsruhe.de

Tim Dwight
MCI Telecommunications Corporation,
901 International Parkway
Richardson, Taxas, 75081, USA.
Tel:(214)-498-1484, Fax:(214)-498-1300
E-mail:0006078043@mcimail.com

****

Date : December 11-15, 1995

****

Distribution: Plenary (FEC-BOF)

****

****

1 Scalable Reliable Multicast Architecture

The current multicast applications (e.g., designed for the M-Bone) do not require error-free data delivery, since the applications are error tolerant (e.g., voice and video). However, the needs for reliable multicast service have been growing and are expected to be of high significance soon [Floy95, Holb95, Brut95].
The design principle of a large scale reliable multicast service architecture [Floy95] should be:

1. avoiding packet preservation within the network;
2. avoiding protocol state implosion;
3. avoiding control message implosion;
4. avoiding the re-transmission message implosion;
5. allowing unreliable behavior of down-stream nodes.

Though there should be the other design principles, as discussed in [Floy95], the above requirements are the important points.
In this memo, three possible architectures are discussed.
All architectures satisfy the above requirements.

1.1 wb, distributed whiteboard [Jaco92, McCa92, Floy95]

Each member of the group is individually responsible for detecting loss and requesting retransmission (i.e., negative ACK approach rather than positive ACK approach). In order to implement the negative ACK approach, the sequence number of the message with the highest sequence number is periodically announced. The control message contains timestamps that are used to estimate the distance (time) from each member. Since the protocol to provide reliable data transmission is soft-state and receiver oriented, rather than pt-pt-based sender oriented, the implosion of protocol state will be avoided, and unreliable behavior of receivers will be accepted, i.e., the requirements of (2) and (5) are met.
When receiver(s) detect missing data, they wait for a random time determined by their distance from the original data source, then send a recovery request. The recovery request messages and retransmission messages are always sent as a multicast to the whole group. Therefore, the network does not need to preserve any multicast packet to recover erroneous messages (i.e., requirement (1)).
Hosts that are missing data, but did not yet experience the timeout of a recovery request message sending timer, may receive the recovery request message issued by another host and then reset their recovery request message sending timer, suppressing to send their recovery message. This operation will avoid the implosion of control message, i.e., requirement (3).
Any host that has a copy of the requested recovery message will prepare to answer the request using a multicast channel. It will set a repair timer to a random value depending on its distance from the sender of the request message and multicast the recovery message when the timer is expired. Other hosts that did receive the message correctly and scheduled the transmission of this recovery message will cancel the scheduled event when they receive the recovery message from the multicast channel. This operation will avoid the implosion of re-transmission messages (i.e., requirement (4)).

1.2 Log-Based Receiver-Reliable Multicast (LBRM) [Holb95]

This concept is based on designated servers (i.e., distributed log-servers) to perform packet retransmission. These log-server(s) will not co-exist with routers, but exist outside the network. Therefore, we could not say this approach satisfies requirement (1). Each member of the group knows the individual address of its corresponding log-server(s). Among the log-servers and source nodes, the conventional positive ACK
approach (like TCP) is applied.

Each member of the group is individually responsible for detecting loss and requesting retransmission (i.e., negative ACK approach rather than positive ACK approach). In order to implement a negative ACK approach, heartbeat packets or keep-alive packets are periodically transmitted.
Since the protocol to provide reliable data transmission is soft-state and receiver oriented, rather than pt-pt-based sender oriented, the implosion of protocol state will be avoided and the unreliable behavior of receivers will be accepted, i.e., requirements of (2) and (5).
When receiver(s) detect missing data, they send a recovery request message to the corresponding log-server. The retransmission of missing data is performed based on point-to-point unicast channels. This retransmission scheme does not cause the implosion of control messages and retransmission messages (i.e., requirements (3) and (4)). However, when the number of nodes missing a given packet is large, the log-server must transfer many packets. A local multicast channel could be used for packet retransmission. However, in this case, a local multicast channel, as well as global multicast channel would be required.

1.3 Protocol state maintenance at dedicated intermediate nodes

LBRM discussed above could be modified in a way that a log-server does not preserve packets nor retransmits missing messages. When the dedicated intermediate nodes only maintain protocol state of the down-stream nodes, the network does not need to preserve packets (i.e., requirement (1)). It would be possible that only the sender process preserves packets and performs packet retransmission. When a dedicated intermediate node merges protocol state indicated to the up-stream nodes, the implosion of control messages can be avoided (i.e.,requirement (3)). Also, when the protocol state management at each dedicated intermediate nodes and at sender processes is based on soft-state (i.e., negative ACK), the implosion of protocol state information can be avoided (i.e., requirement (2)). Since the packet retransmission for the missing data is performed by the sender process, a multicast channel can be used, and we can avoid the implosion of retransmission messages (i.e., requirement (4)). Since the protocol state management can use a negative ACK (soft-state) scheme, and membership management for receivers can be soft-state and in a distributed fashion, unreliable receivers can be allowed (i.e.,requirement (5)).

2 Issues of reliable multicast in the ATM network

As discussed above, we could provide a scalable reliable multicast. However, when we provide reliable multicast over the ATM network, there are additional issues to be considered:

a) packets are fragmented into multiple cells. (A packet may be fragmented into several hundred cells.) For example, with a default MTU of 9,180 Bytes for IP over ATM as defined by RFC1626, a packet of maximum size corresponds to 192 cells.
b) ATM is a switch-based platform and not a broadcast-based platform. After passing a branching point (i.e., a copying point), each copied cell is transferred individually and independently.

Let assume M as the number of cells in a packet, N as the number of receivers, and m as the number of receivers in a given broadcast-based platform. The diameter (d) of the multicast tree with the broadcast-based platform will be approximately in the order of log_{m} N. d would be usually less than a few decades.
With a network consisting of broadcast-based subnetworks, the packet error probability for a sender process will degrade according to the order of d, which is in the order of log_{m} N. On the other hand, in a switch-based platform (like ATM), the packet error probability for a sender process will degrade according to the order of N*M*d. This means that in the switch-based ATM network, the packet error probability will severely degrade with an increase of N and M. Here, the packet error probability for a sender process means the probability that at least one receiver requires the retransmission of a packet.

This can be illustrated by the following examples:

```
 I.   CLR=10**-6, Datagram error probability=10**-6;
      Packet error probability for sender
        + broadcast-based platform ; 10**-5 (=10**-6 x 10)
        + ATM platform             ; 10**-2 (=10**-6 x 10 x 100 x 100)
 II.  CLR=10**-8, Datagram error probability=10**-8;
      Packet error probability for sender
        + broadcast-based platform ; 10**-7 (=10**-8 x 10)
        + ATM platform             ; 10**-4 (=10**-8 x 10 x 100 x 100)
 III. M=100, CLR=10**-8, Datagram error probability=10**-8;
      Packet error probability for sender
        + broadcast-based platform ; 10**-7 (=10**-8 x 10)
        + ATM platform             ; 10**-3 (=10**-8 x 10 x 1,000 x 100)
```

3 Relation with packet level FEC

When the packet error probability for the sender process is not acceptable, there are two possible solutions that apply FEC. One is the improvement of datalink level transmission quality (e.g., by using an AAL-level FEC in ATM), and the other is using packet level forward error correction (e.g., packet level FEC). Both solutions can be applied simultaneously.

As discussed in the previous section, ATM networks may show a particularily high scalability problem. Since, in general, an application does not know when data travels via an ATM channel, FEC needs to be applied in all cases in which the existence of ATM channels with significant cell loss would lead to an unacceptable high packet loss rate.

However, when the packet error probability is improved by an AAL level FEC scheme, the application does not need to initiate additional FEC in the case where the data flow travels over an ATM network section.

We should not force the use of a packet level FEC scheme within end-stations just because ATM is introduced in the Internet, due to significant processing and implementation costs of a packet level FEC scheme.

Of course, with large scale multicast, the packet error quality may be expected to be insufficient. In this case, the application may use packet-level FEC. Even in this case, additional AAL-level FEC can be applied, in order to solve a particular quality degradation problem of an ATM subnetwork.

4 Conclusion

For a large scale reliable multicast service over ATM networks, we need an AAL-level FEC scheme to obtain a sufficient probability of packet delivery for the sender process. Though, there are some scalable reliable multicast architectures, ATM networks may have a problem without the AAL-level FEC scheme, due to packet fragmentation into cells and due to switch-based data transmission.

In order to efficiently provide a large scale reliable multicast service over an ATM network, we have to use the AAL-level FEC scheme in all cases with significant cell loss rates.

5 Reference

[Brut95] D.Brutzman, M.Macedinia, M.Zyda, „Internetwork Infrastructure Requirements for Virtual Environments“, Virtual Reality Modeling Language (VRML) Symposium, December 1995.

[Floy95] S.Floyd, V.Jacobson, S.McCanne, C.Kiu, L.Zhang, „A Reliable Multicast Framework for Light-weight Session and Application Level Framing“, SIGCOMM95, September, 1995.

[Holb95] H.Holbrook, S.Singhal, D.Cheriton, „Log-Based Receiver-Reliable Multicast for Distributed Interactive Simulation“, SIGCOMM95, September, 1995.

[Jaco92] V.Jacobson, „Multimedia Conferencing on the Internet“, tutorial 4, SIGCOMM94, August, 1994.

[McCa92] S.McCanne, „A Distributed Whiteboard for Network Conferencing“, UC Berkeley CS 268 Computer Networks term project, May, 1992.

[RFC1626] R.Atkinson: „Default IP MTU for use over ATM AAL5“, RFC1626, May, 1994.