USING MULTICAST IN THE
GLOBAL COMMUNICATIONS INFRASTRUCTURE
FOR GROUP COMMUNICATION

Deborah A. Agarwal, Information and Computing Sciences Division,
Ernest Orlando Lawrence Berkeley National Laboratory

ABSTRACT

International Monitoring System (IMS) stations and the International Data Centre (IDC) of the Preparatory Commission for the Comprehensive Nuclear-Test-Ban Treaty Organization generate data and products that must be transmitted to one or more receivers. The application protocols used to transmit the IMS data and IDC products will be CD-x and IMS-x and the World Wide Web (WWW). These protocols use existing Internet applications and Internet protocols to send their data. The primary Internet applications in use are electronic mail (e-mail) and the file transfer protocol (ftp). The primary Internet communication protocol in use is the Transmission Control Protocol (TCP), which provides reliable delivery to the receiver. These Internet applications and protocol provide *unicast* (point-to-point) communication. A message sent using unicast has a single recipient; any message intended for more than one recipient must be sent to each recipient individually. In the current design, the IDC and the National Data Centres (NDC's) provide data forwarding to the appropriate receivers. The overhead associated with using unicast to transmit messages to multiple receivers either directly or through a forwarder increases linearly with the number of receivers. In addition, using a forwarding site introduces possible delays and possible points of failure in the path to the receivers.

Reliable multicast provides communication services similar to TCP but for a group of receivers. The reliable multicast protocol provides group membership services and message delivery ordering. If an IMS station were to send its data using reliable multicast instead of unicast, only sites that are members of the multicast group would receive the data at approximately the same time. This might provide an efficient means of disseminating station data or IDC data products to all receivers and eliminate or greatly reduce the need for data forwarding.

Several commercial and research reliable multicast protocols exist for the Internet. Each of these protocols is designed to serve a specific community of users and applications. The author has undertaken a study to determine if reliable multicasting is appropriate for use in the Global Communication Infrastructure (GCI).

Key Words: communications, multicast, reliable

OBJECTIVE

Reliable multicast is a basic capability that has been proposed for use as a communication mechanism in the Global Communication Infrastructure (GCI). A study has been undertaken by the author to determine whether reliable multicast can be used in the Comprehensive Nuclear Test-Ban Treaty (CTBT) monitoring network. There are two aspects to providing reliable multicast in the GCI network. The reliable multicast capability would need to be available in the network and the International Data Center (IDC) application protocols would need to be converted to make use of the reliable multicast capability before any benefits would be realized.

The Global Communication Infrastructure (GCI)

The International Monitoring System (IMS) under the Comprehensive Nuclear Test-Ban Treaty (CTBT) comprises 321 IMS stations and 16 radionuclide laboratories worldwide. The International Data Center (IDC) in Vienna will collect, store, and process the monitoring data from these stations. The IDC is responsible for distribution of the data and the data products to the National Data Center (NDC) of each interested Member State. The closed and secure data network under construction to carry this data is called the Global Communications Infrastructure (GCI). The GCI will provide data communication links for the IMS stations, the NDC's and the IDC. When the network is in full operation, it is expected to carry approximately 10 gigabytes per day.

Topologically, the GCI is a private network composed of frame relay and satellite links configured as a two-level tree rooted at the IDC. The major network nodes are the IDC in Vienna and four (or more) satellite hubs in Germany, Italy (2) and the United States. High-speed terrestrial Frame Relay private virtual circuits (PVC) with integrated services digital network (ISDN) backups connect the IDC to each of the hubs. There are four space satellites; each VSAT hub provides communication for a different satellite. Typically, an IMS station is connected by a VSAT (Very Small Aperture Terminal earth station) to one of the satellites and thus one of the four hubs. The NDC's are connected to the GCI via VSAT, frame relay, or the Internet. There are also three methods of connecting IMS stations to the GCI. These methods are called respectively the Basic Topology, Partitioned Subnetwork and Independent Subnetwork. In the Basic Topology, the monitoring data originating at an IMS station uses a GCI VSAT satellite link to a VSAT hub, and from there, it travels a GCI Frame Relay PVC to the IDC. In a Partitioned Subnetwork, the physical connection of the GCI at the IMS stations is the same as in the Basic Topology except that data from the IMS stations is routed through the NDC before going to the IDC. In an Independent Subnetwork, the GCI does not have a direct connection to the IMS stations. The network between the IMS stations and the NDC is installed and operated by the Member State. Data from the IMS stations in an Independent Subnetwork is routed through the NDC before going to the IDC. The GCI is expected to meet stringent performance requirements. For example, the one-way delay target is 5 seconds or less for at least 99.5% of the time. The virtual circuit availability is set at 99.5% over one year.

The protocols and applications that have been developed for the general Internet will be used by the GCI to send the CTBT data and data products through the GCI network. The Internet protocols are implemented in the hosts and routers of the network and they define a method for sending data through the network. The Internet protocols come standard on routers and hosts and provide a standard and effective means of communicating in diverse environments. The Transmission Control Protocol (TCP) and the User Datagram Protocol (UDP) are the Internet protocols used to provide unicast (point-to-point) communication[4]. The TCP protocol provides reliable transmission of messages, error detection, and flow control. The TCP protocol uses an explicit connection between the sender and receiver to send the data. The two ends of the connection explicitly coordinate to determine connection parameters, what has been received, buffer space, and lost messages. UDP provides an unreliable, unregulated message transmission capability. UDP provides no coordination between the sender and the receiver. The Internet protocols are used to exchange data across the network. They transport the data to its destination; they do not attempt to interpret the data inside the messages.

The Internet applications use the underlying Internet protocols (primarily TCP) to provide higher level capabilities like electronic mail (e-mail), file transfer protocol (ftp), and World Wide Web (WWW). By using the already existing Internet standards as the GCI communication mechanisms, the GCI is able to utilize all of the capabilities of the general Internet but in its own private network. These applications come standard in most operating systems and are interchangeable across different implementations.

The IMS data will be transmit across the GCI using proprietary IDC application protocols. These application protocols are the CD-x and IMS-x protocols (the x when replaced by a number identifies a particular version of the protocol). The CD-x protocol will provide reliable transmission of continuous data from primary seismic, hydroacoustic, and infrasound sensors. Non-continuous data from the IMS stations and the IDC will likely be sent using the IMS-x protocol. The IMS-x protocol will use a combination of conventional e-mail, ftp, and WWW.

The CD-1 protocol, which is already implemented, uses the TCP protocol to send data. The CD-2 protocol is being designed now and is expected to eventually replace the CD-1 protocol. The CD-2 protocol will provide many enhancements including authentication and a more advanced data format. It is being designed to be able to run over either the TCP or UDP protocol. The reliability mechanisms in CD-2 will provide application level reliability and provide reliable delivery beyond what is available in TCP. The TCP reliable message delivery works only while there is a connection between the sender and the receiver. The CD-2 protocol will provide the mechanisms for determining what the receiver is missing and recovering the missing data. CD-2 will also add mechanisms for making sure that the data reaches stable storage at the receiver. In addition, the CD-2 protocol will be responsible for forwarding data at the IDC and the NDC's of Independent Subnetworks and Partitioned Subnetworks.

All of the IMS data must be received at the IDC regardless of whether it is via the Basic Topology, a Partitioned Subnetwork, or an Independent Subnetwork. The IDC is also responsible for forwarding this data to Member States that request the data or data product. Since the CTBT monitoring network implementation is only just underway, the number of NDC's that will request data and particular data products from the IDC is not well known. Many of the Member States will have relatively low data rate connections from the IDC to their NDC, so it is unlikely that they are expecting to request significant amounts of continuous data. However, at least one Member State has requested all the continuous and non-continuous data from the IDC.

Multicast

As the Internet has grown so has the need for additional communication protocols. In particular a new class of applications has emerged that require the ability to send messages to a large group of receivers efficiently. This need has led to the development of the IP multicast capability in the Internet. IP multicast provides an unreliable communication mechanism that allows a single message to be sent to a group of receivers. IP multicast is a service implemented in the hosts and routers of the network. The multicast addresses are a separate address range recognized by the routers as multicast groups. Multicast packets are sent addressed to a multicast address. Applications that wish to receive the multicast packets open a connection to the multicast address and their host automatically transmits a join message to the nearest router. The router then adds the host to the multicast dissemination tree for the group. The multicast tree is dynamic and provides an efficient means of transmitting the packet through the network to reach all the receivers without traveling any link more than once. The routers at branch points in the tree duplicate the packet and send it down all the tree branches. For more detail regarding the multicast routing protocols, see [5] and [11].

The IP multicast communication mechanisms are gradually becoming a standard part of the Internet protocol suite and they co-exist with the TCP and UDP mechanisms. The IP multicast mechanisms do not replace the unicast mechanisms; they instead provide an additional service. Since IP multicast is still a relatively new technology routers do not come with IP multicast enabled by default. The router administrator must enable it before it can be used in the network. Unfortunately, not many commercial applications are making use of IP multicast so many Internet Service Providers (ISP) have not yet enabled the capability.

IP multicast messaging is an unreliable service similar to UDP; it provides unreliable message delivery. Reliable multicast is effectively the multicast equivalent to the TCP protocol. Reliable multicast provides reliable delivery of messages to multiple receivers. Reliable multicast uses IP multicast to provide the actual message dissemination capability and it adds reliable delivery mechanisms. A reliable multicast protocol provides several potential advantages over TCP when there are in fact multiple receivers. With reliable multicast the receivers in a group can be reached by sending a single message. Using TCP the messages would need to be sent to each receiver individually by the original sender or a site acting as a forwarder. In the case of the CTBT network the forwarding site is the IDC (and the NDC in the case of a Partitioned or Independent Subnetwork).

Reliable multicast is not yet a standard communication protocol that is part of the operating systems of hosts but it can run on the hosts. Reliable multicast is an end-to-end protocol that is run at each of the senders and receivers participating in a reliable multicast session. It usually uses IP multicast for its underlying communication mechanism. There are several commercial and freeware reliable multicast protocols available today. Some of the available reliable multicast protocols are described in [1-3], [6-10], and [12].

The reliable multicast protocols are not yet standard on the Internet. One reason is that there are competing views regarding what reliable message delivery guarantees an application should be provided. The types of applications that are supported by the reliable multicast protocols range from distributed databases to multicast ftp so they do not have a uniform set of requirements. The reliable multicast protocols are each designed for use in a particular class of applications. The application classes differ in the message delivery reliability and ordering properties required and the number of participants in a multicast group. The reliable multicast protocols also differ in their methods of indicating message loss, determining membership, and retransmitting lost messages. The underlying mechanisms and message formats of each reliable multicast protocol are different; so the different implementations can not communicate with each other. This means that sites participating in the same reliable multicast session need to run the same reliable multicast protocol. The network routers do not participate in the reliable multicast protocol; they only need to handle IP multicast messages.

RESEARCH ACCOMPLISHED

Reliable multicast can only provide benefits to the CTBT if it is implemented in the GCI network and it is used by the IDC application protocols such as CD-x and IMS-x. The potential benefits of using reliable multicast in the network could be increased reliability from eliminating data forwarding operations, network bandwidth savings and improved scalability to multiple receivers. The reliability of the continuous data collection and dissemination might be improved by eliminating the forwarding operations. But, eliminating some or all of the forwarding operations may be disallowed by the CTBT. Determination of whether particular multicasting uses are allowed by the treaty is an important issue but it is outside the scope of this paper. The legal and political issues of multicasting in the GCI will not be addressed by this paper. This paper will instead focus only on the technical issues of multicasting. Since the IDC application protocols can only make use of reliable multicast if it is provided in the GCI network, this section start with a description of how IP multicast and reliable multicast would be implemented in the GCI network.

IP Multicast in the GCI

As with the Internet, the IP multicast capabilities of the GCI are implemented in the routers and the end hosts. In order for IP multicast to work in the network the routers at the satellite hubs and in the frame relay network would need to have IP multicast enabled. The broadcast nature of the satellite portion of the GCI network provides some relatively natural multicast mechanisms. The satellite hub already broadcasts all unicast and multicast data it sends to its remote VSAT sites. If the data in a message is not addressed to any of the machines located at a particular VSAT then the VSAT throws the message away after receiving it; the VSAT passes messages sent to a multicast address. Multicasts from the satellite hub are sent once over the satellite link and received by all the remote VSAT sites connected to that hub. In the other direction the satellite link is not broadcast; multicasts sent by a remote VSAT site are received only at the satellite hub. If there are other VSAT located members of the multicast group on that same satellite hub then the hub retransmits the multicast to the remote VSAT's. This retransmission is handled by the router located at the satellite hub. Since a multicast from a VSAT to the hub reaches only the hub, no network bandwidth on this piece of the route will be saved. However, a multicast in the direction from the hub to the VSAT reaches all the VSAT's with hosts joined to the multicast group; the multicast uses only the network bandwidth required to reach one VSAT. The network bandwidth savings from using multicast increase with the number of receivers located at VSAT sites connected to the same satellite hub.

The routers in the frame relay network are located at the ends of the frame relay PVC's. These routers would also need to be configured to allow IP multicast before IP multicast would work in the frame relay portion of the GCI network. The PVC's between the routers simply act as ordinary network links. Multicast operates in a frame relay network the same way it does on the Internet; the routers build a message dissemination tree and forward multicast messages down this tree. The principle savings IP multicast could provide are from avoiding sending data down any

network link more than once. In the case of transoceanic or transcontinental links this may produce significant savings. The frame relay network can also be configured to contain a multicast mesh. Enabling a multicast mesh effects only the multicast traffic, it has no effect on the unicast traffic in the frame relay network. The unicast messages travel through the frame relay PVC's and the multicast traffic uses the mesh. The benefit of using a multicast mesh is realized when a more efficient multicast dissemination tree can be built. The mesh allows the tree to branch at any place in the frame relay network rather than only at the ends of the PVC's. The decision of whether to share the unicast PVC's or define a multicast mesh in the frame relay network is an engineering design decision that does not effect the external behavior of the network.

The GCI Integration Laboratory within the IDC contains a test network composed of all the essential components of the GCI but isolated from the GCI so that it can be used for testing without impacting the GCI. The GCI Integration Laboratory is composed of three remote sites that are connected via VSAT to a satellite hub. The satellite hub is connected to the IDC using a frame relay PVC. The routers and software in the GCI Integration Laboratory network are the same types and versions as is used in the rest of the GCI. The IP multicast capabilities were enabled on the GCI Integration Laboratory routers temporarily for feasibility tests to allow testing of IP multicast.

Extensive tests measuring throughput, loss, round trip time, and packet size were conducted in the GCI Integration Laboratory. The test software was designed to allow packets of a designated size to be generated and sent using a periodic rate. The software had the ability to send the data either one-way or with the receiver(s) bouncing the traffic back using another IP multicast address to allow round trip times to be measured. The tests indicate that the GCI Integration Laboratory components are able to pass IP multicast traffic at the full capacity of the links without problem.

One difficulty encountered is that the hosts located at the end of the frame relay link are located behind a firewall. The firewall does not allow IP multicast packets to pass through it so these machines cannot exchange IP multicast messages with the VSAT connected hosts. If the GCI network will contain firewalls between machines that must exchange IP multicasts, a way to get multicasts through the firewall will need to be determined. Most firewall products do not directly recognize IP multicast packets. The successful tests of the IP multicast capabilities and performance in the GCI Integration Laboratory were conducted by placing a machine on a LAN directly connected to the frame relay link. This machine was outside of the firewall and was thus able to exchange IP multicast messages with the VSAT located hosts.

CTBT Application Protocols

The IDC application protocols (CD-x and IMS-x) were also studied to determine feasibility of using reliable multicast as their transport mechanism. The overall data rate in the GCI is expected to be on the order of 10 gigabytes per day. Approximately 8.5 gigabytes per day of this data is continuous data sent using the CD-x protocol. Since the CD-1 protocol is already in use and is expected to be phased out the study instead focused on the CD-2 protocol, which is still in the design phase.

The current CD-2 design specifications assume that data will be transmitted using unicast. This assumption is most apparent in the specification of the connection setup and termination. With unicast connections there is only one sender and one receiver so the connection is either up or down. Multicast introduces the concept of multiple receivers and a multicast group. The members of the group are notified of membership changes and the current membership of the group. The determination of whether the sender or a specific receiver is in the group is made by examining the group membership rather than the connection state. The CD-2 protocol design would likely require minor modifications to accommodate this change. The other aspect of the CD-2 design that would be affected by reliable multicast is the reliable delivery of frames. Acknowledgements and retransmission requests would be coming from multiple receivers if reliable multicast were used under CD-2. This may affect the specification but it will need to be considered in the CD-2 protocol implementation. Since there are relatively few subscribers to the CD-2 data from the IMS stations reliable multicast is unlikely to provide any scaling benefits. The benefits would more likely be from bandwidth savings and any improvements in reliability that can be achieved by removing forwarding operations.

The IMS-x protocol will likely be used to disseminate the data products generated by the IDC and to transmit any non-continuous IMS station data. The e-mail, ftp and WWW services will likely be provided by the standard Internet applications available in most operating systems today. The Internet standard versions of these applications are based on unicast communication. There are, however, commercial and research versions of ftp and WWW that are based on reliable multicast. The network news application also has versions available that support reliable multicast and it could potentially be used instead of e-mail. But, in order to use reliable multicast-based services in the IMS-x protocol the reliable multicast-based versions of each of these reliable multicast enabled applications would need to be installed at each of the sites participating in IMS-x data exchanges. Currently the only data that will likely be transmitted using IMS-x and have a significant number of recipients is the data products generated by the IDC. These data products are sent to only to NDC's subscribed to the data product. The connections from the IDC to the various NDC's subscribed to a particular data product are likely to be a combination of Internet, satellite and frame relay links. Thus, the network bandwidth savings may not turn out to be significant. More information than is available today about the subscription list of each of the data products would be needed to make this determination. In addition, it may take a while to get IP multicast enabled between the IDC and the NDC's connected via the Internet depending on the Internet Service Providers involved.

Reliable Multicast in the GCI

The reliable multicast protocols were also evaluated to determine whether an existing reliable multicast protocol could serve the needs of the CTBT. One of the critical delineators between different reliable multicast protocols is whether they support coordination between multiple senders in a group. The IMS-x and CD-x would not require coordination between different sites sending messages so the reliable multicast protocol would only need to support a single sender per reliable multicast group. The reliable multicast group sizes (receiver set) will also likely be relatively small (less than 200) for all IDC applications and quite small (less than 10) for the CD-x application protocol.

Some of the existing reliable multicast protocols that provide the characteristics described in the paragraph above are the Multicast Dissemination Protocol (MDP)[8] and the Reliable Multicast Transport Protocol (RMTP)[7]. These protocols are available, they are deployed in applications, and they have on-going development projects. Many characteristics differentiate the above protocols and these characteristics could be used to narrow the choices. If the decision to use a reliable multicast protocol in the GCI is made, then additional criteria such as membership, congestion control, and security will also need to be considered in determining which reliable multicast protocol to use.

The MDP protocol which is a freely available protocol was obtained via the WWW and installed in the GCI Integration Laboratory to test the basic capability of the reliable multicast protocols to run in the GCI network. Reliable multicast transfers of jpeg images were performed using the MDP protocol. The images were transferred back and forth between remote VSAT located sites and an IDC located machine. The MDP protocol provided reliable multicast over these GCI links without requiring any modification or tuning to account for the satellite links. The other reliable multicast protocols were not tested in the GCI Integration Laboratory. The MDP is source code available and free making it a prime candidate for use in the GCI. The RMTP protocol is a commercial product so it has associated licensing fees and is not source code available. In addition, the RMTP protocol contains mechanisms for scaling to exceedingly large groups. These mechanisms would not be needed in the GCI but, they would require additional configuration and maintenance.

CONCLUSIONS AND RECOMMENDATIONS

Reliable multicast protocols provide a capability to send a single message to multiple receivers. Using reliable multicast protocols can sometimes improve scalability, efficiency or reliability of an application. This paper reports on a study that has been undertaken by the author to determine whether use of reliable multicast can benefit the CTBT monitoring network. Since the CTBT monitoring network is not yet fully established there are many details of the networks and protocols that have not been finalized. This made studying the possible effect of reliable multicast in the CTBT monitoring network more difficult. The study was, however, able to determine that IP multicast and reliable multicast can run in the GCI network. Some candidate reliable multicast protocols for use in the network were presented and a likely candidate, the MDP protocol, was tested in the GCI Integration Laboratory.

The CD-2 protocol may be able to gain benefits from using reliable multicast in the GCI. But, a determination of what uses of reliable multicast are allowed under the CTBT would need to be made. If reliable multicast is to be used in the GCI, then a final decision regarding which reliable multicast protocol to use would need to examine the detailed needs of the application protocol.

REFERENCES

[1] D. Agarwal, P. Melliar-Smith, L. Moser, and R. Budhia, "Reliable Ordered Delivery Across Interconnected Local-Area Networks," Transactions on Computer Systems, vol. 16, no. 2 (May 1998).

[2] Y. Amir and J. Stanton, "The Spread Wide Area Group Communication System," Johns Hopkins University Technical Report, CNDS-98-4, available from ftp://ftp.cnds.jhu.edu/pub/papers/spread.ps.

[3] K. Birman and T. Joseph, "Reliable communication in the presence of failures," ACM Transactions on Computer Systems, vol. 5, no. 1, pp. 47-76, February 1987.

[4] D. Comer, *Internetworking with TCP/IP: Principles, Protocols, and Architecture*, 2nd ed, Volume I, New Jersey: Prentice-Hall, 1991.

[5] D. Estrin, D. Farinacci, A. Helmy, D. Thaler, S. Deering, M. Handley, V. Jacobson, P. Sharma, L. Wei, "Protocol Independent Multicast," IETF RFC 2362, available from ftp://ftp.isi.edu/in-notes/rfc2362.txt

[6] S. Floyd, V. Jacobson, C. Liu, S. McCanne, and L. Zhang, "A Reliable Multicast Framework for Light-weight Sessions and Application Level Framing," IEEE/ACM Transactions on Networking, December 1997, Volume 5, Number 6, pp. 784-803.

[7] J. Lin and S. Paul, "RMTP: A Reliable Multicast Transport Protocol," IEEE INFOCOM '96, March 1996, pp. 1414-1424.

[8] J. Macker and W. Dang, "The Multicast Dissemination Protocol version 1 (mdpv1) Framework," Technical white paper, US Naval Research Laboratory, available from http://tonnant.itd.nrl.navy.mil/docs/mdpv1.ps.

[9] K. Miller, K. Robertson, A. Tweedly, M. White, "StarBurst Multicast File Transfer Protocol (MFTP) Specification," IETF Draft Specification, draft-miller-mftp-spec-03.txt, dated April 1998.

[10] W. Strayer, S. Gray, and R. Cline, Jr., "An Object-Oriented Implementation of the Xpress Transfer Protocol," Proceedings of the Second International Workshop on Advanced Communications and Applications for High-Speed Networks (IWACA), Heidelberg, Germany, September 26-28, 1994.

[11] D. Waitzman, C. Partridge, S. Deering, "Distance Vector Multicast Routing Protocol," IETF RFC 1075, available from ftp://ftp.isi.edu/in-notes/rfc1075.txt

[12] B. Whetten, M. Basavaiah, S. Paul, T. Montgomery, N. Rastogi, J. Conlan, T. Yeh, "The RMTP-II Protocol," Internet Draft, draft-whetten-rmtp-ii-00.txt and draft-whetten-rmtp-ii-app-00.txt, dated April 1998.