DISCUSSION OF RELIABLE MULTICAST DEPLOYMENT PROGRESS FOR THE CONTINUOUS DATA PROTOCOL

Deborah A. Agarwal

Ernest Orlando Lawrence Berkeley National Laboratory

Sponsored by National Nuclear Security Administration
Office of Nonproliferation Research and Engineering
Office of Defense Nuclear Nonproliferation

Contract No. DE-M9AL-66156.501

ABSTRACT

The International Monitoring System (IMS) seismic sensor data are currently collected using point-to-point networking protocols. Multicast communication allows a single transmission of the data from a sensor to be received by multiple sites (point-to-multipoint). This capability has the potential to improve fault tolerance and possibly efficiency of the sensor data collection and dissemination process. An experiment was conducted to demonstrate the collection and dissemination of seismic sensor data using reliable multicast communications. Telcordia and SAIC created a prototype multicast-capable version of the Continuous Data (CD-1) protocol as an experiment. This prototype version used the RMTP reliable multicast protocol available from Talarian Corporation to transport the sensor data. The experiment demonstrated that reliable multicast is a viable technology for use in transmitting the seismic data.

The CD-1.1 protocol has since been developed and released. The CD-1.1 protocol incorporates many enhancements that make it better suited than the CD-1 protocol to the use of reliable multicast. Initial studies into the feasibility of implementing a reliable multicast version of the CD-1.1 protocol have begun. This paper provides an overview of the current progress in the study of the possible use of reliable multicast in the transmission of continuous data from IMS stations and the current state of reliable multicast development.

KEY WORDS: communications, multicast, reliable

OBJECTIVE

Reliable multicast is a communication capability that can be used in the network to allow a message to be sent from a single sender to multiple receivers. It uses IP multicast for the transmission of message on the network. IP multicast is a simple communication mechanism that allows a single message to be sent to a group of receivers at the network level. With reliable multicast the receivers in a group can be reached by sending a single message. Using unicast the messages would need to be sent to each receiver individually by the sender or a site acting as a forwarder.

IP multicast is an unreliable messaging service implemented in the hosts and routers of the network. Multicast packets are sent addressed to an address in the multicast address range. Applications that wish to receive the multicast packets open a connection to the multicast address. The multicast capability provides an efficient means of transmitting a packet through the network to reach all the receivers. The IP multicast communication mechanisms are now a standard part of the Internet protocol suite and they co-exist with the unicast Transmission Control Protocol (TCP). and User Datagram Protocol (UDP) mechanisms. The IP multicast mechanisms do not replace the unicast mechanisms; they instead provide an additional service.

Reliable multicast is effectively the multicast equivalent of the TCP protocol. Reliable multicast provides reliable delivery of messages to multiple receivers. It uses IP multicast to provide the message dissemination capability and adds reliable delivery mechanisms. Reliable multicast is not yet a standard

communication protocol that is part of the operating systems of hosts. Reliable multicast is instead run as an application-level protocol. An instance of the reliable multicast software is run at each of the senders and receivers participating in a reliable multicast session. The software then uses IP multicast for its underlying communication mechanism. There are several commercial and freeware reliable multicast protocols available today. Two of the existing reliable multicast protocols are the Multicast Dissemination Protocol (MDP)[7] and the Reliable Multicast Transport Protocol (RMTP)[8].

The CD-1 protocol is the protocol currently in use for sending continuous data from the IMS seismic sensors. The CD-1 protocol runs at the sender of the data and at the receiver and is responsible for transmission of the data to the receiver. The CD-1 protocol is designed to provide transmission of continuous data between a sender and receiver pair. The CD-1 protocol retrieves the data from a Last In First Out (LIFO) Heap at the sending side and stores it in a Disk Loop at the receiving site. The protocol uses a TCP connection to transmit the data from the LIFO Heap to the receiver. TCP provides unicast, reliable, source-ordered delivery of messages between the sender and receiver. CD-1 delivers data only while there is a TCP connection between the sender and the receiver. When the connection is down the sender buffers data locally in the LIFO heap waiting until a connection can be re-established to the receiver. Some amount of data may be lost by the CD-1 protocol when a TCP connection is closed due to a failure.

The objective of the work reported in this paper is to study the potential use of reliable multicast as a communication mechanism for the CD-x protocol. The expected benefits of this change would be improved fault-tolerance and efficiency.

RESEARCH ACCOMPLISHED

Early work on the study focused on the IP multicast capabilities of the network and some suggestions for appropriate reliable multicast protocols. The CD-1 protocol, the IMS data rates and data delivery criteria were also studied. The findings from these early studies are reported in [2], [11]. At the completion of these preliminary studies an experiment was conducted. In the experiment a prototype multicast capable version of the continuous data protocol (CD-1) was created[1]. The purpose of the experiment was to provide a small-scale technical feasibility trial of the use of reliable multicast as a transport mechanism for the continuous data. A detailed discussion of the experiment is contained in [3].

A Multicast Experiment Using the CD-1 Protocol

At the beginning of the experiment the CD-1 protocol was the only implemented version of the continuous data protocol. The CD-1 protocol contains no end-to-end reliability mechanisms, so it was not an ideal candidate for long-term use with reliable multicast. However, the use of CD-1 for a technical feasibility prototype allowed rapid development for testing. To reduce development cost and time, a goal of the prototype multicast-enabled implementation of CD-1 was to use as much of the existing code as possible. Another goal of the experiment was to have a comparable frame loss rate to that exhibited by the existing CD-1 protocol implementation.

At the beginning of the experiment, the MDP and RMTP-II reliable multicast protocols were evaluated. These are the two protocols identified by the earlier study to be the best candidates for use with the CD-x protocol. The principle deciding factors were the application-programming interface (API) and the availability of commercial support. The RMTP-II protocol provided both these features and was thus chosen for use in the experiment. The experiment also provided an opportunity to test the robustness of the RMTP-II from Talarian Corporation.

---

[1] The design, implementation, and testing of the multicast-enabled version of the CD-1 protocol were carried out by Telcordia Technologies and SAIC. Sponsored by U.S. Department of Defense, Defense Threat Reduction Agency, Contract No. DTRA01-99-C-0025.

The CD-1 protocol's original design was based on an assumption that the underlying data transmission was unicast. The LIFO heap at the CD-1 sender is used to buffer up data waiting for transmission to the receiver. The multicast-enabled version of CD-1 replicates the sender's LIFO heap at the receiver. Reliable multicast is then inserted between the LIFO heaps and is used to transfer the data between the sender's and receivers' LIFO heaps. This allowed the changes to the existing CD-1 protocol implementation to be minimized.

During failures, there is more needed to provide behavior equivalent to the original behavior of the CD-1 system. If the sender crashes and recovers, the sending of data in the unicast CD-1 was discontinued during the crash and resumed after the recovery. The behavior of the multicast-enabled CD-1 is the same in this case. The crash of a receiver in the unicast CD-1causes the sender to buffer up data until the receiver recovers and re-establishes a connection. But, if the multicast enabled version of the CD-1 software were to stop transmitting when any one receiver was down then the perceived reliability of the CD-1 software from the other receiver(s) would be less than the original unicast CD-1 software. In the multicast version, if a receiver crashes, the frame sending continues and the operational receivers continue to receive frames. When the receiver recovers, it rejoins the multicast group and resumes receiving frames. The missed frames were stored in a "catch-up" LIFO Heap at the sender and the frames are sent to the recovered receiver using a unicast TCP connection in parallel with the ongoing multicast connection. "Catch-up" frames and new frames are merged at the receiver.

The prototype multicast-based CD-1 implementation was tested in several configurations to evaluate its performance. The initial tests of the system were performed at Telcordia using two receivers. These tests only sent data through to the LIFO Heap at the receiver. The next series of tests were between Telcordia and the Prototype International Data Center (PIDC). In these tests, the sender was at Telcordia and the receivers were at Telcordia and the PIDC. This configuration allowed testing with a moderate latency link over the Internet. They also allowed the software to be tested all the way through to the Disk Loop Manager (DLMan).

The tests between Telcordia and the PIDC were run continuously for eight days. Six times during the eight days the receiver at the PIDC became unreachable from the sender at Telcordia. In each of these cases, the catch-up channel was activated when the receiver rejoined the multicast channel. The correct transfer of the merged catch-up and multicast data to the DLMan at the receiver was also tested.

The final tests used a version of the RMTP-II protocol that allowed the network loss and latency characteristics to be emulated. In these tests characteristics representative of the satellite network were emulated. All the test configurations used one sender and two receivers. In these emulations, the network delay between the sender and the receivers used an exponential distribution with a mean of 1.2 seconds, a standard deviation of 0.2 seconds and cut-offs at 1.0 and 1.7 seconds. The loss probability was set to 0.5%. The CD-1 prototype performed well in this test. Successful tests of all combinations of system start-up and recovery/catch-up were also performed. The tests between Telcordia and the PIDC were left running for four weeks. The data transfer ran without a problem during that time and showed results comparable to the original unicast CD-1.

The CD-1.1 Protocol

At the beginning of the multicast study, the CD-1.1 version of the protocol existed only in specification form. The CD-1.1 protocol has since been implemented and the implementation of the CD-1.1 protocol is currently undergoing testing at the PIDC using TCP communication[2]. The CD-1.1 protocol includes several features that should make it easier to use other protocols besides TCP for data transmission. These features include retransmission request mechanisms, self-describing data frames, and a more flexible connection setup method. These features are expected to make the integration of reliable multicast with CD-1.1 more straightforward than was the case with CD-1. In particular, the connection setup frame contains a field that can be used for specifying a multicast address for use in data transmission. The retransmission request mechanisms provide a means of retrieving lost data. These are directly of benefit to

[2] A beta version of the software is available from http://www.pidc.org/

both a unicast and multicast-based version of CD-1.1 since they will allow any missing frames to be retrieved directly through CD-1.1. In CD-1 missing frames could only be retrieved through bulk data request mechanisms.

Multicast Standards Activities

Reliable multicast and IP multicast development efforts have made significant advances since the beginning of the CD-1 multicast experiment. There are many groups within the Internet Engineering Task Force (IETF) working on multicast related standards for the Internet. The two groups of primary interest to this study are the Source Specific Multicast (SSM) working group[3] and the Reliable Multicast Transport (RMT) working group[4].

Source-specific multicast is an IP multicast capability that allows members of a multicast group to subscribe to a specific source[5]. This allows the receiver to restrict the multicast to delivery of messages generated by that specific source only. This capability is being added to the Internet to improve handling of well-known multicast sources, access control, and scalability of the multicast address space. The first two of these are directly applicable to the IMS network since the IMS sites are normally well-known sources. This capability will allow the receivers to subscribe specifically to the data sources of interest and not receive traffic from the other members of the multicast group. The SSM working group is likely to finish its work quickly since many of the router vendors already have working versions of source-specific multicast.

The RMT working group is tackling the problem of providing Internet standards for reliable multicast protocols. Since reliable multicast protocols are generally built with application specific goals in mind, the protocols have different message delivery properties and different methods of achieving reliability[6]. For example, some reliable multicast protocols use retransmission requests to retrieve missed messages and others use forward error-correcting codes (FEC) to eliminate the need for retransmissions. With FEC, redundant data is placed in each message and the net effect is that the receiver will be able to reconstruct the entire data stream despite missing some of the packets. Some reliable multicast protocols provide bulk-data transfer capabilities and some are intended for support of real-time applications. This difference is generally seen in the timeliness of retransmissions. Some bulk-data transfer protocols wait to send retransmissions until all the data has been sent once. Some reliable multicast protocols acknowledge messages that have been received and others send negative acknowledgements indicating what messages are missing.

Despite the differences between reliable multicast protocols, there are many common tasks that are shared by these protocols. The RMT working group is defining standardized building blocks for reliable multicast protocols[9]. The intent of these building blocks is to identify the common components and standardize these components. There are currently several building blocks under development. These building blocks address congestion control, message reliability mechanisms, and mechanisms for router assistance. The idea is for each protocol to be composed of a subset of the building blocks along with any specialized components it requires. The benefit of this approach is that the building blocks are many of the underlying core components of a reliable multicast protocol and standardizing these improves the robustness of all the protocols that use them. The Talarian Corporation personnel involved in RMTP-II are also heavily involved in these standards activities and will likely be quick to adopt the standards.

CONCLUSIONS AND RECOMMENDATIONS

The use of reliable multicast as a method of sending continuous data from the IMS stations continues to be studied. Initial work studied the network, IMS data rates, and the continuous data protocols. An experiment was conducted using a version of the CD-1 protocol that was modified to use reliable multicast communication as the underlying data transport mechanism instead of TCP. The experiment demonstrated

[3] The SSM working group charter can be viewed at http://www.ietf.org/html.charters/ssm-charter.html. The group's working documents are available at http://sith.maoz.com/SSM/.
[4] The RMT working group charter can be viewed at http://www.ietf.org/html.charters/rmt-charter.html.

the feasibility of using reliable multicast for transmission of continuous data. The fault-tolerance introduced by this approach was also demonstrated in the experiment. The CD-1.1 protocol is now available as a beta release. The CD-1.1 protocol provides many new features that are likely to ease the integration of multicast data transmission capabilities. A possible future activity will be to develop a reliable multicast-based version of the CD-1.1protocol.

Over the past several years there have also been several developments in reliable multicast and IP multicast capabilities for the Internet. There are several important multicast related standards activities within the IETF. The building blocks being designed to standardize reliable multicast are likely to provide robust versions of components critical to reliable multicast protocols. Source-specific multicast is another IETF activity that promises to provide improved access control and multicast routing.

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] D. Agarwal, "Using Multicast in the Global Communications Infrastructure for Group Communication," in the Proceedings of the 22nd Annual Seismic Symposium, Las Vegas, Nevada, September 1999.

[3] D. A. Agarwal, R. Stead, J. E. Burns, N. Shah, and N. Kyriakopoulos, "Initial Results Of The CD-1 Reliable Multicast Experiment," published at the Global Communications Infrastructure Workshop, Vienna, Austria, October 2000.

[4] K. Berket, D. A. Agarwal, P. M. Melliar-Smith, and L. E. Moser, "Overview of the InterGroup Protocols," Proceedings of the 2001 International Conference on Computational Science. LNCS 2073. Springer-Verlag. pp. 316-25.

[5] S. Bhattacharyya, C. Diot, L. Giuliano, R. Rockell, J. Meylor, D. Meyer, G. Shepherd, B. Haberman, "An Overview of Source-Specific Multicast(SSM) Deployment," Internet Draft <draft-ietf-ssm-overview-00.txt>. Working Group URL - http://sith.maoz.com/SSM/.

[6] M. Handley, S. Floyd, B. Whetten, R. Kermode, L. Vicisano, and M. Luby, "The Reliable Multicast Design Space for Bulk Data Transfer," Internet Draft, Internet Engineering Task Force, RFC 2887, August 2000.

[7] 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.

[8] 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.

[9] B.Whetten, L. Vicisano, R. Kermode, M. Handley, S.Floyd, and M. Luby, "Reliable Multicast Transport Building Blocks for One-to-Many Bulk-Data Transfer," Internet Draft, Internet Engineering Task Force, RFC 3048, January 2001.

[10] "Formats and Protocols for Continuous Data CD-1.1," Published by Scientific Applications International Corporation as part of the International Data Centre Documentation, July 2000. Available from http://www.pidc.org/librarybox/idcdocs/downloads/343.pdf.

[11] "Multicasting in the GCI - A Report of Study Results," available on the CTBT Expert's Communication System as CTBT/WGB/TL-3/9/Rev.1/Amend.1, 2 September 1999.