IETF
Internet-Draft
Intended status: Best Current Practice
Expires: December 14, 2013

G. Shepherd, Ed.
Cisco Systems
June 12, 2013

Multicast UDP Usage Guidelines for Application Designers
draft-shepherd-multicast-udp-guidelines-01

Abstract

The multi-recipient nature of Multicast prevents the use of any pont-to-point connection-oriented transport, therefore restricts all Multicast data to be sent over the User Datagram Protocol (UDP). UDP provides a minimal message-passing transport that has no inherent congestion control mechanisms. Because congestion control is critical to the stable operation of the Internet, applications and upper-layer protocols that choose to use Multicast UDP as an Internet service must employ mechanisms to prevent congestion collapse and to establish some degree of fairness with concurrent traffic. This document provides guidelines on the use of UDP for the designers of multicast applications and higher-level protocols.

Status of This Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on December 14, 2013.

Copyright Notice

Table of Contents

1. Introduction

The User Datagram Protocol (UDP) [RFC0768] provides a minimal, unreliable, best-effort, message-passing transport to applications and upper-layer protocols (both simply called "applications" in the remainder of this document). [RFC5405] is scoped to provide guidelines for unicast applications only, but all of the general requirements, references, and use cases apply to multicast [RFC1112][RFC4607] UDP application designers as well. This document chooses to only make recommendations in requirements, use cases, and references where they differ from [RFC5405]or are unique for applications sending multicast UDP data (simply called "multicast" in the remainder of this document).

1.1. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119]

2. Multicast UDP Usage Guidelines

2.1. Congestion Control Guidelines

[RFC2309] discusses the dangers of congestion-unresponsive flows and states that "all UDP-based streaming applications should incorporate effective congestion avoidance mechanisms". Many large-scale multicast deployments are within a single administrative domain, and are provisioned over a bandwidth-reserved path or paths where congestion control is less relevant. But there are a growing number of deployment cases where multicast is transiting multiple domains, is tunneled across the unicast Internet, or transits the Internet through a unicast overlay network. This document is only concerned with the latter case of multicast data transiting the larger Internet, either as native IP multicast or encapsulated in a unicast tunnel and does not apply to administratively scoped deployments.

When the multicast traffic exits the administrative domain of a single network or the bi-laterally agreed path between networks, or is tunneled across the unicast Internet either to another multicast network or to an end device, the application SHOULD provide a TCP-compatible aggregate flow over the end-to-end path to each leaf.

There are currently two models of multicast delivery: the Any-Source Multicast (ASM) model as defined in [RFC1112] and the Source-Specific Multicast (SSM) model as defined in [RFC4607]. ASM group members will receive all data sent to the group by any source, while SSM constrains the distribution tree to only one single source. Many congestion-controlled transport protocols are often not applicable to multicast distribution services, or simply won’t scale well to very large multicast trees since they require bi-directional communication and adapt the data-rate to accommodate the network conditions to a single receiver. Multicast distribution trees can often fan out to massive numbers of receivers limiting the scalability of an in-band return channel to control the data-rate, and the one-to-many nature of multicast distribution trees prevent adapting the data-rate to individual receiver requirements. For this reason, TCP-compatible aggregate flow for Internet multicast data, either native or tunneled, is the responsibility of the application.

2.1.1. Bulk Transfer Applications

Applications that perform bulk transmission of data over a multicast distribution tree, i.e., applications that exchange more than a small number of UDP datagrams per maximum receiver RTT, SHOULD implement Asynchronous Layered Coding (ALC) [RFC5775], TCP-Friendly Multicast Congestion Control (TFMCC) [RFC4654], Wave and Equation Based Rate Control (WEBRC) [RFC3738], NACK-Oriented Reliable Multicast (NORM)

transport protocol [RFC5740], File Delivery over Unidirectional Transport (FLUTE) [RFC6726], Real Time Protocol/Control Protocol (RTP /RTCP), [RFC3550] or another congestion control scheme following the guidelines of [RFC2887]and utilizing the framework of [RFC3048].

Bulk transfer applications that choose not to implement [RFC4654], [RFC5775], [RFC3738], [RFC5740], [RFC6726], or [RFC3550] SHOULD implement a congestion control scheme that results in bandwidth use that competes fairly with TCP within an order of magnitude. Section 2 of [RFC3551] suggests that applications SHOULD monitor the packet loss rate to ensure that it is within acceptable parameters. Packet loss is considered acceptable if a TCP flow across the same network path under the same network conditions would achieve an average throughput, measured on a reasonable timescale, that is not less than that of the UDP flow. The comparison to TCP cannot be specified exactly, but is intended as an "order-of-magnitude" comparison in timescale and throughput.

Finally, some bulk transfer applications may choose not to implement any congestion control mechanism and instead rely on transmitting across reserved path capacity. This might be an acceptable choice for a subset of restricted networking environments, but is by no means a safe practice for operation in the Internet. When the multicast traffic of such applications leaks out on unprovisioned Internet paths, it can significantly degrade the performance of other traffic sharing the path and even result in congestion collapse. Applications that support an uncontrolled or unadaptive transmission behavior SHOULD NOT do so by default and SHOULD instead require users to explicitly enable this mode of operation.

2.1.2. Low Data-Volume Applications

All of the recommendations in section 3.1.2 of [RFC5405] are applicable to multicast as well.

2.1.3. UDP Tunnels

All of the recommendations in section 3.1.3 of [RFC5405] are applicable to multicast carried inside of unicast UDP tunnels. There are, however deployment cases and solutions where the outer header of a UDP tunnel contains a multicast destination address, such as [RFC6513], but these are primarily deployed in bandwidth reserved environments within a single administrative domain, or between two domains where a bi-laterally agreed upon path and bandwidth is in place and so congestion control is not an issue.

2.1.4. Message Size Guidelines

IP fragmentation lowers the efficiency and reliability of Internet communication. The loss of a single fragment results in the loss of an entire fragmented packet, because even if all other fragments are received correctly, the original packet cannot be reassembled and delivered. This fundamental issue with fragmentation exists for both IPv4 and IPv6, unicast and multicast packets. In addition, some network address translators (NATs) and firewalls drop IP fragments. The network address translation performed by a NAT only operates on complete IP packets, and some firewall policies also require inspection of complete IP packets. Even with these being the case, some NATs and firewalls simply do not implement the necessary reassembly functionality, and instead choose to drop all fragments. Finally, [RFC4963] documents other issues specific to IPv4 fragmentation.

Due to these issues, a multicast application SHOULD NOT send UDP datagrams that result in IP packets that exceed the effective MTU as described in section 3 of [RFC6807]. Consequently, an application SHOULD either use the effective MTU information provided by the Population Count Extensions to Protocol Independent Multicast [RFC6807] or implement path MTU discovery itself [RFC1191][RFC1981][RFC4821] to determine whether the path to each destination will support its desired message size without fragmentation.

If the multicast application is incapable of, or choose not to implement a worst-cast path MTU solution, the application SHOULD assume the maximum MTU of any link will be affected by multiple levels of encapsulation and SHOULD NOT send any packet larger than 1280 bytes.

3. Acknowledgements

This template was derived from an initial version written by Pekka Savola and contributed by him to the xml2rfc project.

This document is part of a plan to make xml2rfc indispensable [DOMINATION].

4. IANA Considerations

This memo includes no request to IANA.

All drafts are required to have an IANA considerations section (see the update of RFC 2434 [I-D.narten-iana-considerations-rfc2434bis] for a guide). If the draft does not require IANA to do anything, the section contains an explicit statement that this is the case (as above). If there are no requirements for IANA, the section will be removed during conversion into an RFC by the RFC Editor.

5. Security Considerations

All drafts are required to have a security considerations section. See RFC 3552 [RFC3552] for a guide.

6. References

6.1. Normative References

[RFC0768] Postel, J., "User Datagram Protocol", STD 6, RFC 768, August 1980.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[min_ref] authSurName, authInitials., "Minimal Reference", 2006.

6.2. Informative References

[DOMINATION] Mad Dominators, Inc., "Ultimate Plan for Taking Over the World", 1984, <http://www.example.com/dominator.html>.

[I-D.narten-iana-considerations-rfc2434bis] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", draft-narten-iana-considerations-rfc2434bis-09 (work in progress), March 2008.

[RFC1112] Deering, S., "Host extensions for IP multicasting", STD 5, RFC 1112, August 1989.

[RFC1191] Mogul, J. and S. Deering, "Path MTU discovery", RFC 1191, November 1990.

[RFC1981] McCann, J., Deering, S., and J. Mogul, "Path MTU Discovery for IP version 6", RFC 1981, August 1996.

[RFC2309] Braden, B., Clark, D., Crowcroft, J., Davie, B., Deering, S., Estrin, D., Floyd, S., Jacobson, V., Minshall, G., Partridge, C., Peterson, L., Ramakrishnan, K., Shenker,

S., Wroclawski, J., and L. Zhang, "Recommendations on Queue Management and Congestion Avoidance in the Internet", RFC 2309, April 1998.

[RFC2629] Rose, M., "Writing I-Ds and RFCs using XML", RFC 2629, June 1999.

[RFC2887] Handley, M., Floyd, S., Whetten, B., Kermode, R., Vicisano, L., and M. Luby, "The Reliable Multicast Design Space for Bulk Data Transfer", RFC 2887, August 2000.

[RFC3048] Whetten, B., Vicisano, L., Kermode, R., Handley, M., Floyd, S., and M. Luby, "Reliable Multicast Transport Building Blocks for One-to-Many Bulk-Data Transfer", RFC 3048, January 2001.

[RFC3550] Schulzrinne, H., Casner, S., Frederick, R., and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications", STD 64, RFC 3550, July 2003.

[RFC3551] Schulzrinne, H. and S. Casner, "RTP Profile for Audio and Video Conferences with Minimal Control", STD 65, RFC 3551, July 2003.

[RFC3552] Rescorla, E. and B. Korver, "Guidelines for Writing RFC Text on Security Considerations", BCP 72, RFC 3552, July 2003.

[RFC3738] Luby, M. and V. Goyal, "Wave and Equation Based Rate Control (WEBRC) Building Block", RFC 3738, April 2004.

[RFC4607] Holbrook, H. and B. Cain, "Source-Specific Multicast for IP", RFC 4607, August 2006.

[RFC4654] Widmer, J. and M. Handley, "TCP-Friendly Multicast Congestion Control (TFMCC): Protocol Specification", RFC 4654, August 2006.

[RFC4821] Mathis, M. and J. Heffner, "Packetization Layer Path MTU Discovery", RFC 4821, March 2007.

[RFC4963] Heffner, J., Mathis, M., and B. Chandler, "IPv4 Reassembly Errors at High Data Rates", RFC 4963, July 2007.

[RFC5405] Eggert, L. and G. Fairhurst, "Unicast UDP Usage Guidelines for Application Designers", BCP 145, RFC 5405, November 2008.

[RFC5740] Adamson, B., Bormann, C., Handley, M., and J. Macker, "NACK-Oriented Reliable Multicast (NORM) Transport Protocol", RFC 5740, November 2009.

[RFC5775] Luby, M., Watson, M., and L. Vicisano, "Asynchronous Layered Coding (ALC) Protocol Instantiation", RFC 5775, April 2010.

[RFC6513] Rosen, E. and R. Aggarwal, "Multicast in MPLS/BGP IP VPNs", RFC 6513, February 2012.

[RFC6726] Paila, T., Walsh, R., Luby, M., Roca, V., and R. Lehtonen, "FLUTE - File Delivery over Unidirectional Transport", RFC 6726, November 2012.

[RFC6807] Farinacci, D., Shepherd, G., Venaas, S., and Y. Cai, "Population Count Extensions to Protocol Independent Multicast (PIM)", RFC 6807, December 2012.

Appendix A. Additional Stuff

This becomes an Appendix.

Author's Address

Greg Shepherd (editor)
Cisco Systems
Tasman Drive
San Jose
USA

Email: gjshep@gmail.com