conex Working Group S. Krishnan
Internet-Draft Ericsson
Intended status: Standards Track M. Kuehlewind
Expires: September 13, 2012 IKR University of Stuttgart
C. Ucendo
Telefonica
March 12, 2012

IPv6 Destination Option for Conex
draft-ietf-conex-destopt-02

Abstract

Conex is a mechanism by which senders inform the network about the congestion encountered by packets earlier in the same flow. This document specifies an IPv6 destination option that is capable of carrying conex markings in IPv6 datagrams.

Table of Contents

1. Introduction

Conex [CAM] is a mechanism by which senders inform the network about the congestion encountered by packets earlier in the same flow. This document specifies an IPv6 destination option [RFC2460] that can be used for performing conex markings in IPv6 datagrams.

2. Conventions used in this document

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].

3. Background

The Conex working group came up with a list of requirements that had to be met by any marking mechanism. It then considered several alternative mechanisms and evaluated their suitability for conex marking. There were no mechanisms found that were completely suitable, but the only mechanism that came close to meeting the requirements was IPv6 destination options. The analysis of the different alternatives can be found in [draft-krishnan-conex-ipv6].

4. Requirements for marking IPv6 packets

R-1: The marking mechanism needs to be visible to all conex-capable nodes on the path.

R-2: The mechanism needs to be able to traverse nodes that do not understand the markings. This is required to ensure that conex can be incrementally deployed over the Internet.

R-3: The presence of the marking mechanism should not significantly alter the processing of the packet. This is required to ensure that conex marked packets do not face any undue delays or drops due to a badly chosen mechanism.

R-4: The markings should be immutable once set by the sender. At the very least, any tampering should be detectable.

Based on these requirements four solutions to implement the ConEx information in the IPv6 header have been investigated: Hop-by-Hop options, destination options, using header bits, and Option Headers. Only the use of Destination Option could fulfill the requirements.

5. Conex Destination Option (CDO)

The Conex Destination Option (CDO) is a destination option that can be included in IPv6 datagrams that are sent by conex-aware senders in order to inform conex-aware nodes on the path about the congestion encountered by packets earlier in the same flow. The CDO has an alignment requirement of (none).

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
                                +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
                                |  Option Type  | Option Length |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|X|L|E|C|                       Reserved                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 1: Conex Destination Option Layout

Option Type

8-bit identifier of the type of option. The option identifier for the conex destination option will be allocated by the IANA.

Option Length

8-bit unsigned integer. The length of the option (excluding the Option Type and Option Length fields). This field MUST be set to the value 4.

X Bit

When this bit is set, the transport sender is using ConEx with this packet. If it is reset, the sender is not using ConEx with this packet.

L Bit

When this bit is set, the transport sender has experienced a loss. If it is reset, the sender has not experienced a loss.

E Bit

When this bit is set, the transport sender has experienced ECN-signaled congestion. If it is reset, the sender has not experienced ECN-signaled congestion.

C Bit

When this bit is set, the transport sender is building up congestion credit. Otherwise it is not.

All packets of a ConEx-capable connection MUST carry the CDO.

If the X bit is the zero all other three bits are undefined. If the X bit is zero that means that the connection is ConEx-capable but this packet SHOULD NOT be accounted to determine ConEx information in an audit function. This can be the case for e.g. pure control packets not carrying any user data. As an example in TCP pure ACKs are usually not ECN-capable and TCP does not have an mechanism to announce the lost of a pure ACK to the sender. Thus congestion information about the ACKs are not available at the sender.

If the X bit is set, all three other bit (L, E, C) MAY be set. When ever one of this bits is set, the number of bytes carried by this IP packet (incl. IP header) SHOULD be accounted when determining congestion or credit information. In IPv6 the length can easily be calculated by the value given in the Payload Length header field

(payload length + option space) plus a fixed value of 40 Bytes for the IP header itself.

In principle all of these three bits (L, E, C) MAY be set in the same packet. In this case the packet size MUST be accounted more than once for each respective ConEx information counter. In many cases if congestion occurs the sender will not sent additional credit, but if e.g. a sender assumes losses because of an audit function or needs to maintain a certain sending rate to make an application layer service work, the occurrence of credit bits (c) in parallel to congestion exposure bit (L, E) is reasonable.

If a network node extracts the ConEx information from a connection, this node is usually supposed to hold this information byte-wise, e.g. comparing the total number of bytes sent with the number of bytes sent with ConEx congestion mark (L, E) to determine the current whole path congestion level. When equally sized packets can be assumed, the accounting of the number of packets (instead the number of bytes) should deliver the same result. But a network node must be aware that this estimation can be quite wrong, if e.g. different sized packed are send, and thus is not reliable.

6. Implementation in the fast path of ConEx-aware routers

The conex information is being encoded into a destination option so that it does not impact forwarding performance in the non-conex-aware nodes on the path. Since destination options are not usually processed by routers, the existence of the CDO does not affect the fast path processing of the datagram on non-conex-aware routers. i.e. They are not pushed into the slow path towards the control plane for exception processing.

The conex-aware nodes still need to process the CDO without severely affecting forwarding. For this to be possible, the conex-aware routers need to quickly ascertain the presence of the CDO and process the option if it is present. To efficiently perform this, the CDO needs to be placed in a fairly deterministic location. In order to facilitate forwarding on conex-aware routers, conex-aware senders who send IPv6 datagrams with the CDO MUST place the CDO as the first destination option in the destination options header.

7. Compatiblility with use of IPsec

In IPsec transport mode no action need to be taken as the CDO is visible to the network. When accounting ConEx information the size of the Authentication Header (AH) SHOULD NOT be accounted as this

information has been added later. In the IPsec Tunnel model the CDO SHOULD be copied to the outer IP header as this information is end-to-end. Only the payload of the outer IP header minus the AH SHOULD be accounted.

8. Acknowledgements

The authors would like to thank Marcelo Bagnulo, Bob Briscoe, Ingemar Johansson, Joel Halpern, John Leslie and Wesley Eddy for the discussions that led to this document.

9. Security Considerations

This document does not bring up any new security issues.

10. IANA Considerations

This document defines a new IPv6 destination option for carrying conex markings. IANA is requested to assign a new destination option type in the Destination Options registry maintained at http://www.iana.org/assignments/ipv6-parameters <TBA1> Conex Destination Option [RFCXXXX] The act bits for this option need to be 10 and the chg bit needs to be 0.

11. Normative References

[CAM] Briscoe, B., "Congestion Exposure (ConEx) Concepts and Abstract Mechanism", draft-ietf-conex-abstract-mech-01 (work in progress), March 2011.

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

[RFC2460] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", RFC 2460, December 1998.

Authors' Addresses

Suresh Krishnan
Ericsson
8400 Blvd Decarie
Town of Mount Royal, Quebec
Canada

Email: suresh.krishnan@ericsson.com

Mirja Kuehlewind
IKR University of Stuttgart

Email: mirja.kuehlewind@ikr.uni-stuttgart.de

Carlos Ralli Ucendo
Telefonica

Email: ralli@tid.es