ConEx Working Group
Internet-Draft
Intended status: Standards Track
Expires: August 18, 2014

S. Krishnan
Ericsson
M. Kuehlewind
IKR University of Stuttgart
C. Ucendo
Telefonica

February 14, 2014

IPv6 Destination Option for ConEx
draft-ietf-conex-destopt-06

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.

Status of This Memo

This Internet-Draft is submitted 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 August 18, 2014.

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.

The ConEx information can be used by any network element on the path to e.g. do traffic management or egress policing. Additionally this information will potentially be used by an audit function that checks the integrity of the sender's signaling.

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. Requirements for the coding of ConEx in IPv6

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 IPv6 header bits (from the flow label), and new extension headers. After evaluating the different solutions, the wg concluded that only the use of a destination option would fulfil the requirements.

4. 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).

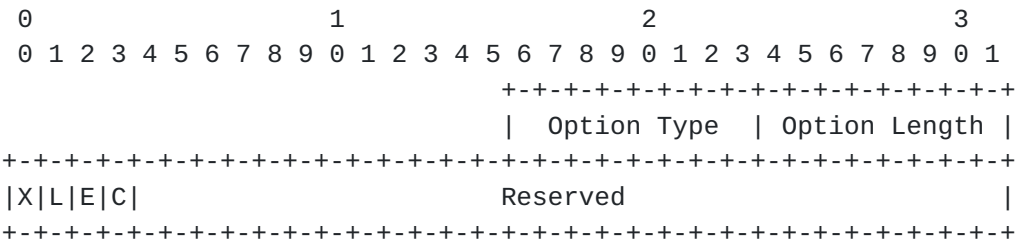

Figure 1: ConEx Destination Option Layout

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 not set, the sender is not using ConEx with this packet.

L Bit

When this bit is set, the transport sender has experienced a loss.

E Bit

When this bit is set, the transport sender has experienced ECN-signaled congestion.

C Bit

When this bit is set, the transport sender is building up congestion credit in the audit.

Reserved

These bits are not used in the current specification. They are set to zero on the sender and are ignored on the receiver.

All packets sent over a ConEx-capable connection MUST carry the CDO. The CDO is immutable. Network devices SHOULD only read the flags. IPSeC Authentication Header (AH) may be used to verify that the CDO has not been modified.

If the X bit is zero all other three bits are undefined and thus should be ignored. The X bit set to zero 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 ACKs are not available at the sender.

If the X bit is set, all three other bit (L, E, C) MAY be set. Whenever one of these bits is set, the number of bytes carried by this IP packet (including the IP header) SHOULD be accounted for determining congestion or credit information. In IPv6 the number of bytes can easily be calculated by adding the number 40 (length of the IPv6 header in bytes) to the value present in the Payload Length field in the IPv6 header.

Credits are sent previous to the occurence of congestion (loss or ECN-CE marks) and the amount of credits should cover the congestion risk. Note, the maximum congestion risk is that all packets in flight get lost or ECN marked.

If the L or E bit is set, a congestion signal in form of loss or, respectively, an ECN mark was previously expirienced by the same connection.

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.

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. For ConEx-aware node processing, the CDO MUST use the Payload length field of the preceding IPv6 header for byte-based accounting. 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.

A ConEx sender SHOULD set the reserved bits in the CDO to zero. Other nodes SHOULD not interpret these bits.

5. 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.

6. Compatibility with use of IPsec

In IPsec transport mode no action needs 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.

If the transport network can not be trusted authentication SHOULD be used to ensure integrity of the ConEx information. If an attacker would be able to remove the ConEx marks, this could cause an audit device to penalize the respective connection, while the sender cannot easily detect that ConEx information is missing.

7. DDoS mitigation by using preferential drop

If a router queue experiences very high load so that it has to drop arriving packets, it MAY preferentially drop packets within the same Diffserv PHB using the preference order given in Table 1 (1 means drop first). Additionally, if a router implements preferential drop it SHOULD also support ECN-marking. Preferential dropping can be difficult to implement on some hardware, but if feasible it would discriminate against attack traffic if done as part of the overall policing framework as described in [RFC6789]. If nowhere else, routers at the egress of a network SHOULD implement preferential drop (stronger than the MAY above).

	Preference
Not-ConEx or no CDO	1 (drop first)
X (but not L,E or C)	2
X and L,E or C	3

Table 1: Drop preference for ConEx packets

A flooding attack is inherently about congestion of a resource. As load focuses on a victim, upstream queues grow, requiring honest sources to pre-load packets with a higher fraction of ConEx-marks.

If ECN marking is supported by the downstream queues preferential dropping provides the most benefits because if the queue is so congested that it drops traffic, it will be CE-marking 100% of the forwarded traffic. Honest sources will therefore be sending 100% ConEx E-marked packets (and therefore being rate-limited at an ingress policer). Senders under malicious control can either do the same as honest sources, and be rate-limited at ingress, or they can understate congestion. If the preferential drop ranking is implemented on queues, these queues will preserve E/L-marked traffic until last. So, the traffic from malicious sources will all be automatically dropped first. Either way, the malicious sources cannot send more than honest sources.

8. Acknowledgements

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

Special thanks to Bob Briscoe who contributed text and analysis work on preferential dropping.

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] Mathis, M. and B. Briscoe, "Congestion Exposure (ConEx) Concepts and Abstract Mechanism", draft-ietf-ConEx-abstract-mech-05 (work in progress), July 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.

[RFC6789] Briscoe, B., Woundy, R., and A. Cooper, "Congestion Exposure (ConEx) Concepts and Use Cases", RFC 6789, December 2012.

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