conex Working Group	S.K. Krishnan
Internet-Draft	Ericsson
Intended status: Standards Track	M.K. Kuehlewind
Expires: May 02, 2012	IKR University of Stuttgart
	C.R.U. Ucendo
	Telefonica
	October 30, 2011

IPv6 Destination Option for Conex
draft-ietf-conex-destopt-01

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

Copyright Notice

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. 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 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                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

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.

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 MUST be zero as well. 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. An audit function MUST be aware of this possibility and SHOULD ensure that not a large amount of data is sent as not-ConEx capable with a ConEx capable connection.
If the X bit is set, all three other bit (L, E, C) MAY be set. When ever one if 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 ca 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 practice loss and ECN marks can not occur at the same time, so there should usually be a way to signal the respective ConEx information in different packets. In many cases if congestion occurs the sender will not sent additional credit bit, 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 accounting the number of packets (and comparing the total number to marked once) should deliver the same result. But a network node MUST be aware that this estimation can be quite wrong and thus is not reliable if e.g. different sized packed are send.

5. 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.

6. Security Considerations

This document does not bring up any new security issues.

7. 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.

8. References

[RFC2119]	Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
[RFC2460]	Deering, S.E. and R.M. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", RFC 2460, December 1998.
[CAM]	Briscoe, B, "Congestion Exposure (ConEx) Concepts and Abstract Mechanism", Internet-Draft draft-ietf-conex-abstract-mech-01, March 2011.

Authors' Addresses

Suresh Krishnan Krishnan Ericsson 8400 Blvd Decarie Town of Mount Royal, Quebec Canada EMail: suresh.krishnan@ericsson.com

Mirja Kuehlewind Kuehlewind IKR University of Stuttgart EMail: mirja.kuehlewind@ikr.uni-stuttgart.de

Carlos Ralli Ucendo Ucendo Telefonica EMail: ralli@tid.es