Workgroup: anima Working Group
Internet-Draft:
draft-richardson-anima-l2-friendly-acp-03
Published: 11 July 2022
Intended Status: Standards Track
Expires: 12 January 2023
Authors: M. Richardson
Sandelman Software Works
W. Pan
Huawei Technologies

Autonomic Control Plane design for Layer-Two Switched Networks

Abstract

This document proposes a design for an L2 aware Autonomic Control Plane that can be deployed easily to layer-two (Ethernet) switched technologies that are common on Campus/Enterprise network architectures.

This document leverages the hop-by-hop announcement used in LLDP, but runs bulk data over normal IPv6 Link-Local unicast ethernet frames.

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 https://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 12 January 2023.

Table of Contents

1. Introduction

The creation and maintenance of the Autonomic Control Plane described in [RFC8994] requires creation of hop-by-hop discovery of adjacent systems. There are Campus L2 systems that are not broadcast safe until they have been connected to their Software Defined Networking (SDN) controller. The use of the stable connectivity provided by [RFC8368] can provide the SDN connectivity required.

There is a bootstrap interlocking problem: the network may be unsafe for ACP discovery broadcasts without the support of Spanning Tree Protocol (STP) or similar mechanisms until configured, yet it can not be automatically configured until the ACP discovery (and onboarding process) is done. Meantime, because of STP complicated topological calculations, the convergence can be very slow for larger networks. This can delay on-boarding.

In addition, forming a campus-wide network by default and using enabling STP does not work. STP is not secure and could be easily spoofed by malicious or untrusted devices. On manually configured networks today, STP is turned off on "access" ports, and enabled only for trunk ports. But in an autonomic network, it is not possible to know a-priori which ports will be trunk ports, so STP would have to be on by default if it is was to be used.

What is needed is a way to send IPv6 traffic between these L2 switching devices in a way that is never forwarded, regardless of how the network is eventually configured. This is not just an inital

configuration problem: devices may be added and removed at any time, due to needed expansion of capacity, planned upgrades, or devices failures.

This document proposes using LLDP for what it is good at: announcing capabilities, while using normal EtherType 0x86DD IPv6 frames for the normal ACP transport.

1.1. Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

2. Protocol

A new TLV for LLDP is allocated and called the GRASP-DULL. The contents of the new TLV are the payload of the normal [RFC8994] GRASP DULL M_FLOOD, AN_ACP message.

The LLDP subsystem in the control plane CPU needs to forward these messages along to the ACP GRASP daemon, and it needs to also include the source MAC address (and port number) from which the LLDP message was received.

The ACP GRASP daemon can see the origin IPv6 Link-Local address from the GRASP DULL packet, and can now create an IPv6 neighbour cache entry (NCE) for that combination. By forcing this NCE entry, the node avoids the need to do an unsafe multicast IPv6 Neighbor Discovery.

The node SHOULD unicast a Neighbor Advertisement to the corresponding node to establish that node's NCE.

At this point it is possible to initiate the right key management daemon (IKEv2, etc.) using unicast IPv6 datagrams that only need unicast Ethernet packets.

It is likely that many L2 switching fabrics may not support IPsec ESP, or L3 routing. It was always the case that the ACP might have to be implemented as a software fabric in a control plane CPU. This is not a significant hurdle, as the ACP is not intended to be used for customer data, only control plane communication, and often only as a last resort.

In addition to normal operation, devices may need to be onboarded. [RFC8995] section 4.1.1 defines the AN_PROXY message to be used for

a new pledge to discover which neighbors are willing to act as onboarding proxies.

This M_FLOOD message will fit into the same GRASP DULL M_FLOOD message that contains the AN_ACP message.

After discover of an eligible neighbour, onboarding proceeds with a TCP connection over IPv6 link-local addresses, using unicast Ethernet frames.

LLDP traffic also uses a destination multicast address: 01:80:c2:00:00:0e, 01:80:c2:00:00:03, or 01:80:c2:00:00:00. The use of this destination address facilitiates transmission of the traffic through unmanaged switches ("dumb ethernet switches"), as well as allowing for seperation of provider and customer traffic in provider bridged (IEEE 802.1ad) situations.

3. Other constraints

On broadcast unsafe L2 networks, IPv6 Duplicate Address Detection (DAD) MUST be turned off. Only auto-configured IPv6 link-local addresses using SLAAC or stable-IID [RFC7217] may be used.

A pledge that is in an L2 network that is broadcast unsafe MUST NOT do mDNS queries as described in [RFC8995] appendix B.

4. Privacy Considerations

The LLDP messages commonly contain information that uniquely identifies a specific piece of switching equipment. The addition of the GRASP DULL message will also now reveal the link-local IPv6 addresses of the device. This additional information is either derived from ethernet addresses (so no new information), or will be derived using [RFC7217].

5. Security Considerations

Unclear as yet.

6. IANA Considerations

IANA is asked to allocate a TLV from the "IANA Link Layer Discovery Protocol (LLDP) TLV Subtypes" https://www.iana.org/assignments/ieee-802-numbers/ieee-802-numbers.xhtml#iana-lldp-tlv-subtypes

for the GRASP DULL L2 announcement.

7. Acknowledgements

Paul Congdon was very helpful in understanding how LLDP was actually processed in production equipment.

8. Changelog

1. A specific LLDP method for announcement using normal IPv6 datagrams described.

2. Document renamed, focus changed.

9. References

9.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <https://www.rfc-editor.org/info/rfc2119>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

[RFC8994] Eckert, T., Ed., Behringer, M., Ed., and S. Bjarnason, "An Autonomic Control Plane (ACP)", RFC 8994, DOI 10.17487/RFC8994, May 2021, <https://www.rfc-editor.org/info/rfc8994>.

[RFC8995] Pritikin, M., Richardson, M., Eckert, T., Behringer, M., and K. Watsen, "Bootstrapping Remote Secure Key Infrastructure (BRSKI)", RFC 8995, DOI 10.17487/RFC8995, May 2021, <https://www.rfc-editor.org/info/rfc8995>.

9.2. Informative References

[RFC7217] Gont, F., "A Method for Generating Semantically Opaque Interface Identifiers with IPv6 Stateless Address Autoconfiguration (SLAAC)", RFC 7217, DOI 10.17487/RFC7217, April 2014, <https://www.rfc-editor.org/info/rfc7217>.

[RFC8368] Eckert, T., Ed. and M. Behringer, "Using an Autonomic Control Plane for Stable Connectivity of Network Operations, Administration, and Maintenance (OAM)", RFC 8368, DOI 10.17487/RFC8368, May 2018, <https://www.rfc-editor.org/info/rfc8368>.

Authors' Addresses

Michael Richardson
Sandelman Software Works

Email: mcr+ietf@sandelman.ca

Wei Pan
Huawei Technologies

Email: william.panwei@huawei.com