Layer 2 Gateway (L2GW)  
draft-xia-nvo3-l2gw-00

Abstract

Layer 2 Gateway (L2GW) is used for interconnecting layer 2 overlay network [NVO3FRWK] with layer 2 bridge networks [IEEE802.1Q] to form one layer 2 virtual network. This draft discusses which Layer 2 Control Protocol (L2CP) specified in IEEE802.1 should be supported by the layer 2 overlay network and which not, and specifies how L2GW should deal with L2CP frames.

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), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt.

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html.

This Internet-Draft will expire on September 14, 2014.

Copyright Notice

Copyright (c) 2013 IETF Trust and the persons identified as the document authors. All rights reserved.
1. Introduction

Cloud computing and network virtualization is evolving to the direction of virtualized networks in overlay, which aims fast and easy to create tenant networks, support tenant system mobility, and bring the manageability of all virtualized resources in DC.

Layer 2 (L2) overlay network in NVO3 means an L2 overlay network interconnecting tenant systems, where any pair of Network Virtualization Edges (NVE) are connected by IP tunnels. As a result, it forms a full mesh topology of overlay network, i.e. only one hop between any pair of NVEs. L2 bridge network in this draft is the network specified in IEEE 802.1Q [IEEE 802.1Q] which are widely deployed in DCs. L2 overlay network is used to refer the L2 network specified in NVO3.

During DC network migration, it’s very common that L2 overlay network may be mix used with L2 bridge network in a DC, and the communication between them is required. In another use case, using L2 bridge network to connect non-virtualized devices in DC are mature and well deployed method in DC; these non-virtualized devices are necessary for some applications such as BIG data and are required to communicate with virtualized resources.

To interconnect an L2 overlay network with an L2 bridge network, gateway functions are needed on the device(s)/system(s) connecting two networks. This device is referred as to Layer 2 Gateway (L2GW) in this draft. L2GW processes the encapsulation translation of packets between overlay technologies (e.g., VxLAN [VXLAN], NVGRE [NVGRE]) and the technologies specified in IEEE 802.1Q; it also deals with Layer 2 Control Protocol (L2CP) frames from the bridge network.

L2CP frames are defined by IEEE802.1 to be used for L2 network control, e.g., STP, LACP, etc. An L2CP is identified by one of the following MAC destination addresses:

- 01-80-C2-00-00-00 through 01-80-C2-00-00-0F: Bridge Block of protocols
- 01-80-C2-00-20 through 01-80-C2-00-00-2F: GARP/MRP Block of protocols

All L2CPs are supported in a L2 bridge network, every 802.1Q bridge performs corresponding action, i.e., peer (to be processed by local protocol entity), discard, pass, based on L2CP frame’s MAC destination address (DA) and protocol identifier (Ethertype or LLC).
Address). For L2 overlay VN, most of L2CPs are unnecessary because L2 overlay VN has its own control plane functions to support needed L2 communication such as transport over a tunnel or OAM. It is very useful to document how these service frames should be handled at L2GW to ensure that two networks can interwork.

This draft discusses which L2CP should be supported by L2 overlay network and which not, and specifies how L2GW should deal with L2CP frames.

1.1. 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 RFC-2119 [RFC2119].

1.2. Terminology

This document uses the terms defined in NVO3 framework [NVO3FRWK] and architecture [NVO3ARCH] documents.

2. L2GW Reference Model

The following diagram shows a reference model where L2GW provides an interconnection between L2 overlay network and L2 bridge network.

Figure 1: L2GW Reference Model

L2GW can reside on access switch providing direct connection with physical server, or reside on core switch or edge router providing connection with L2 bridge network.

Note that this draft only addresses the case that L2GW and L2 NVE are co-located at the edge device between two L2 networks, where L2GW can deal with the interworking work. Other cases are not yet covered in this draft.
### 3. L2CP Review and Applicability to L2 Overlay Network

L2CP protocols defined in IEEE 802.1 are listed in Table 1:

<table>
<thead>
<tr>
<th>MAC DA</th>
<th>Assignment</th>
<th>Protocol Type</th>
<th>L2CP Action</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>VLAN-based L2 services PORT-based L2 services</td>
</tr>
<tr>
<td>01-80-C2-00-00-00</td>
<td>Nearest Customer Bridge</td>
<td>STP/RSTP/MS TP, LACP/LAMP</td>
<td>Filter Pass</td>
</tr>
<tr>
<td>01-80-C2-00-00-01</td>
<td>IEEE MAC Specific Control Protocols</td>
<td>PAUSE</td>
<td>Filter Filter</td>
</tr>
<tr>
<td>01-80-C2-00-00-02</td>
<td>IEEE 802 Slow Protocols</td>
<td>LACP/LAMP, Link OAM, ESMC</td>
<td>Filter Filter</td>
</tr>
<tr>
<td>01-80-C2-00-00-03</td>
<td>Nearest non-TPRM Bridge</td>
<td>Port Authentication, LACP/LAMP</td>
<td>Filter Filter</td>
</tr>
<tr>
<td>01-80-C2-00-00-04</td>
<td>IEEE MAC Specific Control Protocols</td>
<td></td>
<td>Filter Filter</td>
</tr>
<tr>
<td>01-80-C2-00-00-05</td>
<td>Reserved for Future Standardization</td>
<td></td>
<td>Filter Filter</td>
</tr>
<tr>
<td>01-80-C2-00-00-06</td>
<td></td>
<td></td>
<td>Filter Filter</td>
</tr>
<tr>
<td>01-80-C2-00-00-09</td>
<td></td>
<td></td>
<td>Filter Filter</td>
</tr>
<tr>
<td>01-80-C2-00-00-A</td>
<td></td>
<td></td>
<td>Filter Filter</td>
</tr>
<tr>
<td>MAC Address</td>
<td>Protocol Type</td>
<td>Function</td>
<td>Filter</td>
</tr>
<tr>
<td>-------------</td>
<td>---------------</td>
<td>-----------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>01-80-C2-00-00-07</td>
<td>MEF ELMI</td>
<td>E-LMI</td>
<td>Filter</td>
</tr>
<tr>
<td>01-80-C2-00-00-08</td>
<td>Provide Bridge Group</td>
<td></td>
<td>Filter</td>
</tr>
<tr>
<td>01-80-C2-00-00-0B</td>
<td>Reserved for Future Standardization</td>
<td></td>
<td>Filter</td>
</tr>
<tr>
<td>01-80-C2-00-00-0C</td>
<td>Standardization</td>
<td></td>
<td>Filter</td>
</tr>
<tr>
<td>01-80-C2-00-00-0D</td>
<td>Provider Bridge MVRP</td>
<td></td>
<td>Filter</td>
</tr>
<tr>
<td>01-80-C2-00-00-0E</td>
<td>Nearest Bridge, Individual LLDP, PTP Peer Delay</td>
<td></td>
<td>Filter</td>
</tr>
<tr>
<td>01-80-C2-00-00-20</td>
<td>through GARP/MRP Block</td>
<td></td>
<td>Pass</td>
</tr>
<tr>
<td>01-80-C2-00-00-2F</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1 L2CP protocols specification

Note:

Different L2CP protocols can use the same MAC DA in above block of 32 addresses, but be differentiated by protocol identifier. MAC DA determines the intended recipient device for the frame;

Filter represent the L2CP action of peer or discard;

Based on whether L2 interface is VLAN-aware, L2 services can divided into two categories: VLAN-based L2 services, PORT-based L2 services. L2CP action (peer, discard, pass) for these two L2 services is also different;
Whether the L2CP frames are peered or discarded is further determined by the configuration of L2 interface.

Further analysis about whether a L2CP protocol is necessary and how it is processed in NVO3 supported L2 VN, is provided in the following sub sections.

### 3.1. STP/RSTP/MSTP

The Spanning Tree Protocol (STP) is a L2 protocol that ensures a loop-free topology for any bridged Ethernet local area network. The basic function of STP is to prevent bridge loops and the broadcast storm that results from them. Rapid spanning Tree Protocol (RSTP) and Multiple Spanning Tree Protocol (MSTP) are all the enhanced xSTP protocols.

L2 overlay network does not need xSTP protocols to prevent bridge loops because it has its own mechanism for it, i.e., NVA, control plane mechanisms, full mesh + split horizon, etc. So, the process of xSTP frames in L2 VN is:

Be in line with L2CP protocols’ specification of Table 1 from IEEE in the L2 sub-networks attached to L2 NVEs;

xSTP frames are filtered in L2 NVEs and should not go into L2 overlay network.

### 3.2. PAUSE

[IEEE 802.3-2005] has specified a L2 flow control mechanism through using the PAUSE frame. This frame uses L2CP MAC DA of 01-80-C2-00-00-01 to be sent to the node at the other end of the link for informing it to halt the frame transmission for a specified period of time.

When L2 NVE is co-located in Hypervisor, PAUSE frame is not necessary in one device. When they are separated, PAUSE frame is only used in layer 2 network between L2 NVE and Hypervisor, there is no need to overlay PAUSE frame between L2 NVEs.

### 3.3. LACP/LAMP

Link Aggregation [IEEE 802.1AXbk-2012] is a mechanism for making multiple point-to-point links between a pair of devices appear to be a single logical link between those devices. Link Aggregation
Control Protocol (LACP) and Link Marker Control Protocol (LAMP) operate between exactly two peer devices for the purpose of creating, verifying, and monitoring the logical link created by aggregating individual links. Specific L2CP frames, known as Link Aggregation Control Protocol Data Units (LACPDUs), are exchanged between the peer devices on each individual link in the aggregation. The protocol identifier used by LACP is an Ethertype with a value of 0x8809 (the “Slow Protocols” Ethertype) and subtype values 01 (for LACP) and 02 (for LAMP). Note that LACP is used to represent LACP and LAMP in the following text.

LACP uses 3 different L2CP MAC DAs to determine the scope of propagation of LACPDUs within a bridged LAN, as Table 2 follows:

<table>
<thead>
<tr>
<th>Assignment</th>
<th>L2CP MAC DA</th>
<th>Peered or discarded by</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nearest Customer</td>
<td>01-80-C2-00-00-00</td>
<td>End Station, Customer Bridge, Provider Edge Bridge</td>
</tr>
<tr>
<td>Bridge</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IEEE 802 Slow</td>
<td>01-80-C2-00-00-02</td>
<td>End Station, Customer Bridge, Provider Edge Bridge, Provider Bridge</td>
</tr>
<tr>
<td>Protocols</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nearest non-TPRM</td>
<td>01-80-C2-00-00-03</td>
<td>Bridges except for Two Port MAC Relay</td>
</tr>
<tr>
<td>Bridge</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2 LACP specification of L2CP MAC DAs

Base on the summary of Table 2, LACPDUs with the L2CP MAC DA of 01-80-C2-00-00-02 are peered or discarded by every node, so this kind of LACPDUs will not be overlaid across the L2 overlay network. For 01-80-C2-00-00-00, it is possible that LACPDUs need to be overlaid across Provider Bridge and L2 NVEs of L2 overlay network to reach the other end Custom Bridge, L2 overlay network maybe need to support to overlay this kind of LACP frame between L2 NVEs. How the L2 overlay network support LACP frame of 01-80-C2-00-00-03 is TBD.

3.4. Link OAM

Lin OAM defined is defined in [IEEE 802.3ah], as mechanisms for monitoring and troubleshooting Ethernet access links. Specifically it defines tools for discovery, remote failure indication, remote and local loopbacks and status and performance monitoring.
The Link OAM frames using L2CP MAC DA of 01-80-C2-00-00-02 are peered or discarded by every node, so this kind of frame will not be overlaid across the L2 overlay network.

3.5. Port Authentication

[IEEE 802.1X] is an IEEE Standard for Port-based Network Access Control (PNAC). It is part of the IEEE 802.1 group of networking protocols. It provides an authentication mechanism to devices wishing to attach to a LAN or WLAN.

Whether or not the L2 overlay network needs to overlay this L2CP frames is TBD.

3.6. E-LMI

Ethernet Local Management Interface (E-LMI) is a protocol between the customer edge (CE) device and the provider edge (PE) device. It runs only on the PE-CE UNI link and notifies the CE of connectivity status and configuration parameters of Ethernet services available on the CE port. E-LMI interoperates with an OAM protocol, such as Connectivity Fault Management (CFM), that runs within the provider network to collect OAM status. CFM runs at the provider maintenance level (UPE to UPE with inward-facing MEPs at the UNI). E-LMI relies on the OAM Ethernet Infrastructure (EI) to interwork with CFM for end-to-end status of Ethernet virtual connections (EVCs) across CFM domains.

The LLDP frames using L2CP MAC DA of 01-80-C2-00-00-07 are peered or discarded by every node except for the Two Port MAC Relay (TPMR) bridge, so this kind of frame will not be overlaid across the L2 overlay network.

3.7. LLDP

The Link Layer Discovery Protocol (LLDP) is a vendor-neutral link layer protocol in the Internet Protocol Suite used by network devices for advertising their identity, capabilities, and neighbors on an IEEE 802 local area network, principally wired Ethernet. The protocol is formally referred to by the IEEE as Station and Media Access Control Connectivity Discovery specified in standards document [IEEE 802.1AB].

The LLDP frames using L2CP MAC DA of 01-80-C2-00-00-0E are peered or discarded by every node, so this kind of frame will not be overlaid across the L2 overlay network.
3.8. PTP Peer Delay

PTP Peer Delay frame is specified in [IEEE 1588-2008] to carry PTP peer time information. It uses L2CP MAC DA of 01-80-C2-00-00-0E and peered or discarded by every node, so this kind of frame will not be overlaid across the L2 overlay network.

3.9. ESMC

Ethernet Synchronization Messaging Channel (ESMC) is specified in [ITU-T Rec. G.8264] for conveying clock information between Synchronous Ethernet (SyncE) bridges.

The ESMC frames using L2CP MAC DA of 01-80-C2-00-00-02 are peered or discarded by every node, so this kind of frame will not be overlaid across the L2 overlay network.

3.10. GARP/MRP Block

Multiple Registration Protocol (MRP), which replaced Generic Attribute Registration Protocol (GARP), is a generic registration framework defined by the [IEEE 802.1ak] amendment to the [IEEE 802.1Q] standard. MRP allows bridges, switches or other similar devices to be able to register and de-register attribute values, such as VLAN identifiers and multicast group membership across a large LAN. MRP operates at the Data Link Layer.

The block of L2CP MAC DA from 01-80-C2-00-00-20 to 01-80-C2-00-00-2F is used for MRP protocol. Now, only 01-80-C2-00-00-20 is for Multiple MAC Registration Protocol (MMPR) and 01-80-C2-00-00-21 is for Multiple VLAN Registration Protocol (MVRP), other L2CP MAC DA of the block are all reserved for future use. Protocol use one address of this block is passed by all the intervening bridges that does not participate in the protocol using this address, and peered or discarded by the bridge that participate in the protocol at last. This forwarding rule maybe requires MRP frames to be overlaid across the L2 overlay network.

4. L2CP Process in L2GW

For all L2CP protocols, several differences exist between L2 overlay network and L2 bridge network on how to process them. As the demarcation point between L2 overlay network and L2 bridge network, L2GW keeps the same action to all L2CP frames as before at the L2 bridge network side on the one hand, but maybe processes some L2CP frames differently at the L2 overlay network side on the other hand. The following sub sections will describe the L2CP process in L2GW.
4.1. L2CP Frames Filtered (Peered or Discarded) in L2GW

Although xSTP protocols using Nearest Customer Bridge address of 01-80-C2-00-00-00 indicate that it can be overlaid across L2 overlay network, they still are not necessary for L2 overlay network because L2 overlay network has its own mechanism to prevent bridge loops. So xSTP frames will be filtered by the L2GW and not go into the L2 overlay network.

Based on the analysis of section 3.3, LACP/LAMP frames using IEEE 802 Slow Protocols of 01-80-C2-00-00-02 are not necessary for L2 overlay network. So, LACP/LAMP frames will be filtered by the L2GW and not go into the L2 overlay network. ESMC frames using the same MAC DA will also be filtered by L2GW.

For Link OAM frames, if OAM functions are necessary for the whole L2 network which interconnects L2 bridge network and L2 overlay network, L2GW needs to support the interworking of OAM as well. This means that L2GW should peer the Link OAM frames of L2 bridge network and perform some actions between NVEs in L2 overlay network. The detailed operation is TBD.

Other L2CP protocols that are filtered by L2GW and do not go into L2 overlay network include PAUSE, E-LMI, LLDP, PTP Peer Delay. The basic reason is that they all require to be processed hop by hop in L2 network strictly, but overlay network breaks this rule.

The action of ‘’filter’’ can be ‘’peer’’, or ‘’discard’’. It depends on the specific service requirement, i.e., does L2GW need to participate in the L2CP protocol, etc. How to determine the specific action is TBD.

4.2. L2CP Frames Passed through L2GW

Excepting for the aforementioned L2CP protocols filtered by L2GW, the left L2CP protocols need to be passed through L2GW. They include:

- LACP/LAMP frames using IEEE 802 Slow Protocols of 01-80-C2-00-00-00;
- GARP/MRP series protocols (i.e., MMRP, MVRP) using the MAC DA block of 01-80-C2-00-00-20 through 01-80-C2-00-00-2F.

All these kinds of L2CP frames are passed through L2GW and traverse across the L2 overlay network and L2 bridge network to arrive the bridges that participate in the L2CP protocols. For MRP protocols, another necessary operation of L2GW is to use the pre-provisioned
VLAN to virtual network instance (VNI) mappings in NVE locally or by getting from NVA to map these MRP frames into corresponding VNIs.

5. Other Interworking Cases

There are other L2 bridge network technologies that use L2 Control Plane protocols such as Provider Bridge [IEEE802.1AD] or Provider Backbone Bridge [PBB] [IEEE802.1AH]. The use case of L2 Overlay Network interworking with these types of bridge networks is for the further study.

Note that VPLS [RFC4761] [RFC4762], EVPN [EVPN], Shortest Path Bridging [IEEE SPB] and TRILL [RFC6325] are also technologies for L2 private network implementation. These technologies rely on the control plane protocol and aim for service provider network. SDN controller interworking with such control plane protocol will be addressed in separate draft.

6. Security Considerations

TBD.

7. IANA Considerations

The document does not require any IANA action.

8. References

8.1. Normative References


8.2. Informative References

[NVO3ARCH] Black, D., Narten, T., et al, "An Architecture for Overlay Networks (NVO3)", draft-narten-nvo3-arch-01, work in progress


[VXLAN] Mahalingam, M., Dutt, D., etc, "VXLAN: A Framework for Overlaying Virtualized Layer 2 Networks over Layer 3 Networks", draft-mahalingam-dutt-dcops-vxlan-05.txt, work in progress


[IEEE 802.1Q] "Virtual Bridged Local Area Networks", 2005

[IEEE 802.3-2005] "Part 3: Carrier sense multiple access with collision detection (CSMA/CD) access method and physical layer specifications"

[IEEE 802.1AXbk-2012] "IEEE Standard for Local and metropolitan area networks--Link Aggregation Amendment 1: Protocol Addressing"


[IEEE 802.1X] "IEEE Standard for Local and metropolitan Area Networks. Port-based Network Access Control"

[IEEE 802.1AB] "IEEE Standard for Station and Media Access Control, Connectivity Discovery"


[IEEE 802.1ak] "IEEE Standard for Local and metropolitan Area Networks - Virtual Bridged Local Area Networks, Amendment 7: Multiple Registration Protocol"
[IEEE 802.1AD], "Virtual Bridged Local Area Networks - Amendment 4: Provider Bridges", 2005


[ITU-T Rec. G.8264] "Distribution of Timing Through Packet Networks"

Authors’ Addresses

Liang Xia
Huawei Technologies

Email: frank.xialiang@huawei.com

Lucy Yong
Huawei Technologies, USA

Email: lucy.yong@huawei.com