IS-IS VPLS for Cloud Data Center Networks

draft-xu-l2vpn-vpls-isis-03

Xiaohu Xu (xuxh@huawei.com)

Himanshu Shah (hshah@ciena.com)

IETF83, Paris

Cloud Data Center Network Requirements

- **LAN Extension**
  - VM migration across multiple racks or pods within a data center.
  - Some cluster applications depending on link-local multicast for cluster member discovery and heartbeat.
- **VPN/Tenant Space Scalability**
  - Tens of thousands of tenants over a shared infrastructure.
- **Forwarding Table Scalability**
  - Millions of VMs within a data center.
- **Bandwidth Utilization Maximization**
  - ECMP.
  - Shortest path forwarding.

Data Centers can benefit from

- **ARP/Unknown Unicast Flood Suppression**
  - Reduce performance impact on networks and servers.
- **Flexibility for Tradeoffs between Bandwidth and State**
  - Each Tenants have different broadcast/multicast characteristics.
  - Tenants/VPN instances with fewer member PEs can use ingress replication while ones with a mass of member PEs benefit from multicast-tree based approaches
- **Simplified Provisioning and Operation**
  - Extend IP as close to the edge as possible (ToR being the ideal candidate).
  - Increased scale mandates keeping service provisioning as simple as possible.
- **Reuse Existing Operating Experiences**
  - Leveraging deployed protocols and the related experience to provision new services –> a great plus.

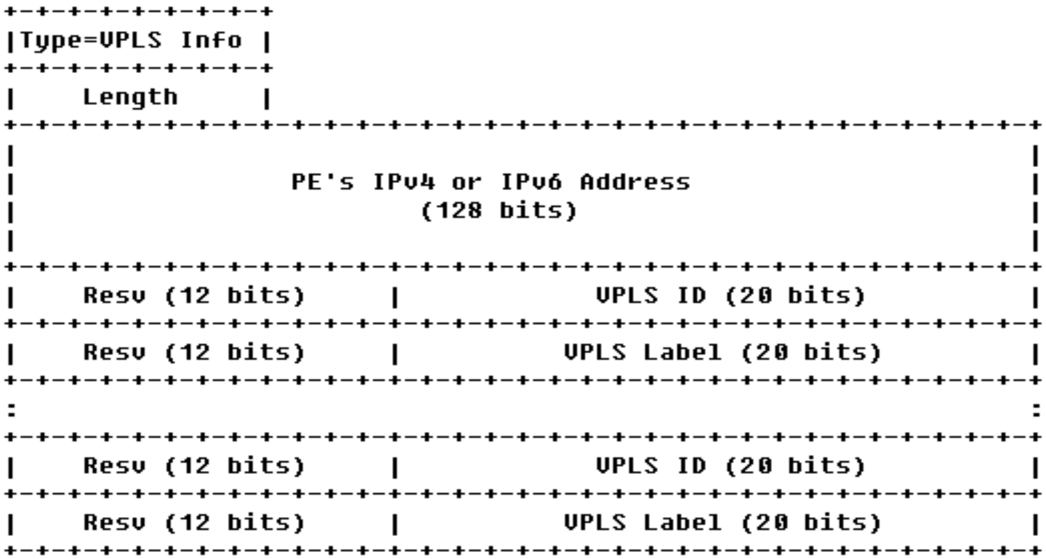
VPLS in Cloud Data Centers

- **VPLS technology provides a credible solution for data center networks.**
- **Following considerations apply to enhance the applicability**
  - Simplified auto-discovery.
  - Ease of service turn up, adding and deleting PE nodes.
  - Low touch, irrespective of the number of PEs in the service – should scale well.
  - Efficient and smart broadcast/multicast delivery.
    - Use ingress replication for those tenants with a few member PEs.
    - Use multicast tree for those tenants with large numbers of member PEs.

IS-IS VPLS at a Glance

- **Leverages the deployed IGP, IS-IS, with incremental extensions that provide auto-discovery as well as signaling of VPLS instance/tenant identifier/Virtual Network Identifiers.**
- **The data plane encapsulations adhere to what has already been defined – no change to forwarding procedures.**
- **The proposed solution make improvements while retaining advantages of the existing VPLS solutions**
  - No PW between PE routers => Scalable.
  - Both ingress replication and P-multicast tree are available=> Flexible.
  - No separate protocol for VPLS => Simple.

VPLS Info TLV



```
+-+-+-+-+-+-+-+-+
|Type=VPLS Info |
+-+-+-+-+-+-+-+-+
|    Length     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
|                  PE's IPv4 or IPv6 Address                    |
|                          (128 bits)                           |
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    Resv (12 bits)     |            VPLS ID (20 bits)          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    Resv (12 bits)     |          VPLS Label (20 bits)         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
:                                                               :
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    Resv (12 bits)     |            VPLS ID (20 bits)          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    Resv (12 bits)     |          VPLS Label (20 bits)         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

- **Auto-discovery and signaling functionalities are accomplished by propagating this TLV across PE routers.**
  - P routers do not process this TLV, but instead synchronizes the Link State PDUs (LSPs) with IS-IS neighbors as normal.
  - There is precedence in other solutions whereby IS-IS protocol is used to distribute non-IP specific information
  - Associated VPLS label is used to identify VPLS instance in the data plane.

Remote MAC Address Learning

- **Date-plane based MAC learning**
  - IP/GRE tunnel is used between PE routers to carry client payload
  - Ingress PE of the received VPLS packet could be identified according to tunnel source address.
- **Control-plane based MAC learning**
  - MAC-reachability TLV defined in [RFC6165] could be reused.
- **IS-IS VPLS allows for a flexible tradeoff between forwarding table state and unknown unicast suppression on a per tenant basis.**

Multicast/broadcast Delivery

- **Ingress replication**
  - MAC-in-MPLS-in-IP/GRE encapsulation.
  - VPLS label assigned by each egress PE is used here as a downstream-assigned label.
- **P-multicast tree**
  - MAC-in-MPLS-in-IP/GRE encapsulation.
  - VPLS label assigned by ingress PE is used here as an upstream-assigned label.
- The proposed solution offers a choice between ingress replication and use of multicast tree based broadcast propagation. A selection could be based on the tradeoff between bandwidth usage and multicast state maintenance on a per tenant basis.

Next Steps

- **A question for support of Active/Active multi-homing was asked during last presentation (IETF82)**
  - We discussed possible solutions internally and reached a conclusion that it is feasible
  - While active/active multi-homing would be a desirable option, it is not a mandatory criteria for all solutions, in order to be accepted as WG documents.
  - Active/Active multi-homing can be added as a section once this draft is adopted as a WG doc, or be submitted as a complementary draft.
- **WG adoption of this draft?**