Network working group
Internet Draft
Category: Informational

X. Xu
Huawei Technologies
Kai Lee
China Telecom

Expires: January 2013

July 9, 2012

Path Optimization for LAN Extension

draft-xu-nvo3-lan-extension-path-optimization-00

Status of this Memo

Copyright Notice

Abstract

This document describes path optimization issues caused by LAN extension across geographically dispersed data centers. In addition, this document also describes requirements for possible solutions to these issues.

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].

Table of Contents

1. Problem Statement

Virtual Machine (VM) migration and geo-clustering across data centers usually require a LAN to be extended across these data centers. Figure 1 depicts a generic data center interconnect architecture where multiple data centers are interconnected with a given LAN extension solution and remote VPN sites (e.g., cloud user sites) are connected to these data centers with L3VPN solution [RFC4364].

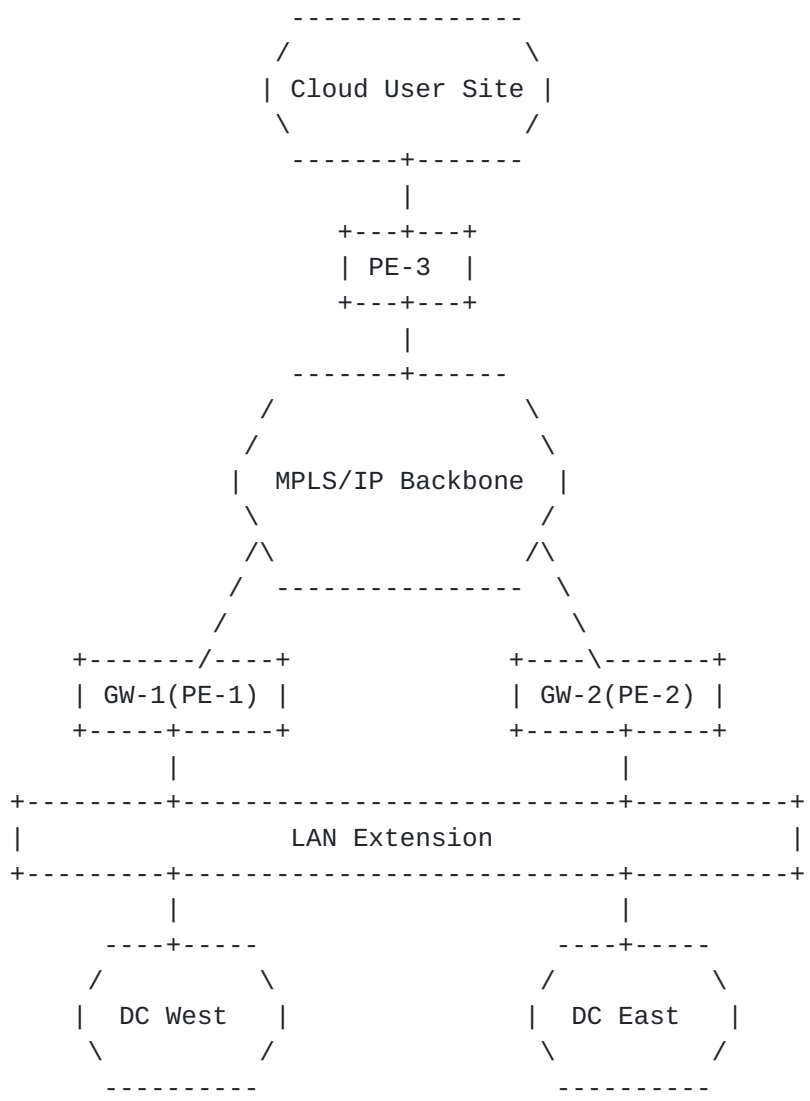

Figure 1: A Generic Data Center Interconnect Architecture

Since the LAN has been extended across multiple data center locations, the IP subnet associated with this LAN is also extended

across these locations. As such, the traffic to/from the extended subnet (e.g., the traffic between cloud user sites and data centers) would encounter suboptimal routing issues as described in the following sub-sections. Such suboptimal routing not only unnecessarily consumes the bandwidth intended for data center interconnect, but also decreases the cloud users' experiences due to increased path latency. Note that here the traffic to/from the extended subnet refers to L3VPN traffic between a remote L3VPN site (e.g., a cloud user site) and data centers, rather than Internet traffic. How to optimize the path for Internet traffic to/from the extended subnet would be explored in the future.

1.1. Suboptimal Routing for Incoming Traffic

Since an IP subnet has been extended across multiple locations, the subnet no longer retains its location semantics. As a result, the incoming traffic towards a given server within the extended subnet could travel through suboptimal paths if the traffic is forwarded based on the corresponding subnet route. For example, assume a server is physically located at data center East of an extended subnet, the incoming traffic towards that server would possibly travel through the default gateway router at data center West when entering that subnet.

1.2. Suboptimal Routing for Outgoing Traffic

Let's assume the existing VPLS solution [RFC4761, RFC4762] is used to achieve LAN extension across multiple data center locations. In this case, VRRP would usually be enabled on default gateway routers of different locations and only one of them would be selected as the VRRP Master for the subnet associated with the extended LAN, which is available for forwarding outgoing traffic of the subnet. In addition, although multiple default gateway routers of different locations could be selected as VRRP masters by filtering VRRP messages among them, since the existing VPLS solution however perform MAC learning as a traditional bridge, the route (e.g., MAC forwarding entry) for a given MAC address would be determined without taking the network distance into account. As a result, if the forwarding path to the VRRP virtual MAC is currently pointed to a default gateway router at data center East, for those servers located at data center West, their outgoing traffic would have to traverse the data center interconnection path so as to reach that default gateway router at data center East, which in turn forwards the traffic out of that subnet.

2. Terminology

This memo makes use of the terms defined in [RFC4364] and [RFC2338].

3. Solution Requirements

3.1. Path Optimization for Incoming Traffic

The basic idea is to allow each default gateway router acting as a L3VPN PE router to propagate host routes for local servers within the extended subnet to remote PE routers. More specifically, a default gateway router at a given data center is allowed to advertise hosts routes only for servers located in that data center, rather than those ones located in other data centers. In this way, remote PE routers would be able to forward traffic destined for a given server within the extended subnet according to the corresponding host route for that server, rather than the subnet route for that extended subnet.

The challenge here is how to make default gateway routers be able to tell which servers within the extended subnet are their local ones. Hence the possible solution for this path optimization issue SHOULD ensure default gateway routers to be able to obtain enough information so as to distinguish local servers from remote ones.

3.2. Path Optimization for Outgoing Traffic

To realize the purposes of default gateway redundancy and VM live mobility across data centers, default gateway routers of a given extended subnet at different locations SHOULD be configured with an identical virtual IP/MAC address pair (i.e., virtual router). As such, servers within the extended subnet could use that virtual router's IP address as their default gateway. To ensure the outgoing traffic with destination MAC address being the virtual router's MAC address to be forwarded to a local default gateway router, rather than any remote default gateway router, just like the anycast manner in IP networks, the LAN extension solution SHOULD be able to select the best route for a given MAC address (e.g., the virtual router's MAC address) among multiple possible routes, e.g., by taking network distance as one factor in the decision-making process of best-route selection.

4. Security Considerations

TBD.

5. IANA Considerations

There is no requirement for IANA.

6. Acknowledgements

TBD.

7. References

7.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

7.2. Informative References

[RFC2338] Knight, S., et al., "Virtual Router Redundancy Protocol", RFC 2338, April 1998.

[RFC4761] Kompella, K. and Y. Rekhter, "Virtual Private LAN Service (VPLS) Using BGP for Auto-Discovery and Signaling", RFC 4761, January 2007.

[RFC4762] Lasserre, M. and V. Kompella, "Virtual Private LAN Service (VPLS) Using Label Distribution Protocol (LDP) Signaling", RFC 4762, January 2007.

[RFC4364] Rosen, E. and Y. Rekhter, "BGP/MPLS IP Virtual Private Networks (VPNs)", RFC 4364, February 2006.

Authors' Addresses

Xiaohu Xu
Huawei Technologies,
Beijing, China.

Phone: +86 10 60610041
Email: xuxiaohu@huawei.com

Kai Lee
China Telecom,
Beijing, China.

Leekai@ctbri.com.cn