Operational Aspects of Virtual Private LAN Service

Kireeti Kompella
Agenda

1. Introduction to VPLS

Operational Issues
4. LAN over a MAN/WAN?
5. MAC Address Scaling
6. Full Mesh Connectivity
7. Loops and Spanning Tree
8. Inter-AS (Inter-Provider) VPLS
9. Deployment Status
1. Introduction to VPLS

- Typical Building/Campus Network
- Frame Relay (ATM) Connectivity
- Ethernet-based Connectivity
- Why Ethernet for External Connectivity?
- Why VPLS?

Summary: Multipoint Ethernet access is a service desired by many enterprises
Typical Building/Campus Network

- Intra-building connectivity via Ethernet
- Broadcast domains (LANs) broken up by routers
- External connectivity via a WAN link from a router
  - Primary theme of talk: WAN link replaced by Ethernet
Frame Relay (ATM) Connectivity

- Intra-building connectivity via Ethernet
- External connectivity via Frame Relay or ATM VCs
- Routing paradigm shift -- multiple point-to-point adjacencies instead of a single multi-point adjacency

SP network looks like a Frame Relay switch

Forwarding is based on VCs
Ethernet-based Connectivity

- Intra-building connectivity via Ethernet
- External connectivity via VPLS – just another Ethernet broadcast domain
- All customer routing is based on multi-point adjacencies over Ethernet; multicast is native Ethernet multicast
Why Ethernet for External Connectivity?

- Most networks inside buildings have Ethernet – this is the most common network connection
  - Ethernet is cheap, fast and simple
- Routing over an Ethernet is easier and more scalable than over N point-to-point links
  - For RIP, one can broadcast or multicast updates
  - For OSPF and IS-IS, form a single adjacency per LAN segment, send one hello and floods LSDB once
- Broadcast and multicast are simpler -- native operation with IGMP instead of PIM
- Native operation for non-IP Ethernet-based applications
Why VPLS (Not Native Ethernet)?

- “Network convergence” -- don’t want a separate network for Ethernet access
- Ethernet is an appealing *access* medium, but it makes a poor Service Provider *infrastructure*
  - Don’t want to carry all customer MAC addresses in every single device -- does not scale, violates privacy
  - Don’t want to run Spanning Tree in SP network
  - Cannot afford even transient layer 2 loops or broadcast storms
2. LAN Over a MAN/WAN?

- Can the SP network emulate an Ethernet well enough? Learn (and age) MAC addresses, flood packets, etc.?
- Will LAN applications work correctly over a MAN or WAN connection?
LAN Over a MAN/WAN?

- The answer to the first question is **absolutely**!
- The answer to the second question is less definite at present
  - This is a new service, and there isn’t enough deployment experience
    - However, many active deployments -- we’ll know soon
    - The attitude is, Ethernet/VPLS deployment and usage is inevitable, so **just make it work**!
  - No issues are anticipated with IP-based applications
  - The main issues are: latency and packet loss
    - These are known problems, and have good solutions
3. MAC Address Scaling

- Will the SP network be able to handle all the customer MAC addresses?
MAC Address Scaling

- The aim is *not* to build a single huge, world-spanning broadcast domain for each customer!
  - Even within a building, there are multiple LANs
- MAC address knowledge for a given VPLS is limited to the PEs participating in that VPLS
  - Analogy: RFC 2547bis IP VPNs
- MAC addresses are *not* exchanged among PEs by any protocol -- they are learned dynamically
- Initial deployments: restrict CE devices to routers, and thus limit the number of MACs
4. Full Mesh Connectivity

- Why do the PEs need to be fully meshed?
- How does one ensure this?
Full Mesh Connectivity

- All VPLS solutions require full mesh connectivity among PEs belonging to a particular VPLS
  - A partial mesh can lead to weird failure modes that are not easy to debug or diagnose
  - This is a rare failure mode in true LAN environments
- This problem is exacerbated if you don’t have an autodiscovery mechanism
  - Greater likelihood of misconfiguration leading to partial mesh creation
Full Mesh Connectivity

- Assume that one connection goes down from the full mesh in the previous diagram
- Suppose that the CE routers are running OSPF
  - CE1 is the DR, CE2 the BDR
  - CE2 stops hearing hellos from CE1, takes over as DR
  - CE3 and CE4 are now thoroughly confused
- Or suppose that CE1 is ARPing for IP addresses
  - Usually, this works, but when the IP address is behind CE2, there is no ARP response
Full Mesh Connectivity of VPLS PEs

- I-BGP messages go to all peers, by definition
  - This is an inherent part of the protocol
- Thus, by definition there will be full mesh connectivity among PEs for a given VPLS
  - A configuration error (e.g., wrong route target) may result in a PE completely missing a given VPLS, but can never result in a partial mesh
  - Easier to diagnose a completely missing site rather than a partial mesh
5. Loops and Spanning Tree

- Service Providers must protect against a layer 2 loop or broadcast storm in the customer network

- Three ways for a SP to do this
  - Rate-limit broadcast, multicast and flooding traffic from the customer devices
  - Run Spanning Tree Protocol on the PE-CE links
  - Whenever possible, keep control of loop avoidance and link selection with the Service Provider
Broadcast Storms

- One *must* rate-limit the flooding of packets to unknown addresses
  - Possible that the source MAC address is never learned
- One *should* rate-limit broadcasting
  - Limit damage due to broadcast storms
- One *should* rate-limiting multicast traffic
  - In principle, less damaging than broadcast
- Ideally, each of these should have independent knobs, to adapt to customer needs
VPLS and BGP Path Selection

A multi-homed CE would normally immediately cause a layer 2 loop. This is usually resolved by having the CE run STP. However, an alternative is to use BGP path selection.

**Path Selection**

Prefer PE 2; install route to PE 2 with VPLS label 94

**PE2 withdraws**

PE4 redoes path selection, picks path via PE 3

2 announcements for site 1 of RED VPLS with different Local Preferences
6. Inter-AS/Inter-provider VPLS

- A strong requirement in R&E Networks
- Defined in 2547bis for IP VPNs, but can be used as is for BGP L2 VPNs and VPLS
- 3 options: option A, option B, **option C**

Summary: MP-BGP offers a scalable Inter-AS solution with Route Reflectors
Route Reflectors For Inter-AS VPLS

Brute force Inter-AS signaling:
Set up sessions between every PE in AS 1 and every PE in AS 2: MxN sessions, authentication nightmare

BGP with Route Reflectors:
Set up sessions between RRs in AS1 and RRs in AS2 -- easier to manage, fewer authentication keys
Loop-free Distribution of VPLS NLRIs

AS path loop detection

AS 1

AS 2

AS 3

AS 4

AS path-based path selection
Multi-hop EBGP Distribution of Labeled VPN Routes Between PE Routers (1)

Inter-Provider VPN/VPLS **Option C** in 2547bis

Multi-As Operations with a **Direct Connection** Between BGP/MPLS VPN Providers

- a direct L2 link
- a L2 VPN pt-to-pt connection
- a GRE/IPSec tunnel
Multi-hop EBGP Distribution of Labeled VPN Routes Between PE Routers (2)

Multi-As Operations with a Direct Connection Between BGP/MPLS VPN Providers

Forwarding Plane

Provider I-BGP

MP-eBGP (for provider’s VPLS NLRI)

R_1

ASBR 1

Provider IGP + LDP

PE_3

Site 1

VPLS NLRI
NH: PE4
to PE4: NH ASBR1

VFT

Push

Push

Push

Swap

Direct E-BGP (for provider’s internal routes)

R_2

ASBR 4

Provider IGP + LDP

PE_4

Site 2

10.2/16

Control Plane

PHP

Pop

Pop

Pop

Push

Pop

Pop

Pop

Site 2

10.2/16

Multi-hop EBGP Distribution of Labeled VPN Routes Between PE Routers (2)

Multi-As Operations with a Direct Connection Between BGP/MPLS VPN Providers

Provider I-BGP

MP-eBGP (for provider’s VPLS NLRI)

R_1

ASBR 1

Provider IGP + LDP

PE_3

Site 1

VPLS NLRI
NH: PE4
to PE4: NH ASBR1

VFT

Push

Push

Push

Swap

Direct E-BGP (for provider’s internal routes)

R_2

ASBR 4

Provider IGP + LDP

PE_4

Site 2

10.2/16

Control Plane

PHP

Pop

Pop

Pop

Push

Pop

Pop

Pop

Site 2

10.2/16
Multi-hop EBGP Distribution of Labeled VPN Routes Between PE Routers (3)

- Advertise labeled Internal Routes (/32) routes into other AS
- Establish LSP between ingress and egress PE
- Use multihop EBGP over established LSP
- If /32 PE addresses not advertised to P router, can use 3-level label-stack
- ASBR is not aware of VPN information (scalable !)
Multi-hop EBGP Distribution of Labeled VPN Routes Between PE Routers (4)

Multi-As Operations with a BGP/MPLS VPN Capable Transit Provider

- **MP-eBGP** (for provider’s VPLS NLRIs)
- **Provider I-BGP**
- **Provider IGP + LDP (AS 1)**
- **Direct E-BGP** (for provider’s internal routes)
- **Transit Provider IGP + LDP**
- **Direct E-BGP** (for provider’s internal routes)
- **Provider I-BGP**
- **Provider IGP + LDP (AS 2)**

**Forwarding Plane**

- PE_3 pushes Three labels

- PE_3 pushes
- R_1
- PE_1
- Direct E-BGP
- (ASBR)
- Provider IGP + LDP
- Site 2
- Site 1
- PE_2
- PE_4
- Site 2
- Site 1
- PE_3
- PE_1
- PE_2
- CE_1
- CE_2
- P
- (ASBR)
- (ASBR)
- (ASBR)
- VRF
- VRF
- VFT
- VFT
- VFT
- VFT

**Forwarding:**

- PE_3 pushes Three labels

- Forwarding Plane

- push
- push
- push
- pop
- swap
- swap
- push
- pop
- swap
- pop
- pop
- pop
Recursive Multi-AS Operations

- **IGP**: Maintains Provider A’s + D’s Internal Routes
- **E-BGP**: Maintains Provider B’s + C’s Internal Routes
- **MP-eBGP**: Maintains Provider B’s + C’s External (Provider A’s + D’s Internal)
- **MP-iBGP**: Transit Provider’s External (Provider B’s + C’s internal)
- **VPLS NLRIs**
- **VRF**: Provider A’s + D’s VPLS NLRIs

Provider A (AS 1)
Provider B (AS 2)
Transit Provider (AS 3)
Provider C (AS 4)
Provider D (AS 5)
Recursive Multi-AS Operations

Can be:
- a direct L2 link
- a L2 VPN pt-to-pt connection
- a GRE/IPSec tunnel
Recursive Multi-AS Operations

- Direct E-BGP (for provider B’s + C’s internal routes)
- IGP
- Maintains VPLS NLRIs

Can be:
- a direct L2 link
- a L2 VPN pt-to-pt connection
- a GRE/IPSec tunnel

- MP-eBGP Provider A’s + D’s VPLS NLRIs
- Provider A’s + D’s Internal Routes
- IGP
- Maintains VPLS NLRIs

Copyright © 2004 Juniper Networks, Inc.
Recursive Multi-AS Operations

This is actually a CPE based VPN!
- Complexity managed by end-users
- Scalability issue
- Do NOT require any VPN service from transit provider (if GRE or IPSec Tunnel)

Can be:
- a direct L2 link
- a L2 VPN pt-to-pt connection
- a GRE/IPSec tunnel
Inter-AS/Inter-provider VPLS

- Exchange VPN information + VPN labels across AS/provider boundary by using BGP between BGP Route Reflectors in each AS/provider
  - Route Reflectors preserve the next hop information and the VPN label across the AS/provider

- PEs learn routes and label information of the PEs in the neighboring ASes through ASBRs
  - Using labeled IPv4 routes

- No VPN information (e.g., VRF, VFT) on ASBRs

Applies to RFC2547 VPN, L2 VPN, and VPLS !!!
7. Status on Deployment

- Korea Telecom and Hutchinson have jointly announced an inter-provider VPLS deployment using BGP for signaling and auto-discovery.
- Major carrier in the US has tested inter-metro VPLS for over 8 months, and ran a beta trial for their customers. Deployment started in June, to reach over 40 US metro areas by end of ‘04.
  - Active dialogue, many features requested and, yes, implemented.
Status on Deployment

- Catch Communications, an Ethernet-centric carrier in Norway tested VPLS, and laid out their design. They have several active customers.
- Another carrier in Norway has a small VPLS deployment for internal use.
- Several Metro Ethernet providers in Europe and Asia are actively testing BGP VPLS.
- Other groups in the US have also begun testing; target is to replace existing LANE networks.
Thank you!

http://www.juniper.net

kireeti@juniper.net