Objectives

- Provide a solution that enables Service Providers to offer VPN service that:
  - Scales to large number of customers
    - 100,000 - 1,000,000 VPNs
  - Most of the customers with little/no IP routing expertise
  - Supports diverse customer population
    - Including large VPNs (100s - 1000s sites per VPN)
  - Value-added service
  - Low cost service
What Is a VPN?

- VPN is a set of sites which are allowed to communicate with each other.
- VPN is defined by a set of administrative policies.

  Policies determine both connectivity and QoS among sites.
  Policies established by VPN customers.
  Policies could be implemented completely by VPN Service Providers.

Using BGP/MPLS VPN mechanisms.

What Is a VPN (Cont.)?

- Flexible inter-site connectivity ranging from complete mesh to hub-and-spoke.
- Sites may be either within the same or in different organizations.
  VPN can be either Intranet or Extranet.
- Site may be in more than one VPN.
  VPNs may overlap.
- Not all sites have to be connected to the same service provider.
  VPN can span multiple providers.
BGP/MPLS VPN—Example

BGP/MPLS VPN Key
Components:

(1) Constrained distribution of routing information + multiple forwarding tables
(2) Address extension
(3) MPLS
BGP/MPLS VPN Key
Components:

(1) Constrained distribution of routing information + multiple forwarding tables
(2) Address extension
(3) MPLS

Constrained Distribution of Routing Information

• Provides control over connectivity among sites
  flow of data traffic (connectivity) is determined by flow (distribution) of routing information
Routing Information Distribution

Step 1: from site (CE) to service provider (PE)
e.g., via RIP, or static routing, or OSPF, or IGRP, or EIGRP, or BGP

Step 2: export to provider’s BGP at ingress PE

Step 3: within/across service provider(s) (among PEs):
via BGP

Step 4: import from provider’s BGP at egress PE

Step 5: from service provider (PE) to site (CE)
e.g., via RIP, or static routing, or OSPF, or IGRP, or EIGRP, or BGP

Constrained Distribution of Routing Information

- Occurs during Steps 2, 3, 4
- Performed by Service Provider using route filtering based on BGP Extended Community attribute

BGP Community is attached by ingress PE at Step 2
per site (per CE), per address prefix granularity
route filtering based on BGP Community is performed by egress PE at Step 4
Amount of routing peering maintained by CE is $O(1)$ - CE peers only with directly attached PE
Independent of the total number of sites within a VPN
Scales to VPNs with large number of sites (100s - 1000s sites per VPN)
In contrast, mesh with any tunnel-based approach (e.g., IPSec) requires $O(n)$ peering (where $n$ is the number of sites)
Does not scale to VPNs with large number of sites

Amount of configuration changes needed to add a new site (new CE) is $O(1)$:
Need to configure only the directly attached PE
Independent of the total number of sites within a VPN
In contrast, mesh with any tunnel-based approach (e.g., IPSec) requires $O(n)$ changes (where $n$ is the number of sites)
Scales worse than the BGP/MPLS VPN
Multiple Forwarding Tables

• How to constrain distribution of routing information at PE that has sites from multiple (disjoint) VPNs attached to it?

  Single Forwarding Table on PE doesn’t allow per VPN segregation of routing information

Multiple Forwarding Tables (Cont.)

• PE maintains multiple Forwarding Tables
  One per set of directly attached sites with common VPN membership
  E.g., one for all the directly attached sites that are in just one particular VPN

• Enables (in conjunction with route filtering) per VPN segregation of routing information on PE
Multiple Forwarding Tables (Cont.)

- Each Forwarding Table is populated from:
  (a) routes received from directly connected CE(s) of the site(s) associated with the Forwarding Table
  (b) routes receives from other PEs (via BGP)
    Restricted to only the routes of the VPN(s) the site(s) is in
    Via route filtering based on BGP Extended Community Attribute

- Each customer port on PE is associated with a particular Forwarding Table
  Via configuration management (at provisioning time)
- Provides PE with per site (per VPN) forwarding information for packets received from CEs
- Ports on PE could be “logical”
  E.g., VLAN, FR, ATM, L2F, etc...
MPLS VPN Example: Full Mesh

- One BGP Community $C_{closed}$
- At PE with directly attached site:
  - Exports all site’s routes into provider’s BGP with Community $C_{closed}$
  - Imports into the forwarding table associated with the VPN (site) only routes with $C_{closed}$
- Could also be realized with more than one BGP Community
  - Useful when VPN spans multiple providers
  - Each provider uses its own BGP Community

MPLS VPN Example: Hub/Spoke VPN

All Spoke Sites Communicate through Hub

- PE at the hub site:
  - **exports** site’s routes with BGP Community $C_{hub}$
  - **import** into the forwarding table associated with the VPN (site) routes with $C_{spoke}$
- PE at a spoke sites:
  - **export** site’s routes with BGP Community $C_{spoke}$
  - **import** into the forwarding table associated with the VPN (site) routes with $C_{hub}$
BGP/MPLS VPN Key Components:

(1) Constrained distribution of routing information + multiple forwarding tables

(2) Address extension

(3) MPLS

Address Extension

- How to support VPNs without imposing constraints on address allocation/management within VPNs (e.g., allowing private address space [RFC1918])?

  Constrained distribution of routing information uses BGP

  BGP is designed with the assumption that addresses are unique
**VPN-IP Addresses**

- New address family: VPN-IP addresses
  
  VPN-IP address = Route Distinguisher (RD) + IP address

  RD = Type + Provider’s Autonomous System Number + Assigned Number

  Assigned Number field is assigned by each service provider

  Each VPN has its own Assigned Number (but may have more than one)

  **No two VPNs have the same RD**

  RD converts non-unique IP addresses into unique VPN-IP addresses

- Reachability information for VPN-IP addresses is carried via multiprotocol extensions to BGP-4

**Converting between IP and A VPN-IP Addresses**

- Performed by PE

  ingress PE - exporting route into provider’s BGP:

  PE is configured with RD(s) for each directly attached VPN (directly attached sites)

  convert from IP to VPN-IP (by prepending RD) before exporting into provider’s BGP

  egress PE - importing route from provider’s BGP:

  convert from VPN-IP to IP (by stripping RD) before inserting into site’s forwarding table

  **Happens Only in the Control Plane, Not in the Data Plane**
**Route Distinguisher vs BGP Communities**

- **Route Distinguisher:**
  - Used to disambiguate IP addresses
  - Via VPN-IP addresses
  - Not used to constrain distribution of routing information (route filtering)

- **BGP Communities:**
  - Not used to disambiguate IP addresses
  - Used to constrain distribution of routing information
  - Via route filtering based on BGP Communities

**BGP/MPLS VPN Key Components**

1. Constrained distribution of routing information + multiple forwarding tables
2. Address extension
3. MPLS
MPLS

• Given that BGP operates in terms of VPN-IP addresses, how to forward IP packets within Service Provider(s) along the routes computed by BGP?
  
  IP header has no place to carry Route Distinguisher

MPLS (Cont.)

• Use MPLS for forwarding
  
  MPLS decouples information used for forwarding (label) from the information carried in the IP header

• Label Switched Paths are bound to VPN-IP routes

• Label Switched Paths are confined to VPN Service Provider(s)
Packet Forwarding—Example

- Logically separate forwarding table (FIB) for each (directly attached) VPN
  - Expressed in terms of IP address prefixes
  - Conversion from VPN-IP to IP addresses happen when FIB is populated from the routing table (RIB)
- Incoming interface determines the FIB

1. Identify VPN
2. Select FIB for this VPN
3. Attach label info and send out

Two-Level Label Stack

- VPN routing information is carried only among PE routers (using BGP)
- BGP Next Hop provides coupling between external routes (VPN routes) and service provider internal route (IGP routes)
  - Route to Next Hop is an internal route
- Top level label is used for forwarding from ingress PE to egress PE
- Bottom level is used for forwarding at egress PE
  - Distributed via BGP (together with the VPN route)

⇒ P routers maintain only internal routes (routes to PE routers and other P routers), but no VPN routes
Scalability—“Divide and Conquer”

(1) Two levels of labels to keep P routers free of all the VPN routing information

(2) PE router has to maintain routes only for VPNs whose sites are directly connected to the PE router

(3) Partition BGP Route Reflectors within the VPN Service Provider among VPNs served by the Provider

⇒ No single component within the system is required to maintain all routes for all the VPNs

⇒ Capacity of the system isn’t bounded by the capacity of an individual component

Security—Protection against Packet Spoofing

Goal:
In the absence of misconfiguration or deliberate interconnection of different VPNs, it should not be possible for hosts/routers in one VPN to gain access to hosts/routers in another VPN
Security—Protection against Packet Spoofing (Cont.)

- Forwarding is based on Label Switching
- Inter-site LSPs originate/terminate at PEs
- LSPs are associated with logical ports on PEs
- Logical port on PE is bound (at provisioning time) to a particular VPN
  at forwarding time determines the Forwarding Table associated with the VPN
  ⇒ Injecting a packet into a VPN could be done only through a port on PE associated with that VPN

Quality of Service (QoS)

- Essential component of any competitive VPN service offering
- Goals:
  Flexible: in order to support a broad range of VPN customers
    E.g., multiple Classes of Service per each VPN, per application/protocol mapping to any Class of Service, etc...
  Scaleable: in order to support large number of VPN customers
VPN Performance Service Models

- Reflect customer’s way of specifying the performance requirements of VPN
  - “Hose” model
  - “Pipe” model

QoS: “Hose” Model

- Provides “relative” (“soft”) performance guarantees:
  - From an origin site to a set of destination sites
    - A port on an ingress PE and ports on egress PEs
  - From a set of origin sites to a destination site
    - Ports on ingress PEs and a port on an egress PE

- Characterized by:
  - Ingress Committed Rate (ICR) - the aggregate traffic from the origin site to any of the destination sites
  - Egress Committed Rate (ECR) - the aggregate traffic from all other sites to a particular site
QoS: “Pipe” Model

- Provides “absolute” (“hard”) performance guarantees for traffic from a specific origin to a destination site from a port on an ingress PE to a port on an egress PE (unidirectional)

Similar to FR/Private Line - easily understood by customers with existing FR/ATM/Private Line VPNs

“Hose” + “Pipe”

- Not mutually exclusive
- Combination may be useful in certain situations, e.g.,

  - each site may send (to all other sites):
    - up to 1 Mbps `gold’ (“hose” model)
    - up to 2 Mbps `silver’ (“hose” model)
    - unlimited best effort (“hose” model)
  - from head office to backup site:
    - 1.5 Mbps guaranteed (“pipe” model)
QoS Mechanisms

- Committed Access Rate (CAR)
- Weighted RED (WRED)
- Weighted Fair Queueing (WFQ)
- MPLS constraint-based routing

Committed Access Rate (CAR)

- A flexible packet marking setting mechanism
  - Mark packets based on:
    - (a) received interface (identifies VPN)
    - (b) TCP and IP headers (application specific)
    - (c) conformance to a token bucket
- Primary functions:
  - (1) classification (via marking) into a particular Class of Service (CoS)
  - (2) policing: within each CoS identify packets in/out of profile
    - Mark “in” with lower drop probability than “out”
      - e.g. use MPLS COS (“experimental”) bits, CLP
Weighted RED (WRED)

- WRED can manage queues that develop in routers
  - Provides RED benefits (shorter queues, less synchronization, improved fairness)
  - Drops most from heaviest users
  - Prioritizes in-profile traffic over the rest
    - Based on in/out of profile marking provided by CAR

![Graph showing P(drop) vs Q_avg](image)

Weighted Fair Queuing (WFQ)

- Basic idea: emulate bit-wise round robin among some number of classes

- Example:
  - Sort traffic into n Classes of Service (CoS)
  - Allocate link percentage to each CoS: \( w_i / \Sigma w \)
  - Any unused bandwidth from one CoS is available to the others
  - Relationship of offered load to \( w_i \) (level of over-subscription) determines how good the CoS is
CAR + WRED + WFQ

- CAR for marking and policing (per VPN) at PEs only
- WRED + Class-based WFQ for service differentiation:
  - both at PEs and at Ps
  - provides per Class of Service:
    - selective discard (WRED)
    - selective scheduling (WFQ)

**PE LSR—Example**

- Operation
  - CAR classifies packets
  - CAR polices packets
  - Per-class WRED
  - Per-class WFQ
Per-Class WRED
Per-Class WFQ

• Per-Class Discard (WRED) at every hop
• Per-Class Scheduling (WFQ) at every hop
• Per-class WFQ allows bandwidth to be reused (by other classes) if class doesn’t need its share
• No per-VPN state on P LSRs

Diff-serv provides selective discard (via WRED), and selective scheduling (via WFQ)

Diff-serv is NOT a mechanism to match resources required to provide QoS with the available resources on a network-wide basis

Matching available with required resources on a network-wide basis is not a queueing problem

Matching available with required resources is a routing problem

Solved by MPLS constraint-based routing
QoS and Scalability

- One queue per Class of Service, not per LSP
- One Diff-Serv class (2 Diff-Serv control points) per Class of Service
- One LSP per Class of Service (not per VPN) between given ingress and egress PE pair

⇒ No per VPN state on P LSRs

Key Features

- No constraints on addressing plans used by VPNs—a VPN customer may:
  (a) use globally unique and routable addresses,
  (b) use globally unique, but non-routable addresses
  (c) use private addresses (RFC1918)
  (d) use addresses that are neither globally unique, nor private
  (e) use any combination of the above
Key Features (Cont.)

• **Security:**

  Basic security is comparable to that provided by FR/ATM-based VPNs

  Without requiring encryption

  Without complexity and overhead of key management and encryption

  For data privacy VPN customer may use IPSec-based mechanisms

  e.g., CE - CE IPSec-based encryption

Key Features (Cont.)

• **Flexible and scalable support for a wide range of performance characteristics:**

  Multiple Classes of Service per VPN

  Per application/protocol mapping to any Class of Service

  “Hose” and “pipe” models - “mix and match”

  No per VPN state in the P LSRs
Key Features (Cont.)

- **Scalability:**
  - Total capacity of the system isn’t bounded by the capacity of an individual component
  - Scale to virtually unlimited number of VPNs per VPN Service Provider
  - Scale to thousands of sites per VPN

Key Features (Cont.)

- **Simplifies network operations and management for VPN customers:**
  - No need for a VPN customer to manage a backbone or “virtual backbone” of any sort
  - Simplified routing strategy for VPN customers
  - E.g., use (single) default route for single-homed VPN sites

⇒ **Minimizes skill investment by VPN customers**
Key Features (Cont.)

• Simplifies operations and management for VPN Service Providers:

  No need for VPN Service Providers to set up and manage a separate backbone or “virtual backbone” for each VPN

  ⇒ More scalable and cost effective than service based on IP tunnels (e.g., IPSec) overlay

Key Features (Cont.)

• Connectivity to the Internet:

  VPN Service Provider may also provide connectivity to the Internet to its VPN customers

  ⇒ Common infrastructure for both VPN and the Internet connectivity services
Key Features (Cont.)

• Media independence:
  No constraints on the Link Layer technology used within the VPN Service Provider network
  No constraints on the Link Layer technology used to interconnect VPN customers to the VPN Service Provider network

BGP/MPLS VPN—Summary

• Supports large scale VPN services
• Increases value add by the VPN Service Provider
• Decreases Service Provider’s cost of providing VPN services
• Mechanisms are general enough to enable VPN Service Provider to support a wide range of VPN customers
Suggested Reading

Network Working Group
Internet Draft
Expiration Date: September 2000

Eric C. Rosen, Yakov Rekhter (Cisco Systems)
Stephen John Brannon, Monique Jeanne Morrow (Swisscom AG)
Marco Carugi (France Telecom)
Christopher J. Chase (AT&T)
Eric Dean (Global One)
Paul Hitchin, Adrian Smith (BT)
Manoj Leelanivas (Juniper Networks, Inc.)
Luca Martini (Level 3 Communications, LLC)
Vijay Srinivasan (Ericsson IP Infrastructure)
Alain Vedrenne (SITA EQUANT)

March 2000

BGP/MPLS VPNs
draft-rosen-rfc2547bis-00.txt

Suggested Reading

• See RFC2547 for more details...
Thank You!