Multicast Deployment Experiences

SANOG VII Mumbai, India 16-24Jan'06

Rajesh Shinde
rajesh.shinde@relianceinfo.com

Reliance Infocomm

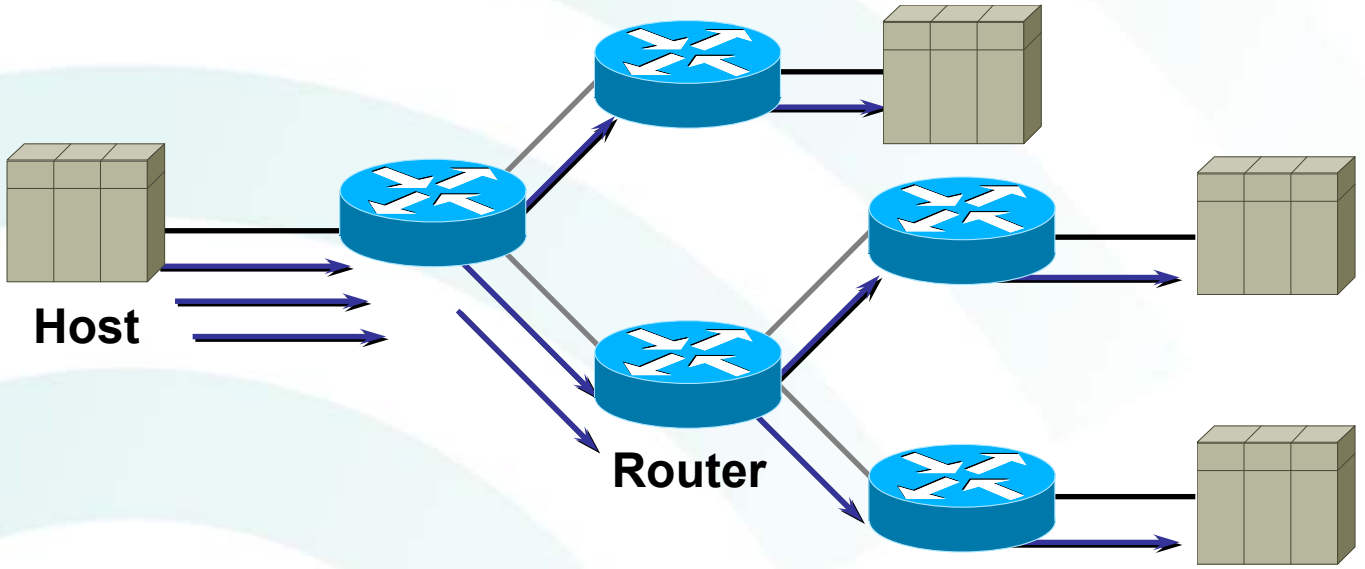
Agenda

- Why Multicast?
- Multicast Service Model
- Multicast Routing Protocols—Characteristics
- Multicast for 2547bis VPNs
- Current Implementations
- A Next Gen MVPN Solution
- Looking Ahead

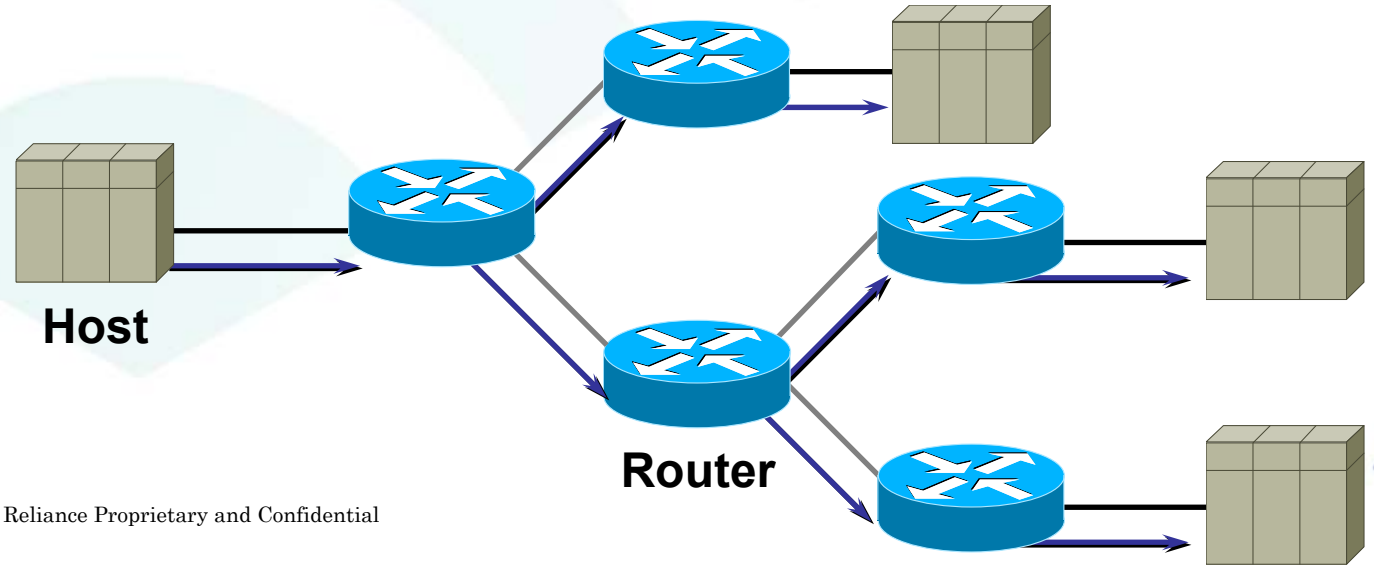
Why Multicast?

- Why multicast?
  - When sending same data to multiple receivers
  - Better bandwidth utilization
  - Lesser host/router processing
  - Receivers' addresses unknown
- Applications
  - Video conferencing
  - Video-On-Demand
  - Resource discovery/service advertisement
  - E-learning
  - Real-time Data Delivery – Financial

Why Multicast? (cont.....)



Unicast



Multicast

Reliance Infocomm

Agenda

- Why Multicast?
- Multicast Service Model
- Multicast Routing Protocols—Characteristics
- Multicast for 2547bis VPNs
- Current Implementations
- A Next Gen MVPN Solution
- Looking Ahead

Multicast Service Model

- RFC 1112
- Each multicast group identified by a class D IP address
- Members of the group could be present anywhere on the network
- Members join and leave the group and indicate this to the routers using IGMP
- Senders and receivers are distinct: i.e., a sender need not be a member
- Routers listen to all multicast addresses and use multicast routing protocols to manage groups

Agenda

- Why Multicast?
- Multicast Service Model
- Multicast Routing Protocols—Characteristics
- Multicast for 2547bis VPNs
- Current Implementations
- A Next Gen MVPN Solution
- Looking Ahead

Multicast Routing Protocols -Characteristics

- Distribution trees
  - Source tree
    - Uses more memory O(S ,G) but you get optimal paths from source to all receivers, minimizes delay
  - Shared tree
    - Uses less memory O(*,G) but you may get suboptimal paths from source to all receivers, may introduce extra delay
- Protocols
  - PIM, DVMRP, MOSPF, CBT, MBGP, MSDP

Multicast Routing Protocols -Characteristics (cont…..)

- Types of multicast mode
  - Dense-mode
    - Push Model
    - Broadcast and prune behavior
    - Similar to radio broadcast
  - Sparse-mode
    - Pull Model
    - Explicit join behavior
    - Similar to pay-per-view

Agenda

- Why Multicast?
- Multicast Service Model
- Multicast Routing Protocols—Characteristics
- Multicast for 2547bis VPNs
- Current Implementations
- A Next Gen MVPN Solution
- Looking Ahead

Multicast for 2547bis VPNs

Pertinent Drafts

- draft-rosen-vpn-mcast-06
- draft-rosen-vpn-mcast-07
- draft-raggarwa-l3vpn-2547-mvpn-00
- draft-raggarwa-l3vpn-mvpn-vpls-mcast-00

RFC 2547 and RFC2547bis originally did not support multicast

Multicast for 2547bis VPN (Cont…)

draft-rosen-vpn-mcast-06

- Solution 1: Multicast Domains
  - P2MP tunneled solution
- Solution 2: Multicast VRF
  - Never implemented
- Solution 3: NBMA/Ingress Replication

Multicast for 2547bis VPN (Cont…)

draft-rosen-vpn-mcast-07

- Removed Solutions 2 and 3
- Filled in the details for Solution 1
  - Additions of PIM-SSM in the P-instance (optional) and SSM SAFI (required)
  - Not interoperable with implementations based on version 06

draft-raggarwa-l3vpn-2547-mvpn-00

- Based on Rosen draft 06, Solution 1 preferred

Multicast for 2547bis VPN (Cont…)

draft-raggarwa-l3vpn-mvpn-vpls-mcast-00

- Completely new solution that addresses the scalability issues with Rosen draft and draft-raggarwa-l3vpn-2547-mvpn-00
- Also addresses multicast in VPLS

Reliance Infocomm

Agenda

- Why Multicast?
- Multicast Service Model
- Multicast Routing Protocols—Characteristics
- Multicast for 2547bis VPNs
- Current Implementations
- A Next Gen MVPN Solution
- Looking Ahead

Current Implementations

- Rosen-06 and draft-raggarwa-l3vpn-2547-mvpn-00
  - Ingress Replication (Solution-3)
  - Multicast Domains (Solution-1)
    - **Default MDT only**
    - **Default MDT with Data MDT**
- Rosen-07
  - SSM in SP core

Current Implementations (Cont…)

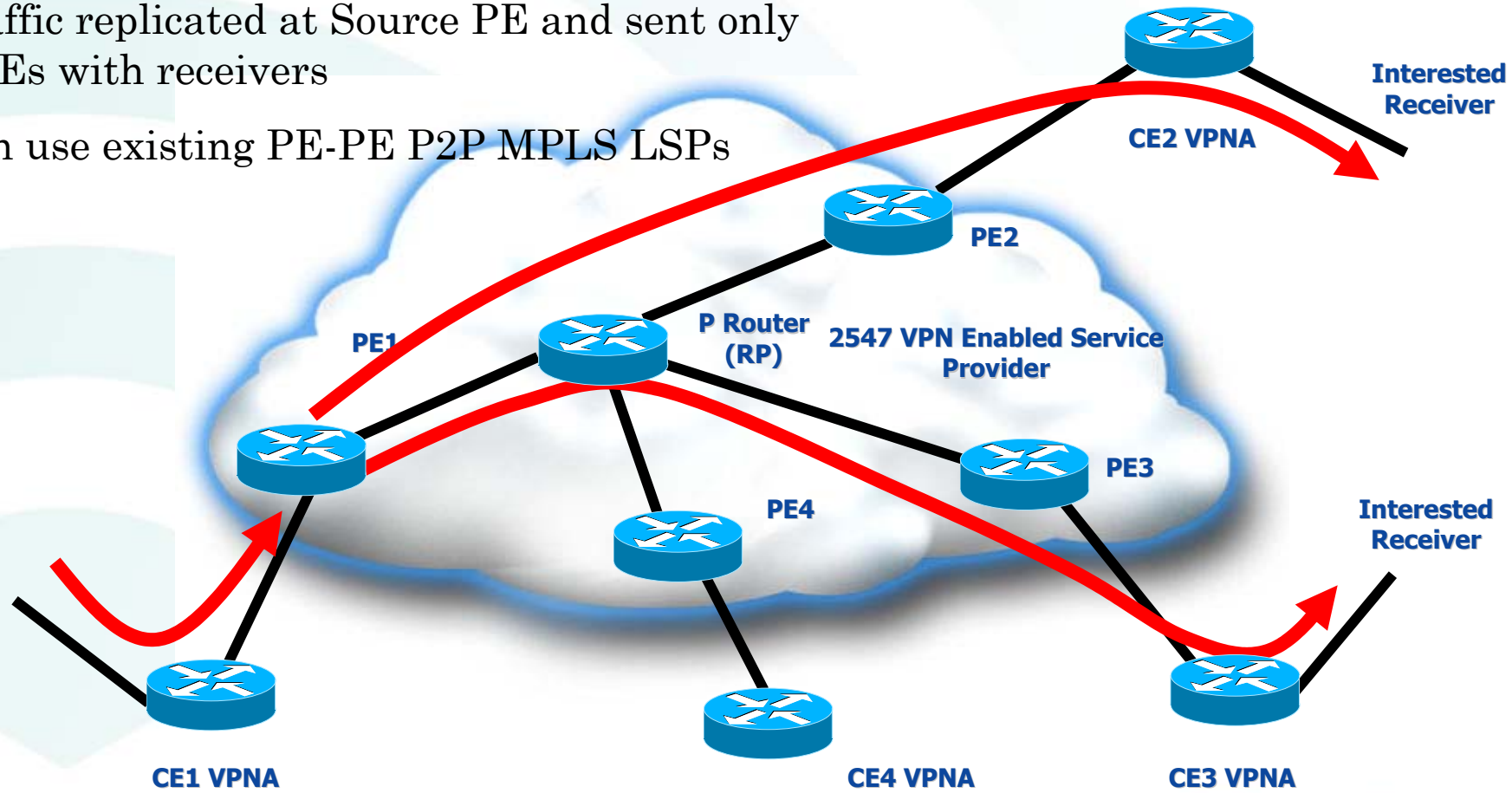
Rosen-06 Solution 3: NBMA/Ingress Replication

- **Multicast control and data sent from PE-PE through P2P tunnels**
  - **Typically, the same LSPs used by VPN unicast**
- **Traffic replicated by source's PE and sent only to PEs that send PIM joins**
- **Very simple**

Current Implementations (Cont…)

MVPN Traffic Flow- Ingress Replication

- Traffic replicated at Source PE and sent only to PEs with receivers
- Can use existing PE-PE P2P MPLS LSPs

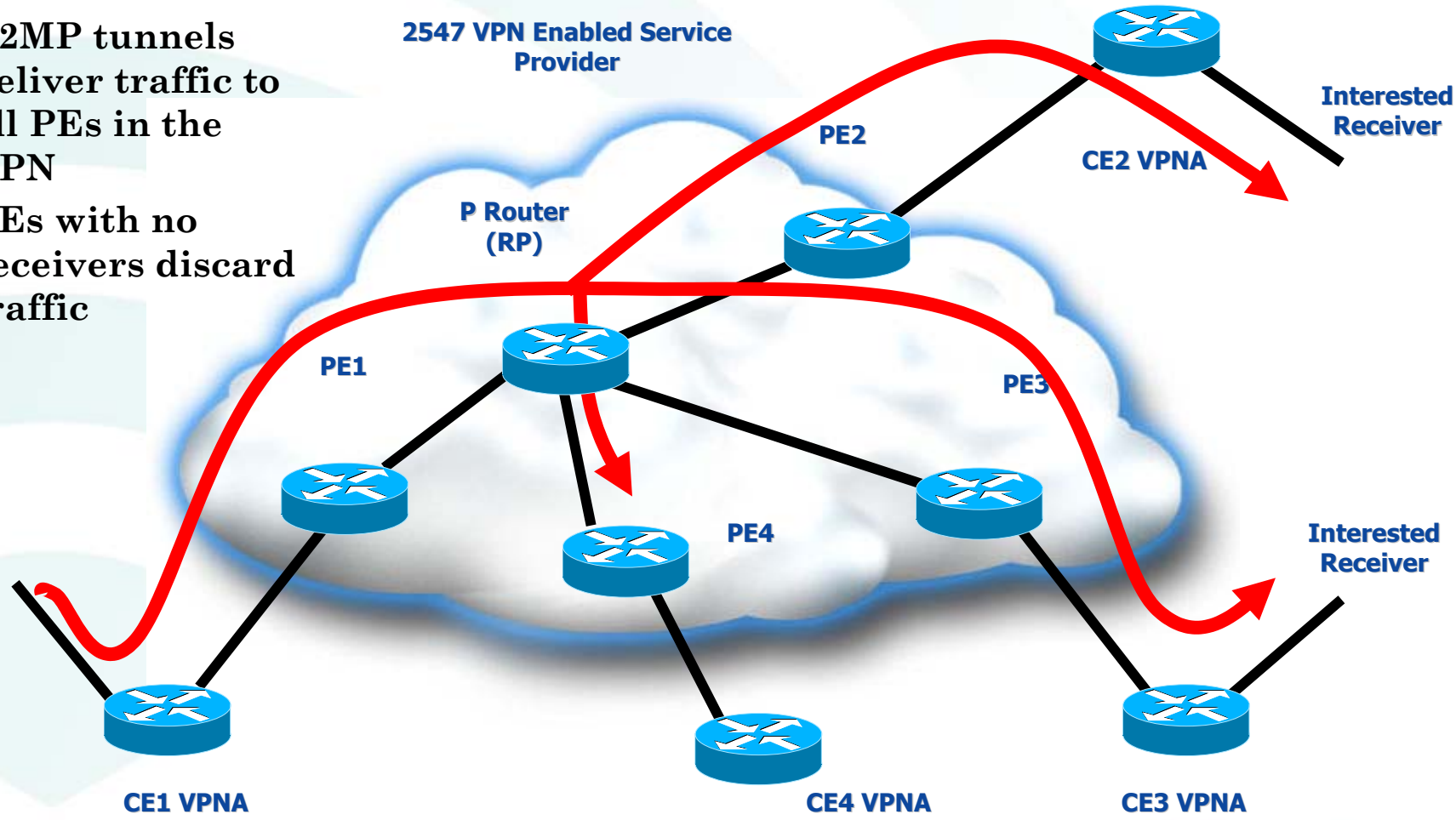


Reliance Infocomm

Current Implementations (Cont…)

MVPN Traffic Flow- Multicast Domains

- P2MP tunnels deliver traffic to all PEs in the VPN
- PEs with no receivers discard traffic



Reliance Infocomm

Current Implementations (Cont…)

Rosen-06 Solution 1: Multicast Domains

- **PIM must run in the core**
  - **Provider PIM instance for creating MDT interfaces**
  - **PE P2MP GRE endpoint per VRF- MDT interfaces**
  - **By default, all control and data sent over a single "default MDT"**
- **Customer instance of PIM between PE-CE**
- **Traffic sent to all PEs with a member of the same VPN**

Current Implementations (Cont…)

Rosen-06 Solution 1: Multicast Domains (Default MDT only)

- Perpetual flooding of mcast data to all PEs with a VPN member
- Most efficient when there are few prunes
  - i.e., when most CEs have at least one receiver
  - Like the old dense mode assumption
- Simple (relatively)
- Efficient when there are many low data rate groups
- Scalability of Provider PIM instance
  - P routers support N groups, where N is # of VPNs
    - With SPT (SSM or ASM), this means N x M state in the core, where M is the number of PEs with a VPN member
    - With shared trees (ASM or Bidir), there is only N state in the core

Current Implementations (Cont…)

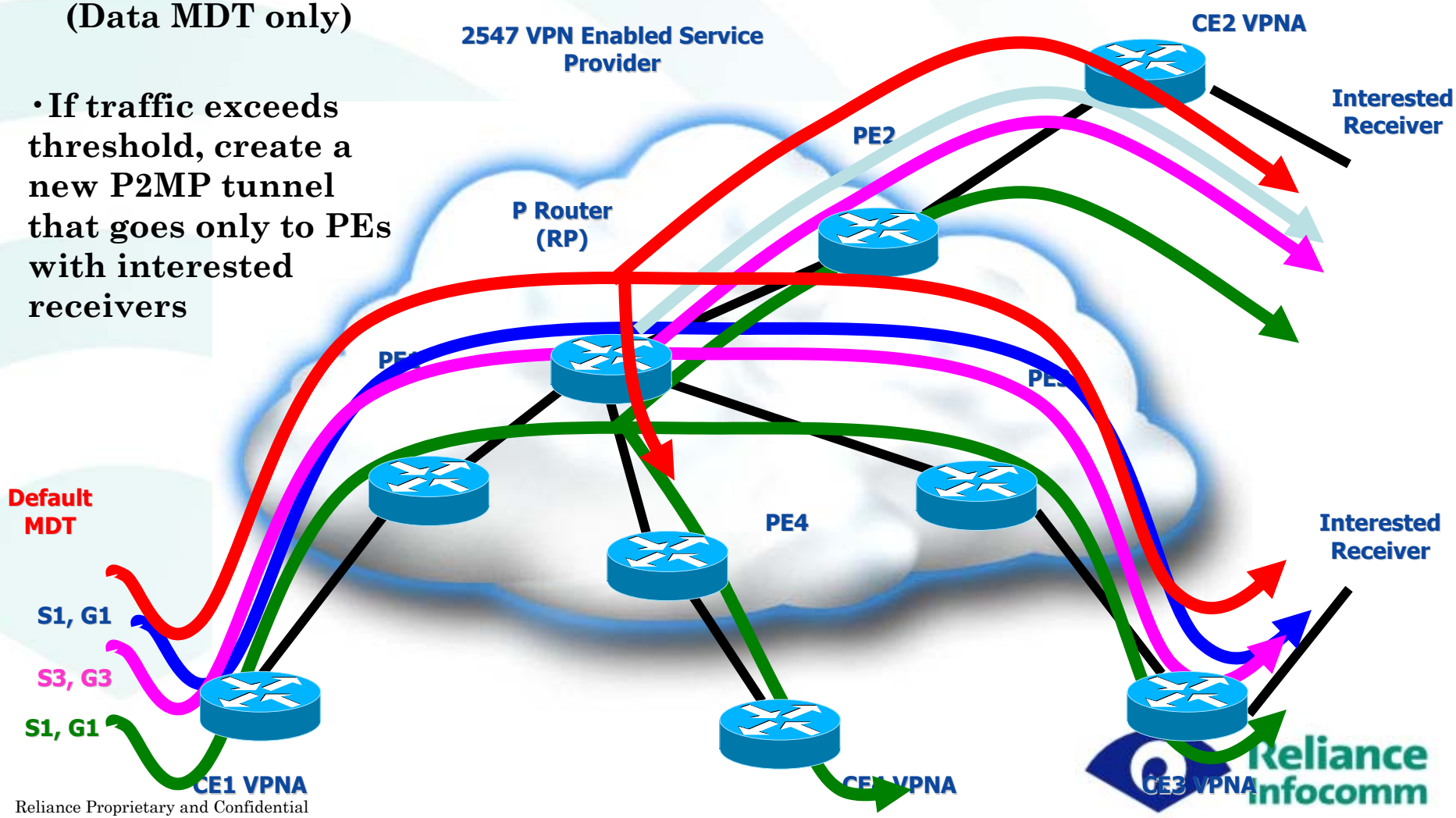
Rosen-06 Solution 1: Multicast Domains (Data MDT only)

- Instead of one MDT per VRF, use multiple MDTs
  - One MDT for control (default), dynamically create new MDT's for data groups in the VPN
- Each high data rate VPN group creates a new data MDT
- Low data rate groups remain on default MDT
- Provides optimal data forwarding for high data rate groups
- Uses UDP signalling between PEs to establish and switch to Data MDT
- Data MDTs initially described in early Rosen drafts in vague terms, details left up to the implementation
  - In Rosen-07, UDP signalling protocol specified

Current Implementations (Cont…)

Rosen-06 Solution 1: Multicast Domains –MVPN Traffic Flow (Data MDT only)

- If traffic exceeds threshold, create a new P2MP tunnel that goes only to PEs with interested receivers



Current Implementations (Cont…)

Rosen-06 Solution 1: Multicast Domains (Data MDT- Disadvantages)

- Each group in the C-VPN could create state in the provider network
- Each group in the C-VPN could create a new MDT interface on the PEs
- Complexity of when to create a new data MDT
- Receiving PEs that are unable to join the data MDT will be blackholed- no feedback mechanism informing the Source-PE that all receiving PEs successfully switched
- No packets and no warning!!

Current Implementations (Cont…)

Simple Comparison of 3 Dimensions of MVPNs

	Data Rate	State	Fanout
Default MDT	Bad	Good	Great
Data MDT	Good	Bad	Great
Ingress Replication	Great	Great	So-so

Data Rate how big is the multicast stream?

State how much routing state do the routers have to keep?

Fanout how many PE routers have a receiver connected?

Current Implementations (Cont…)

Other Problems with Existing Solutions

- PIM neighbors
  - (# of VPNs per PE) x (# of PEs with a site in the VPN)
  - By comparison, unicast requires only 1 BGP session between each PE (or less with route reflectors)
- PIM State
  - P routers must maintain state proportional to the # of VPNs
  - Worsens significantly when you add Data MDT
- PIM Join/Prune processing
  - PIM is a soft state protocol
- Data MDT no feedback for MDT creation mentioned earlier

Current Implementations (Cont…)

Rosen-07: SSM in the SP Core

- Rosen-07 introduces an optional way to use SSM-only in the P-instance using MBGP
  - SSM is optional, but the enabling MBGP SAFI is required

Agenda

- Why Multicast?
- Multicast Service Model
- Multicast Routing Protocols—Characteristics
- Multicast for 2547bis VPNs
- Current Implementations
- A Next Gen MVPN Solution
- Looking Ahead

A Next Gen MVPN Solution

- draft-raggarwa-l3vpn-mvpn-vpls-mcast-00 describes a new set of mechanisms to address each of the issues with Rosen and base draft-raggarwa-l3vpn-2547-mvpn-00
- 2 new solutions:
  - P2P solution- Ingress Replication
  - P2MP solution- P2MP GRE/LSPs

A Next Gen MVPN Solution (Cont…)

MVPNng- P2MP GRE/LSPs

- PIM neighbors- use BGP to eliminate hello processing
- PIM state- aggregate trees
- PIM Join/Prune processing- add refresh reduction to PIM (similar to RSVP)
- Data MDT feedback- RSVP provides feedback
  - Can transmit on both “default” and “data” trees
  - Stop forwarding on “default” tree only after there are no interested PEs
  - “Make before break” similar to SPT switchover

A Next Gen MVPN Solution (Cont…)

MVPNng- Aggregate Trees

- Allow one SP multicast Tree to be shared across multiple VPNs
- Can be setup using PIM-SM
- Requires an inner label to demultiplex a particular VPN
  - 'Upstream' label allocation by the root of the tree
- A flexible tool to reduce state in the SP network
  - A PE may receive un-necessary traffic
- Put MVPNs on par with unicast 2547
- State in the SP network doesn't grow proportional to the number of VPNs
- Use BGP signaling

A Next Gen MVPN Solution (Cont…)

MVPNng- Aggregate Data Trees

- A flexible tool to create separate trees for a set of customer groups to avoid flooding
- To reduce/eliminate unwanted flooding, create aggregate “data” trees
  - Analogous to Data MDT, but can be shared by multiple VPNs
- Setup using PIM-SSM with BGP signaling
- Requires an inner label to demultiplex a particular VPN
  - ‘Upstream’ label allocation by the root of the tree

A Next Gen MVPN Solution (Cont…)

MVPNng- P2MP MPLS

- Can be used to setup Aggregate Tree and Aggregate Data Trees
- Can provide Traffic Engineering
- Fast Reroute
- RSVP-TE connection oriented signaling allows the ingress to avoid state overload
  - Unlike PIM
- RSVP provides feedback to prevent blackholes when moving to Aggregate Data Trees

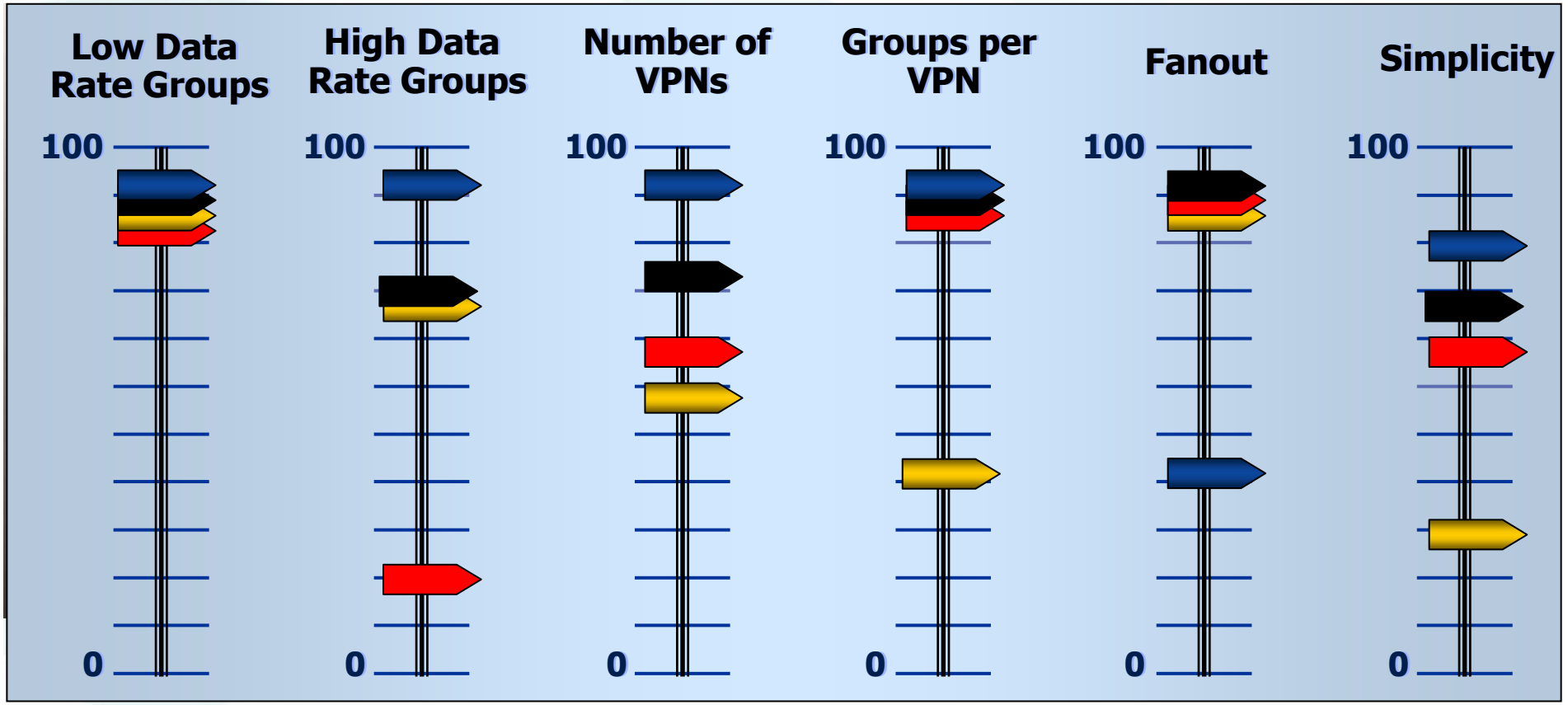
A Next Gen MVPN Solution (Cont…)

MVPNng- Ingress Replication

- PIM neighbors- use BGP to eliminate hello processing
- No flooding, state, PIM, MSDP in the core

A Next Gen MVPN Solution (Cont...)

MVPNng- Scaling MVPNs



Reliance Infocomm

Agenda

- Why Multicast?
- Multicast Service Model
- Multicast Routing Protocols—Characteristics
- Multicast for 2547bis VPNs
- Current Implementations
- A Next Gen MVPN Solution
- Looking Ahead

Looking Ahead

draft-raggarwa-l3vpn-mvpn-vpls-mcast-00

- will be a phased rollout
- Ingress Replication
- BGP for neighbor discovery
- PIM refresh reduction
- Aggregate [Data] Trees
- VPLS- PIM/IGMP snooping

Thank You

Reliance Infocomm