BGP Developments

Timo Liuska <tliuska@juniper.net>

Senior Systems Engineer

June 9th 2016

This statement of direction sets forth Juniper Networks' current intention and is subject to change at any time without notice. No purchases are contingent upon Juniper Networks delivering any feature or functionality depicted in this presentation.

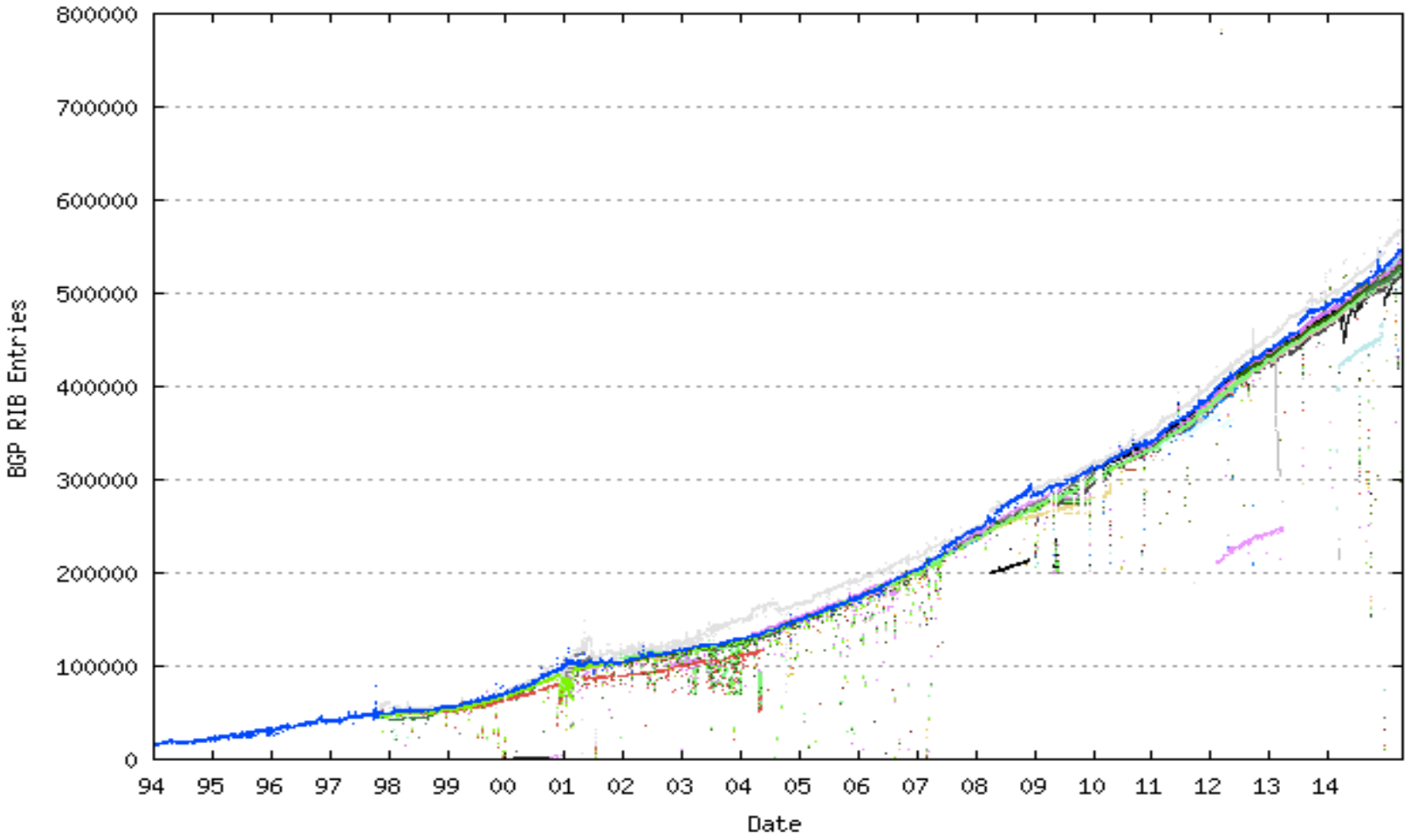
This presentation contains proprietary roadmap information and should not be discussed or shared without a signed non-disclosure agreement (NDA).

AGENDA

- Introduction
- BGP Focus Areas
- BGP Protocol Blocks
- BGP-ORR
- BGP-LS
- BGP Flow Spec

GROWTH OF BGP TABLE FROM 1994 TO PRESENT

Source: bgp.potaroo.net



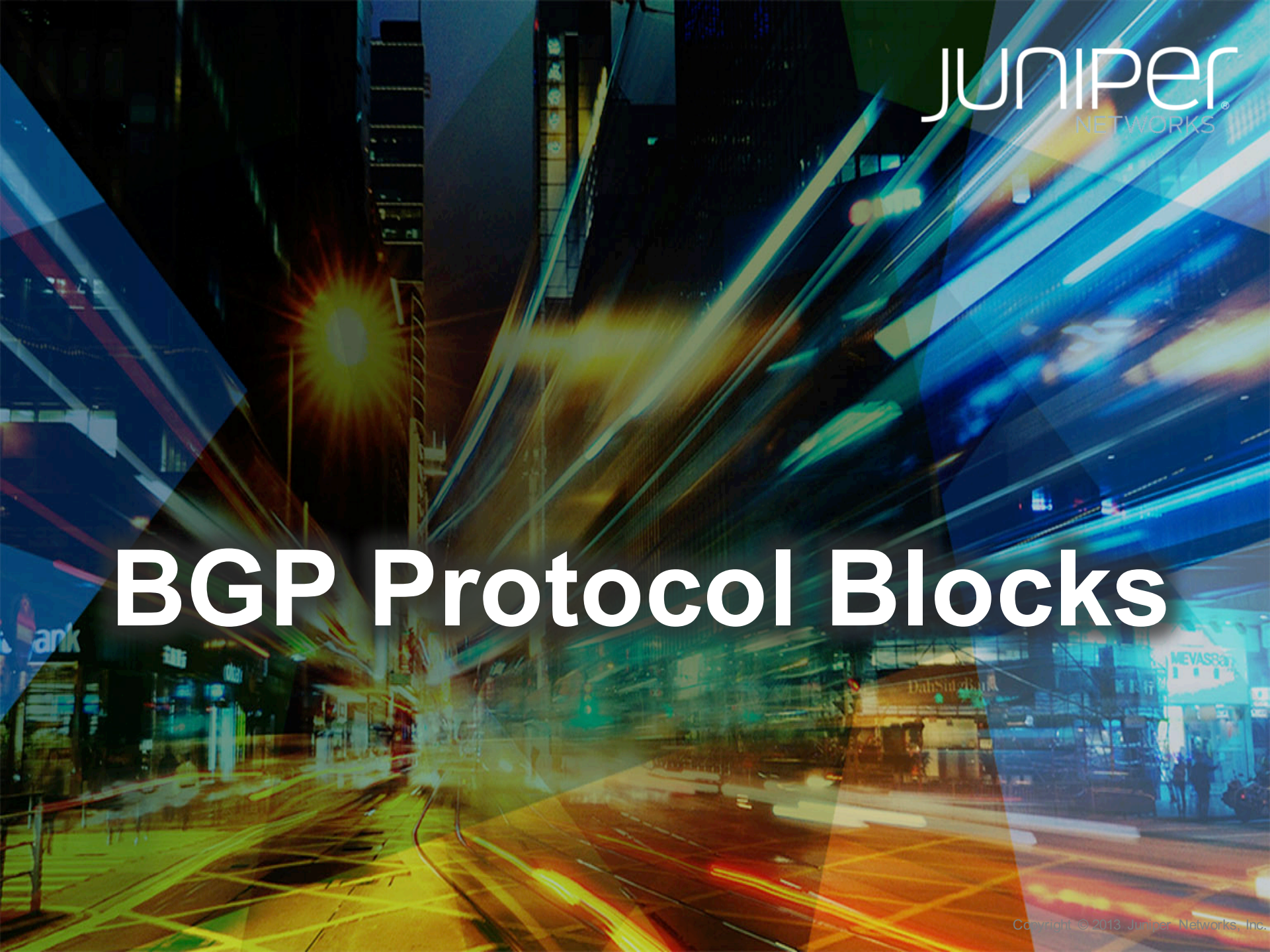
INTRODUCTION

- BGP has evolved from infrastructure to basic connectivity to underlay for advanced service
- Proven protocol since 1989 facilitating Inter domain routing
- Gradually services on the Internet are moving to BGP
  - BGP Multicast VPNs
  - BGP VPLS
  - BGP Flow specifications
- BGP being in Data Center
  - BGP-LU based EPE solution
  - BGP based SPRING solution
  - BGP-LS and BMP for monitoring BGP

BGP Focus Areas

FOCUS AREAS

- BGP Protocol blocks
  - Features to keep the protocol evolving to meet market needs
- BGP based Services
  - BMP, BGP Flow Spec, BGP-LS
  - L3VPN services and associated features
- BGP S&P Initiative
  - Target performance at scale for vRR & PE router
  - Optimize BGP performance for express control plane
- Domain Specific BGP
  - BGP in Data center using BGP-LU based EPE solution
  - BGP-LU based solution for SPRING in Data Center
- Programmable BGP (Concept stage)



BGP Protocol Blocks

PROTOCOL BLOCKS

- Support for long lived GR (BGP LLGR)
  - http://tools.ietf.org/html/draft-uttaro-idr-bgp-persistence-03
- Enhanced BMP capabilities
  - http://tools.ietf.org/html/draft-ietf-grow-bmp-07
- BGP Precision Timer
  - Support for short hold interval timer in BGP keep alives
  - Benefits minimizing NSR dark window during switch over
- BGP 4 bytes AS support

BGP LL GR

- BGP protocol originally designed with focus on correctness
- Increasing use of BGP as a transport for data less associated with packet forwarding
- MPLS tunnels in forwarding reduce the risk of loops
- Persistence complementary to GRES for longer duration failure
- Capability negotiation to exclude routers without capability
- Static environments using BGP as transport
  - BGP used for auto discovery in case of VPLS
  - Filter programming in case of BGP flow spec
  - Support for RTC
- Retain FIB entries with RIB is gone across reboot
  - AFI/SAFIs that do not depend on exchanging BGP state
  - Introduces 3 communities to determine path persistence

BMP: WHAT IS IT?

- Monitoring station to get a dump of routes received from peers
- Provide views for research purposes
- BGP masks implicitly withdraws advertisements

- Add path in principle can provide information on all paths
  - Cost in memory to retain all the routes to a prefix for monitoring
  - Withdraws routes not providing indication of peer down notification

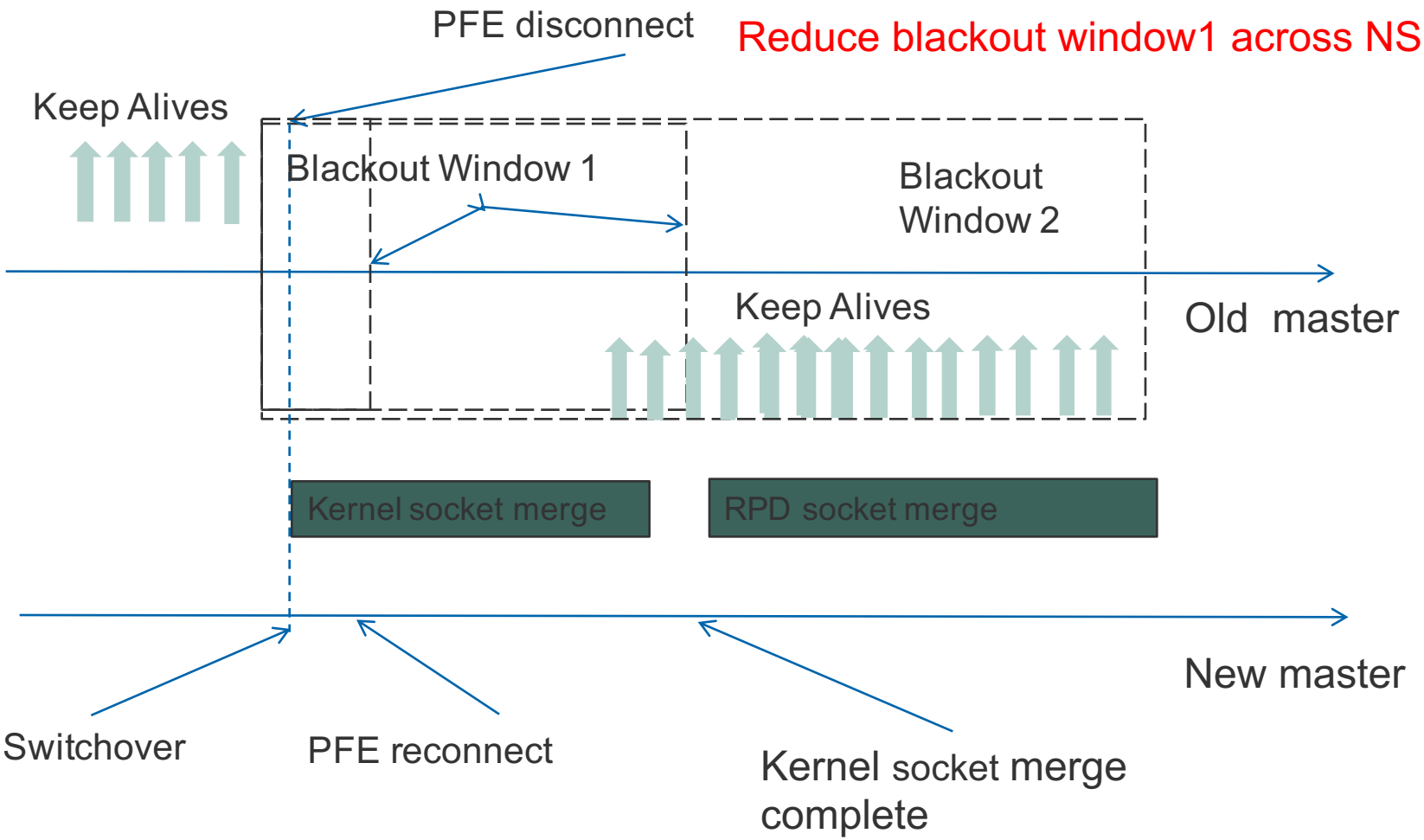
- Provides BGP update messages wrapped in BMP header
  - Timestamp: when route or route withdrawal was received
  - Peer identity: address BGP identifier, RD

- Provides timestamps and operational data beyond routing

BMP: BENEFITS

- BMP protocol provides:
  - Access to the Adj-RIB-In
  - Dump of statistics that can be used by monitoring station
  - Initiation, Peer Down, Peer Up, Route Monitoring and Stats Reports
- Monitoring session does not send message to monitored router
- Monitoring session is tuned to receive messages
- Following the initial dump RM messages are incremental updates
- BMP messages will converge to the correct set of routes

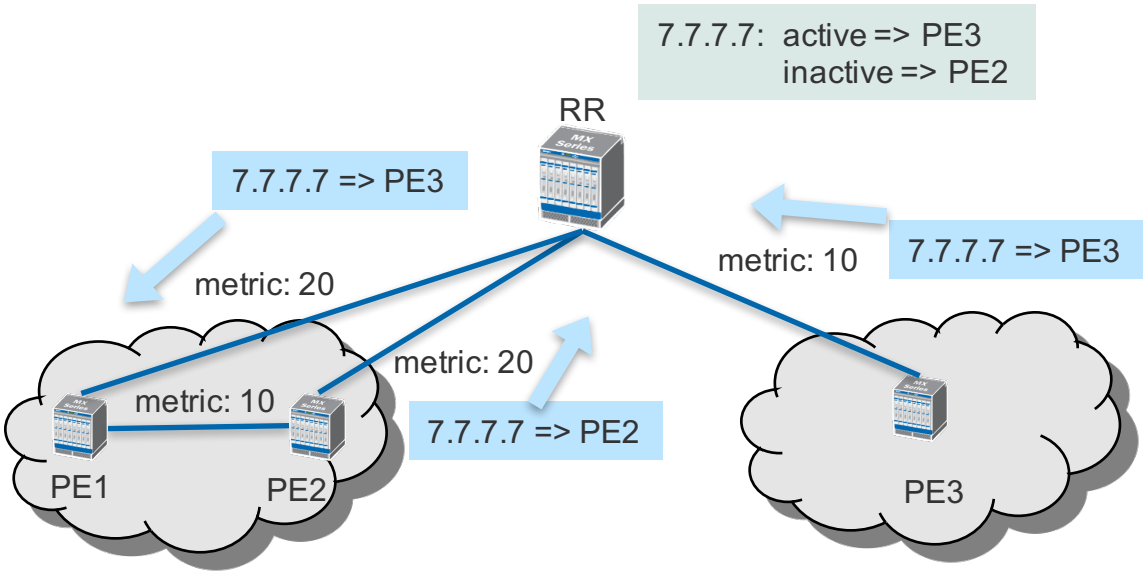
BGP Precision Timer



WHAT IS ORR?

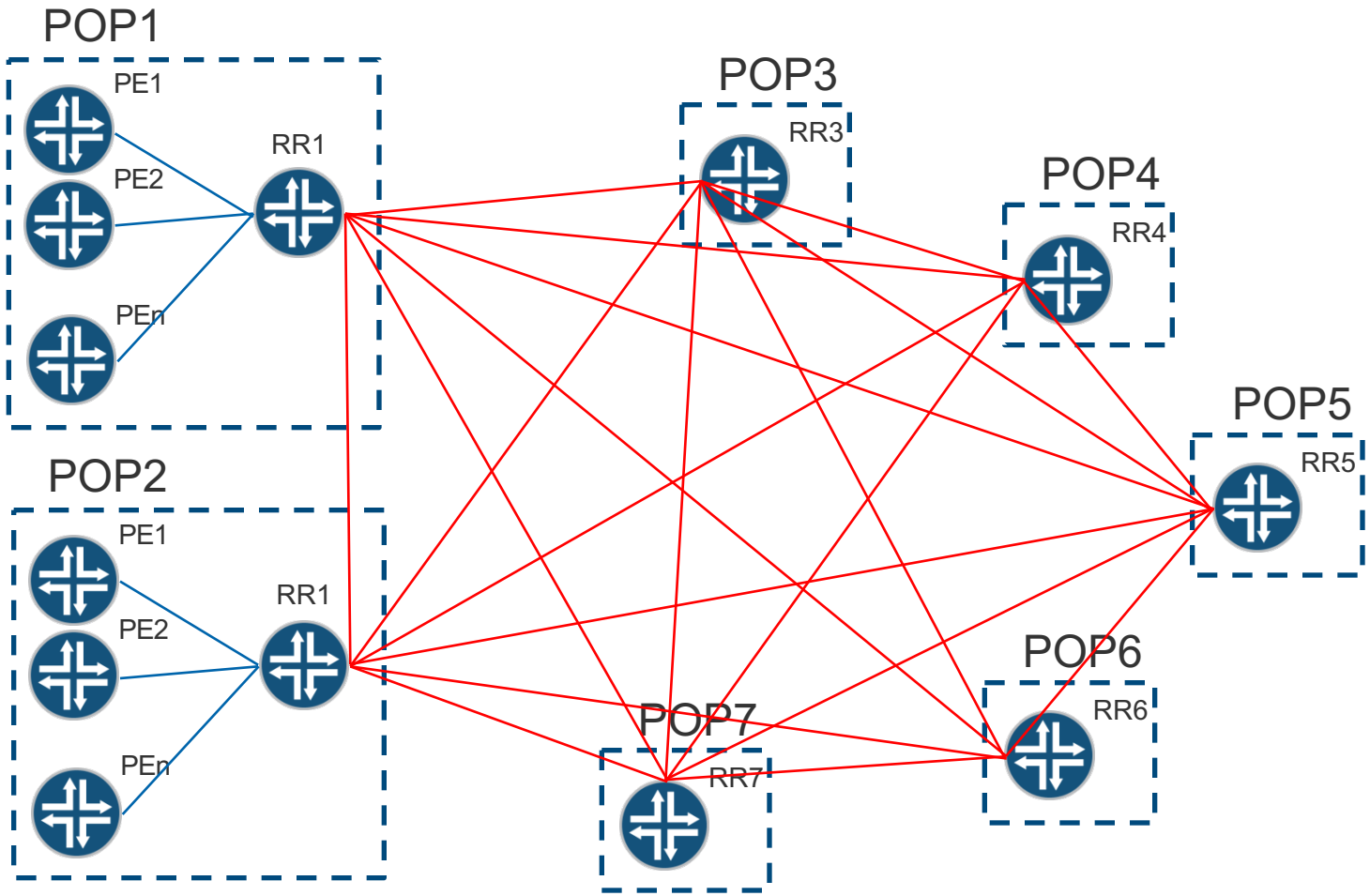
- ORR stands for Optimal Route Reflection
- A route reflector (RR) might receive the same prefix from many egress PEs
- Currently RR chooses the best path from its own perspective (usually the active route) and advertise it to all clients
- ORR wants RR to choose the best path from its client's perspective to advertise to its client
- https://tools.ietf.org/html/draft-ietf-idr-bgp-optimal-route-reflection-08

WHY DO WE NEED ORR?



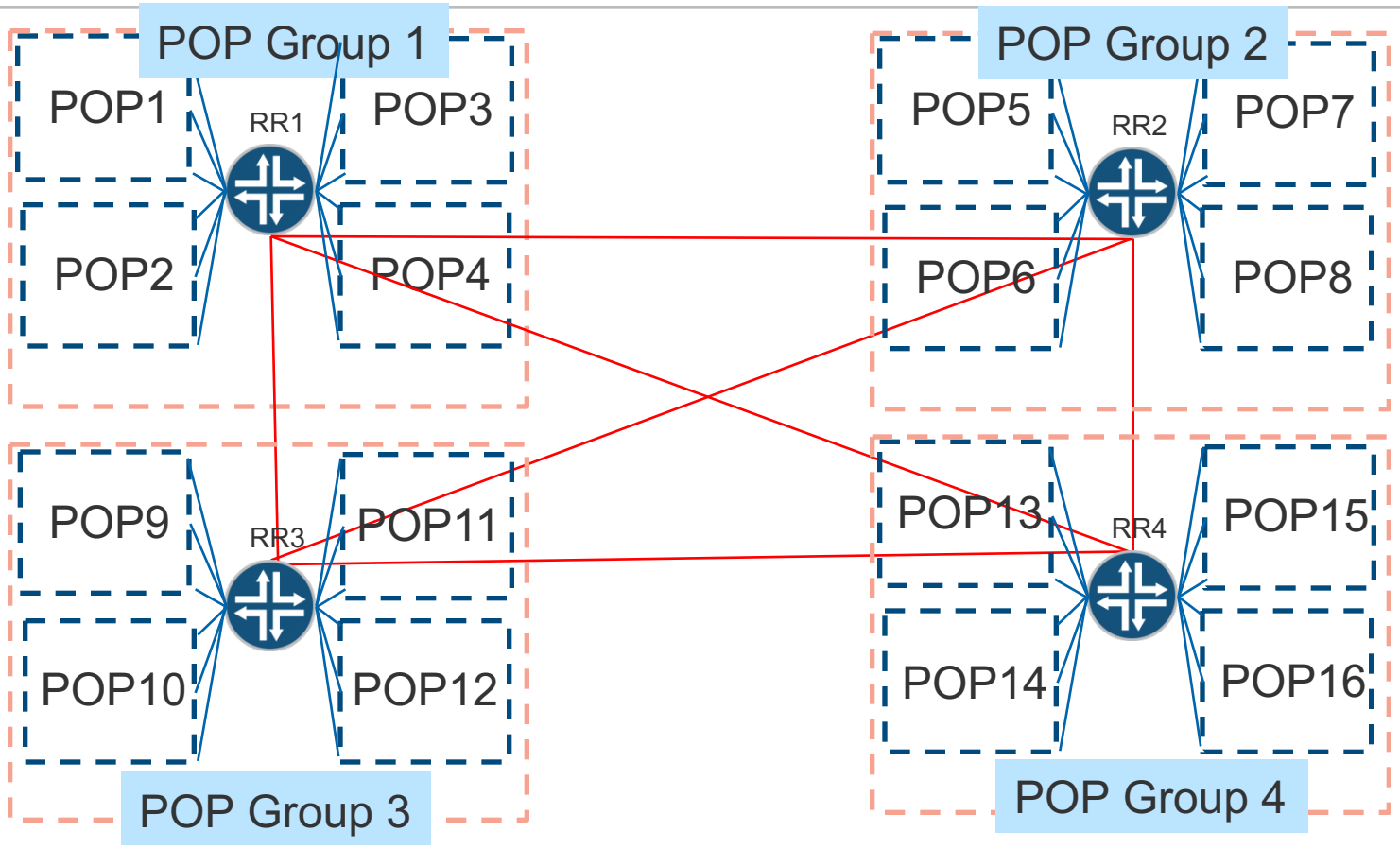
- Without ORR, PE1 will use PE3 as egress PE which is more costly than PE2
- With ORR, RR will reflect the path from PE2 to PE1

ORR REQUEST: CURRENT NETWORK



- There are around 80 RRs, one sitting in each POP

ORR REQUEST: NEW NETWORK



- Multiple POPs are grouped into a POP group and share a single RR (now RR might be geographically far away from some of its clients)
- Initial consolidation ration is 1:4

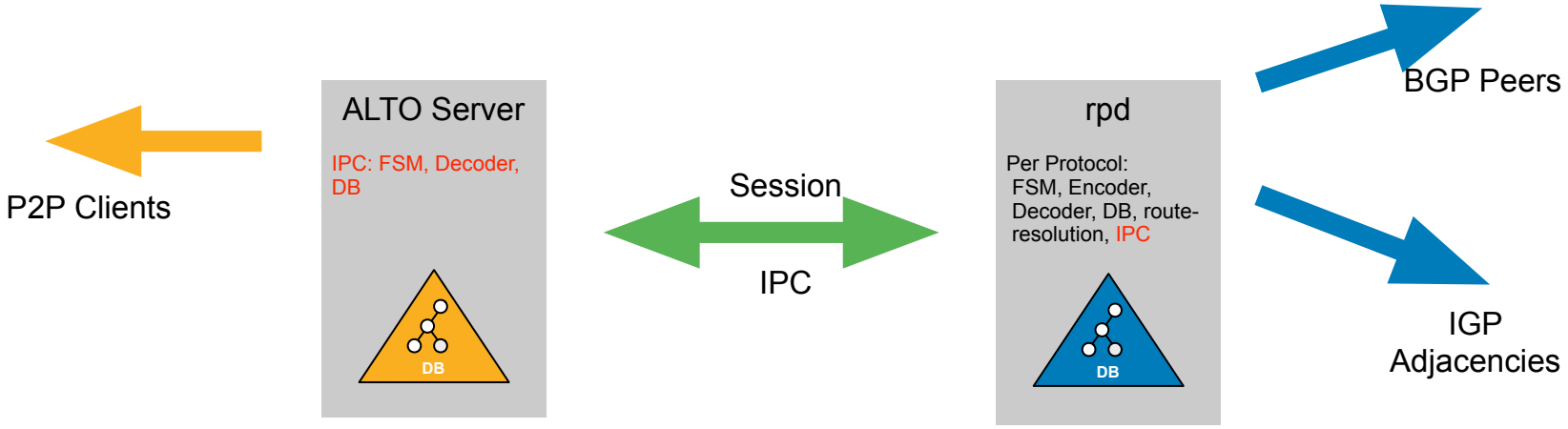
HISTORICAL INSIGHT

External Applications need to access rpd internal data

- BGP Internet (SAFI 1) prefixes
- IGP Topology data

Need to define an API (Session and IPC)

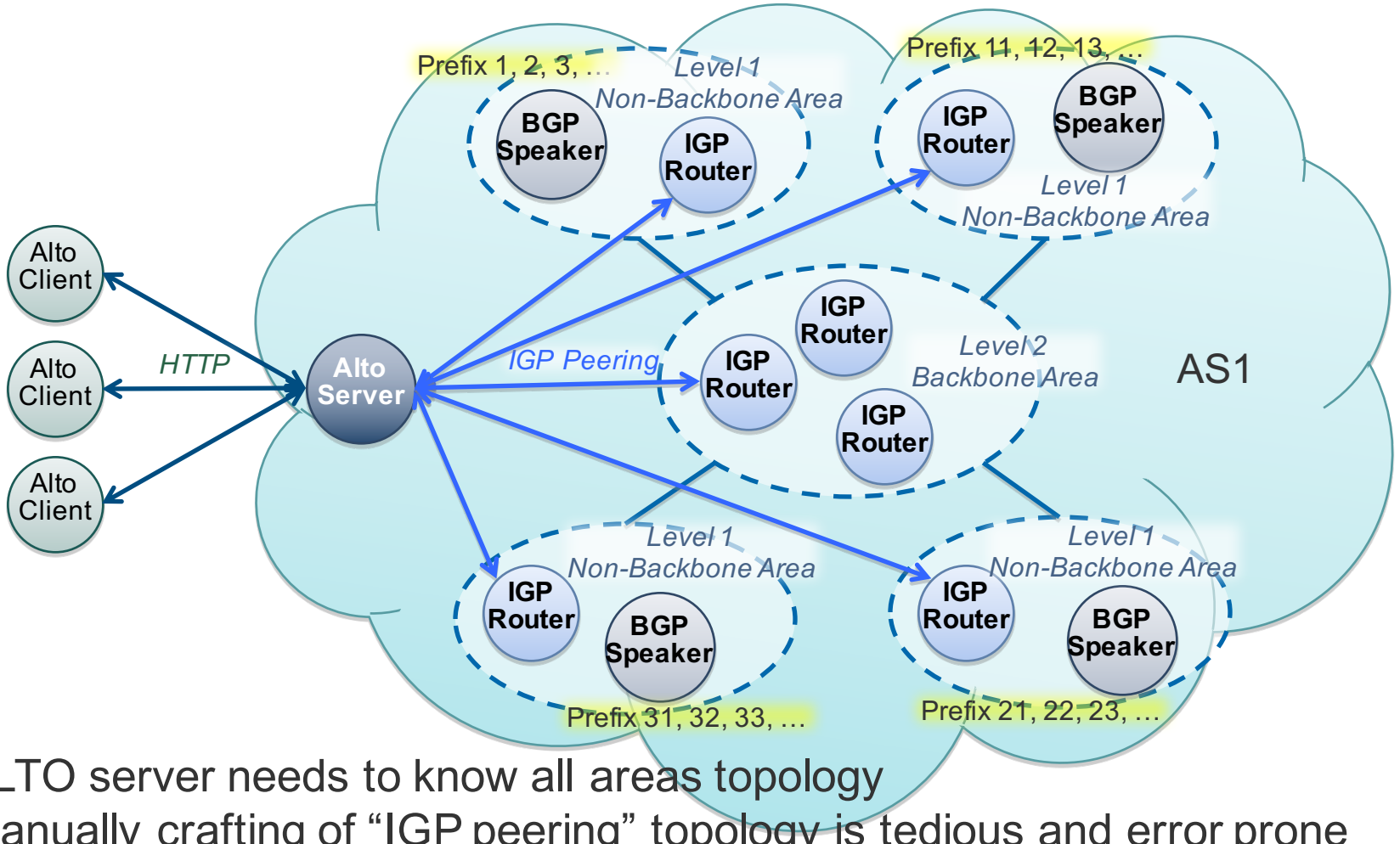
- PUSH or PULL model ?
- IPC format ?
- Would it make sense to “standardize” this API ?



BGP-LS MOTIVATION

- Look across the “fence”
  - “Fence” being IGP area/level or AS boundary
- Gain visibility for application(s) which need ***complete*** topology data
  - ALTO
  - CDNI
  - Inter-{Area, AS} TE
- Unified API, no IGP stack

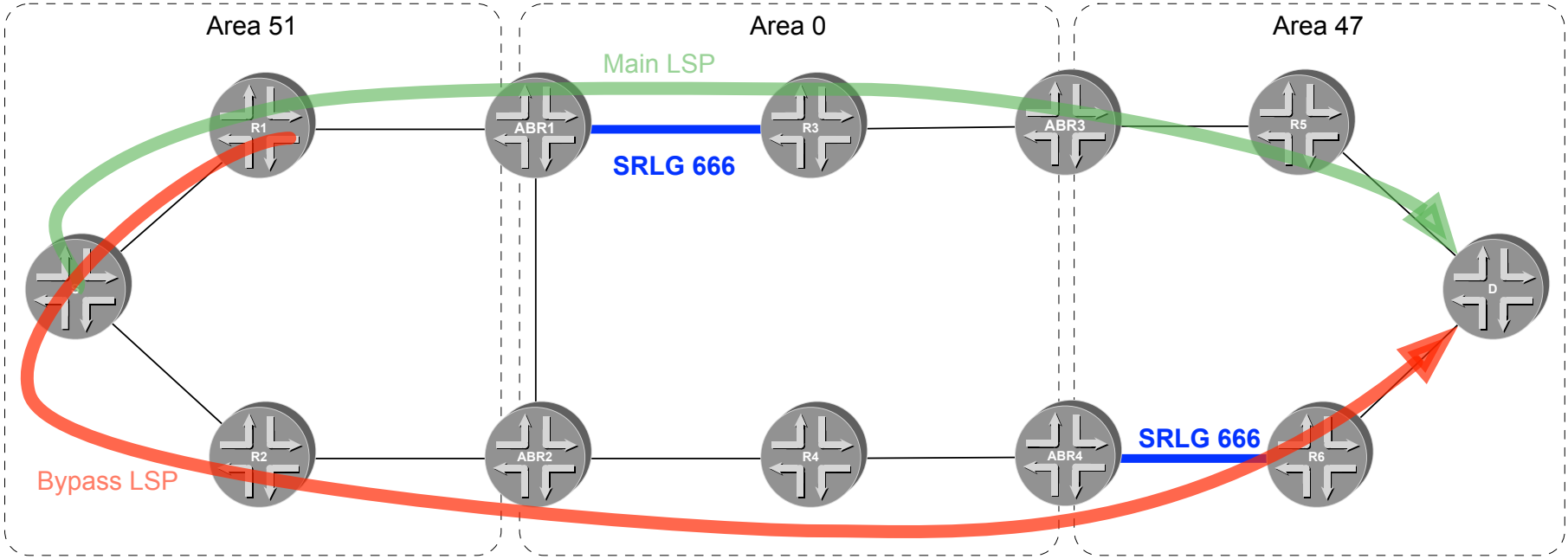
Use case - Multi-area IGP topology



- ALTO server needs to know all areas topology
- Manually crafting of "IGP peering" topology is tedious and error prone

USE CASE – INTER-REGION TRAFFIC ENGINEERING

- RSVP Loose hop expansion has practical deployment limits
- Vanilla RSVP has no crank-back in case it “sees” new information that it did not see at previous hops. (and RFC4920 has a lot of caveats)



BGP-Flow Spec

“FLOW”-BASED BGP NLRI

- Specific information about a flow can now be distributed using this BGP NLRI:
  - AFI/SAFI = 1/133: Unicast Traffic Filtering Applications
  - AFI/SAFI = 1/134: VPN Traffic Filtering Applications
- Route “prefix” contains <destination, source, ports>
  - E.g. 10.0.1/24,*,proto=6 (TCP),port=80
- Flow routes are kept in a separate routing table “inetflow.0”
- The contents of this table are applied as a ingress forwarding-table filter on JUNOS routers
- Flow routes are automatically validated against unicast routing information or via routing policy framework.

WHAT IS IN THE BGP FLOW SPEC NLRI?

- A Flow Specification NLRI is defined which may include several components in order to identify particular flows
  - The NLRI field of the MP_REACH_NLRI and MP_UNREACH_NLRI is encoded as a 1 or 2 octet NLRI length field followed by a variable length NLRI value.
  - The NLRI length is expressed in octets

```
+-----------------------------+
  | length (0xnn or 0xfn nn) |
+-----------------------------+
   | NLRI value (variable) |
+-----------------------------+
```

Type 1 - Destination Prefix

Type 2 - Source Prefix

Type 3 - IP Protocol

Type 4 – Source or Destination Port

Type 5 – Destination Port

Type 6 - Source Port

Type 7 – ICMP Type

Type 8 – ICMP Code

Type 9 - TCP flags

Type 10 - Packet length

Type 11 – DSCP

Type 12 - Fragment Encoding

FLOW SPEC: JUNOS

- Junos supports flow spec for IPv4 and VPNv4
- Flow route installed in flow route table Instance-name.inetflow.0
- Criteria for validating route with unicast routing table
- No-validate to bypass and introduce of operator specific policy
- Once route is added into the inetflow
  - Installed to list of firewall filters in a kernel
  - VPN capability to install flow routes
- Provides framework for
  - match criteria defined on n-tuple match
  - Action criteria defined in RFC 5575 with extensibility built in

HOW DOES BGP FLOW SPEC HELP?

- Flow spec addresses the limitations of existing solutions by allowing the “flow”-based NLRI to convey additional information about traffic filtering rules for traffic that should be discarded

- Since a new address family is defined, filtering information is now separated from the routing information (and in fact this information is kept in a separate RIB: *instance-name*.inetflow.0)

- Provides a tool for Network Operators to quickly react to DDOS attacks, saving valuable time between identification of attack and implementation