IDR Document status
IETF 89

Susan Hares and John Scudder

Idr chairs

Note Well

Any submission to the IETF intended by the Contributor for publication as all or part of an IETF Internet-Draft or RFC and any statement made within the context of an IETF activity is considered an "IETF Contribution". Such statements include oral statements in IETF sessions, as well as written and electronic communications made at any time or place, which are addressed to:

Which is address to:

- The IETF plenary session
- The IESG, or any member thereof on behalf of the IESG
- Any IETF mailing list, including the IETF list itself, any working group or design team list, or any other list functioning under IETF auspices
- Any IETF working group or portion thereof
- Any Birds of a Feather (BOF) session
- The RFC Editor or the Internet-Drafts function
- All IETF Contributions are subject to the rules of RFC 5378 and RFC 3979 (updated by RFC 4879).
- Statements made outside of an IETF session, mailing list or other function, that are clearly not intended to be input to an IETF activity, group or function, are not IETF Contributions in the context of this notice.

Note Well

Please consult RFC 5378 and RFC 3979 for details.

A participant in any IETF activity is deemed to accept all IETF rules of process, as documented in Best Current Practices RFCs and IESG Statements.

A participant in any IETF activity acknowledges that written, audio and video records of meetings may be made and may be available to the public.

Stats of documents

New Documents:

- draft-ietf-idr-te-lsp-distribution-00
- draft-ietf-idr-te-pm-bgp-00
- draft-ietf-idr-mdcs-00
- draft-ietf-idr-mdrs-00
- draft-ietf-idr-as-migration-00

Status of documents

- Pass WG LC
  - draft-ietf-idr-aigp
  - draft-ietf-idr-bgp-enhanced-route-refresh
  - draft-ietf-idr-last-as-reservation (BCP/Proposed)
  - Draft-ietf-idr-ls-distribution (for Early allocation)
- WG LC planned after IETF 89
  - draft-ietf-idr-ls WG LC
  - draft-as-migration WG LC (Chris Morrow Shepherd)

Moving to RFC or Flush

- What does it take to go RFC from WG Draft
  - 2 implementations with Experience
  - Send a note to co-chairs/list requesting WG LC
  - Be ready to answer mail
- Idr flushing old documents
  - We will begin to flush old IDR WG by sending email to authors for Status
  - No response in March 2014 == Flush

Questions

- Errata building on IDR RFCs
- Do you want to revise Base RFCs to pick up Errata and changes?

REVISED ERROR HANDLING FOR BGP UPDATE MESSAGES -06

John Scudder

March 6, 2014

COCONSPIRATORS

Enke Chen

Pradosh Mohapatra

Keyur Patel

BRIEF RECAP OF ERROR HANDLING

BGP “classic” resets session upon encountering an error

- Nice for formal correctness
- Not so nice for network operations

draft-ietf-idr-error-handling revises error handling in cases where the NLRI can be found

- Mostly, treat-as-withdraw
- In some cases, attribute discard
- If NLRI can’t be found or are themselves corrupt, good old session reset

CHANGES IN VERSION -06: ATTRIBUTE FLAGS

Earlier versions mandated that if the attribute flags (optional, transitive) conflict with the attribute type code, the flags should be “fixed”.

- Eric Rosen pointed out that this can be unsafe. (Example: re-setting a flag to transitive might allow a broken attribute to propagate)
- Besides, this is quite a notional class of error
- Draft updated to remove the “fix” behavior and instead define it as malformation just like any other.
- Default: treat-as-withdraw (spec for attribute can override)

CHANGES IN VERSION -06: PICKIER NLRI PARSING

Numerous reviewers pointed out that the sanity of the error-handling approach depends on being able to reliably dig out the NLRI.

It turns out RFC 4271 mandates the NLRI must be syntax-checked, but doesn't define what this means.

Toward this, added some more text

- S 3.1: Error if encoded attributes either exceed overall attribute length (overrun) or fall short (underrun). Mandates that implementations must rely on overall attribute length. Treat-as-withdraw if error.
- S 3.2: Lengths of individual NLRI must be sane. NLRI must not overrun enclosing object (BGP PDU for old-school IPv4, MP_{UN} REACH path attribute for new-style). Session reset if error.

PENDING CHANGES (FOR -07)

For S 3.2 (NLRI syntax) planning to add two more error conditions for MP NLRI:

- Bad attribute flags
- Bad attribute length

As with other NLRI errors, these would cause a session reset.

Thanks to Tony Przygienda

CHANGES IN VERSION -06: RR PATH ATTRIBUTES

Route reflection path attributes were omitted pending update of RR base spec

- Update not (yet?) done, so pulled this in to error-handling

Discard RR path attributes if received from an external peer

Otherwise, treat-as-withdraw if malformed length

DISCUSSION AND NEXT STEPS

"It's basically done"

- Issue -07, update implementations, WGLC

Now would be a very good time to review the draft

- In particular, any with lingering worries about the basic sanity of the approach should speak now or forever hold their peace

THANK YOU

BGP Link-State Information Distribution Implementation Report

draft-gredler-idr-ls-distribution-impl-00

Hannes Gredler hannes@juniper.net
Balaji Rajagopalan balajir@juniper.net
Saikat Ray sairay@cisco.com
Manish Bhardwaj manbhard@cisco.com

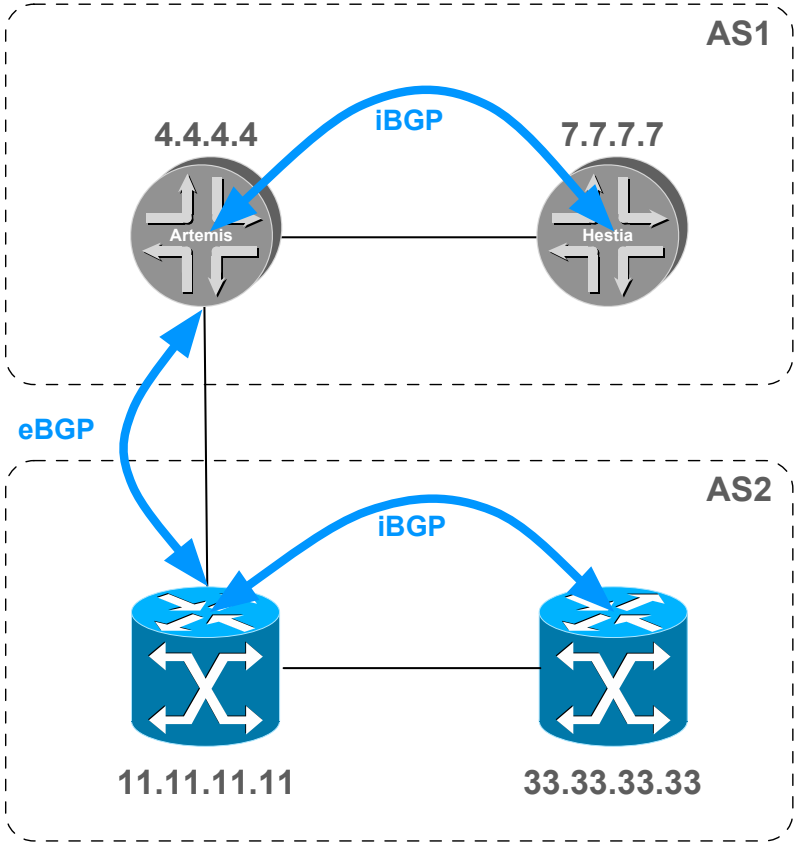
Rationale

- Interop testing of BGP-LS
  - Based on draft-ietf-idr-ls-distribution-04
- Software
  - Cisco IOS-XR (Engineering internal Build)
  - Juniper JUNOS (Engineering private build, 14.2 Base)
- Dec 2-5th, Sunnyvale, Juniper premises

- Verify all the BGP {Update, Withdraw, Refresh, Notification} machinery for the new NLRIs
- Verify correct/consistent generation of LS NLRIs
- Verify all 20+ on-the-wire LS Attributes (encoding, endianness)
- Verify handling of ***unknown*** TLVs (Store and forward)

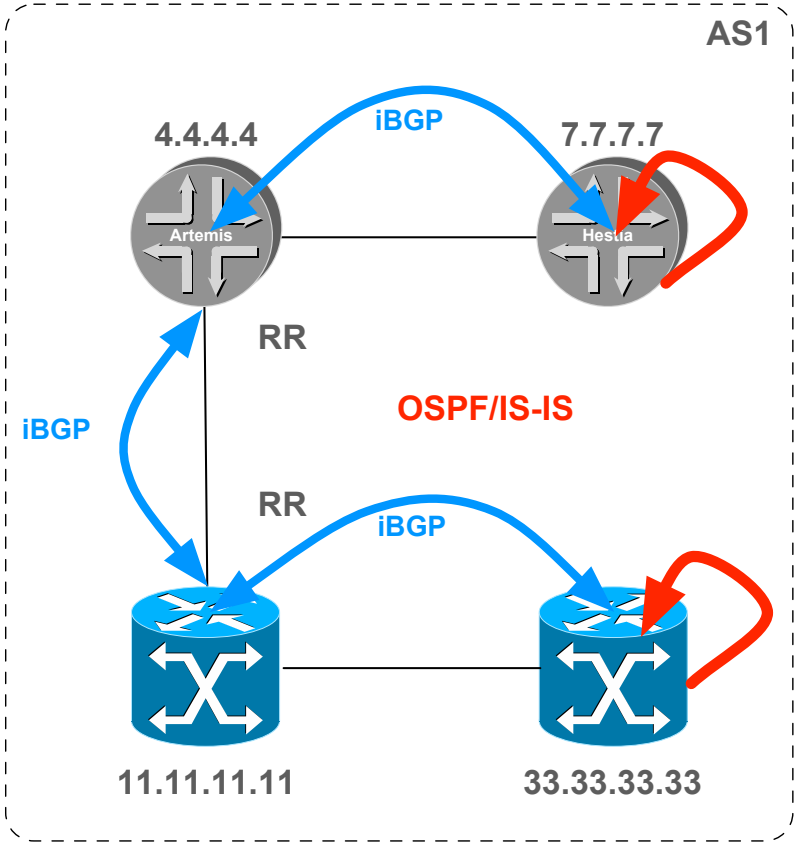
Testsetup #1

- Purpose
  - Check Propagation of <link-state> through a {iBGP, eBGP, iBGP} path
  - Verify proper encoding of BGP-LS attributes
  - Verify update/withdraw logic (key change)



Testsetup #2

- Purpose
  - Check Propagation of <link-state> through a iBGP path
- Check ***consistent*** export of TE/IGP into BGP LS NLRIs



Issues found

- Protocol issues
  - Unnumbered/Numbered Link generation
    - Need Clarification when IDX is key or attribute
- Implementation issues
  - AFI encoded internally as uint8
    - BGP-LS AFI/SAFI is 16388/71 (!)
    - Route refresh broken
  - Endianness for Bandwidth related data
  - Inconsistent Keys for OSPF “Pseudonodes”

draft-gredler-idr-ls-distribution-impl-00

- Document contains list of all tested
  - NLRIs
  - Send/Receive/Originate TLVs
- Plan to include other Implementations (RR, FLOSS implementation)

Impressions

Next Steps

- Release draft-ietf-idr-ls-distribution-05
  - https://github.com/hannesgredler/draft-ietf-idr-ls-distribution
- Questions ?
- Adoption as a WG item ?

Traffic Engineering Database Dissemination for Hierarchical PCE scenarios

CCAMP WG, IETF89,

London

draft-lopez-pce-hpce-ted-01

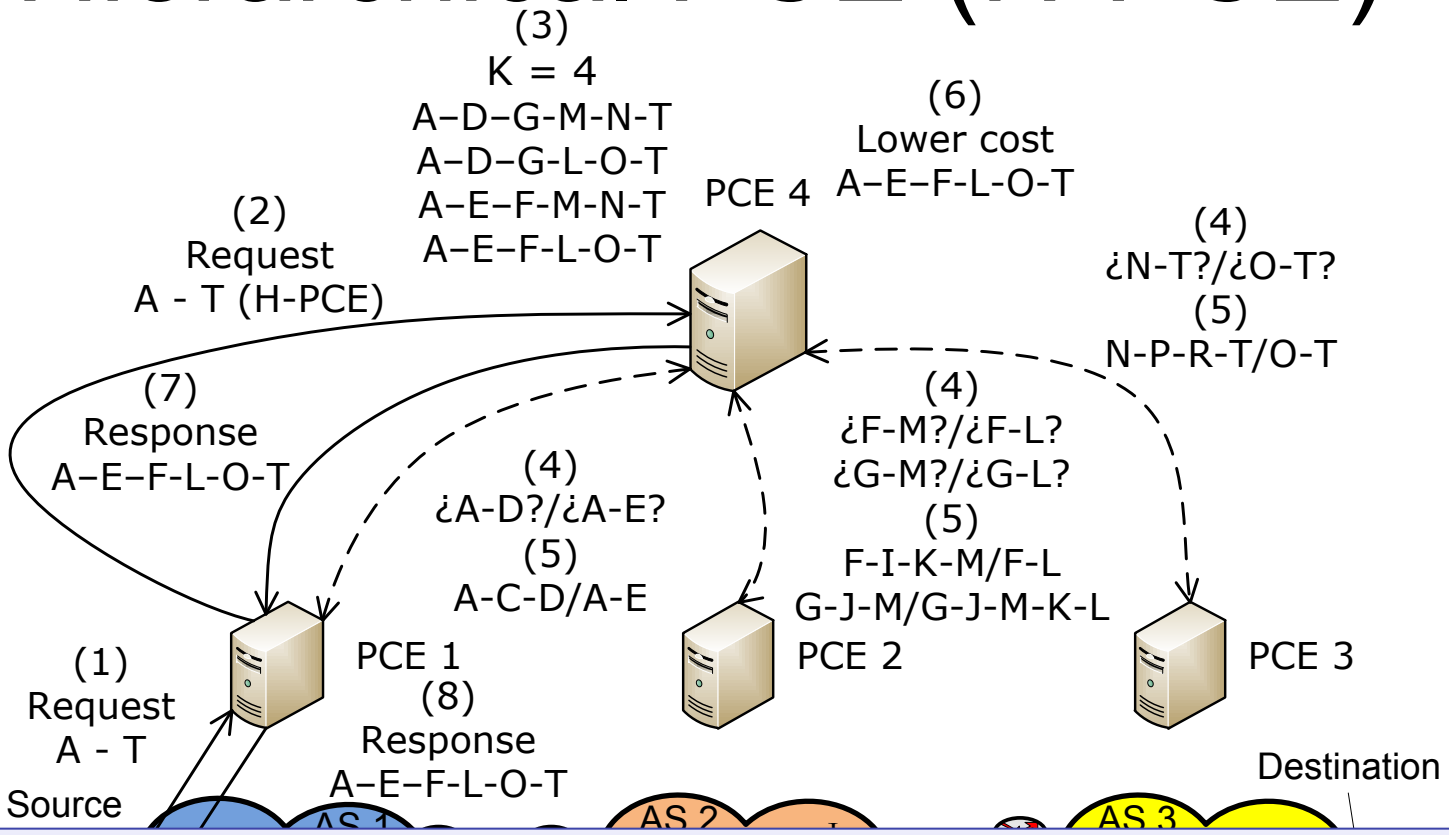
Victor Lopez <vlopez@tid.es>

Oscar Gonzalez de Dios <ogondio@tid.es>

Daniel King <daniel@olddog.co.uk>

Stefano Previdi <sprevidi@cisco.com>

Hierarchical PCE (H-PCE)



Problems:

- Solve the problem with the sequence domains.
- Domain sequence is computed with the topological information of the parent PCE.
- Better network resource utilization than BRPC or Per-Domain approaches.

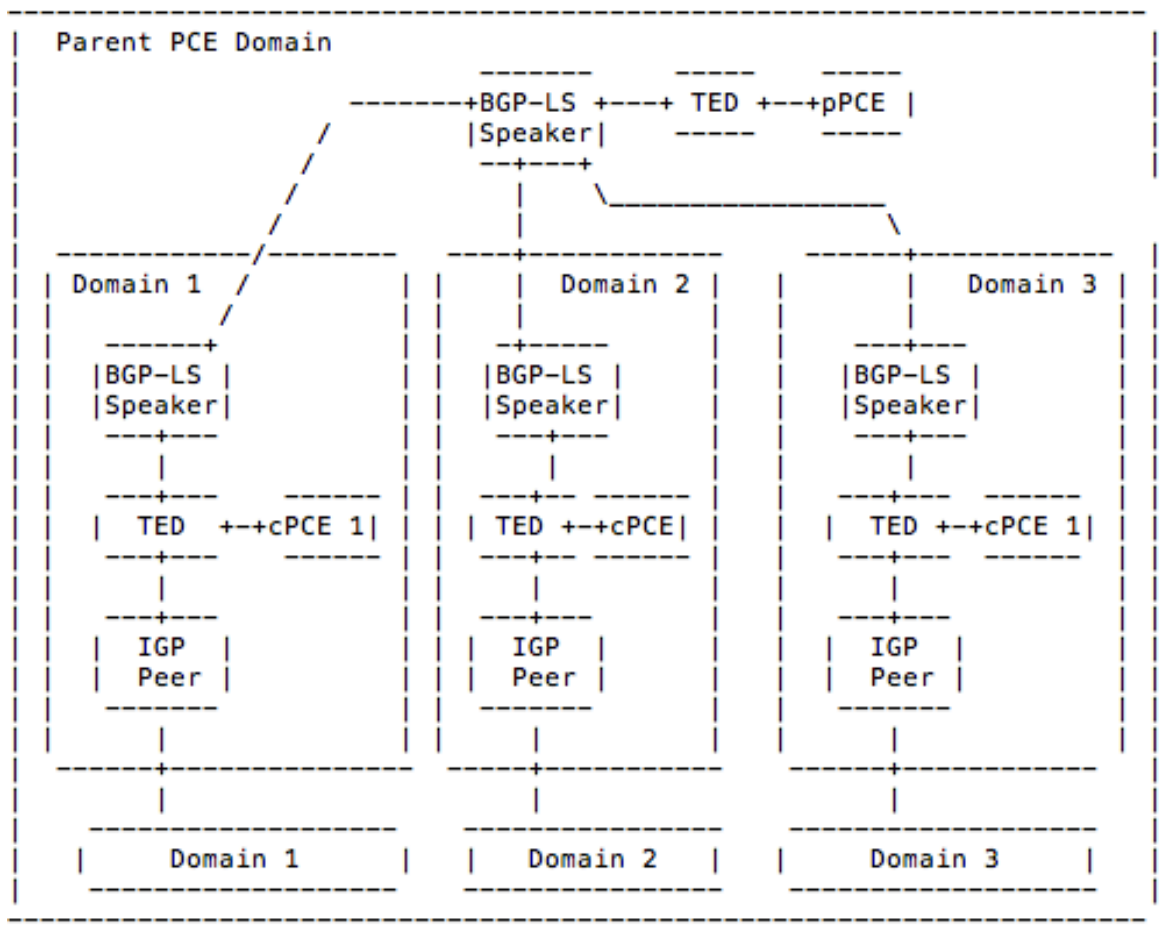
Motivation for the draft

- H-PCE Architecture (RFC6805)
  - Proposal to solve the multi-domain path computation by means of cooperation among different PCEs.
  - Solution draft for H-PCE (draft-ietf-pce-hierarchy-extensions-00)
    - Focus on computation procedures and PCEP protocol extensions.
- Unanswered Questions in the Path Computation Element Architecture (draft-ietf-pce-questions) presents the topology dissemination as an open issue.
- Procedure to build and populate the parent PCE Traffic Engineering Database (TED) is still an open issue.
- Goal of this draft
  - Analyse how topology dissemination mechanisms may be used to provide TE information between Parent and Child PCEs
- Not a goal of this draft
  - Solve the Internet via exposure of all internal domain topoloogies!

H-PCE Topology Dissemination Options

- What needs to be provided?
  - Inter-domain links
  - Edge-to-edge "virtual" TE links created out of (potential) LSPs
- How to provide?
  - Static configuration
  - Join an IGP instance
  - Via PCEP Notifications
  - Separate IGP instance
  - Northbound distribution of TE information (BGP-LS)

H-PCE with BGP-LS architecture



Open issues

- Is BGP-LS the way forward?
- Mapping of OSPF-TE / IS-IS-TE

Next Steps

- Continue to investigate and prototype
- Trigger discussion on which mechanisms should be used and why
  - Application and scenario based?
  - Scalability?
- Receive feedback

BGP Link-State extensions for Segment Routing

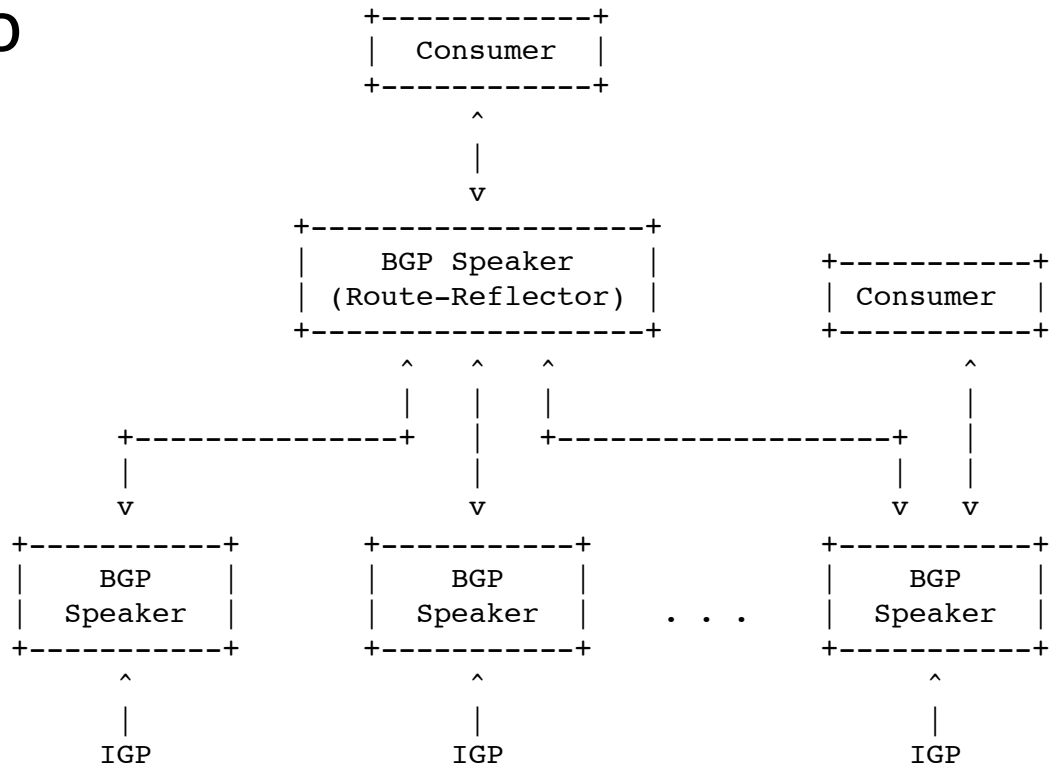
Hannes Gredler
Saikat Ray
Stefano Previdi
Clarence Filsfils
Mach(Guoyi) Chen
Jeff Tantsura

Introduction

- Segment routing
  - A flexible, scalable way of doing source routing
- Segments are “instructions”
  - “Go to node N via shortest path”, “use link L”, etc.
  - Each segment is identified by a “Segment ID” (SID)
- IGPs advertise the <Segments, SID>
- Ingress node adds SID stack to data packets to determine the packet path
  - Per-flow state is only at the ingress node
  - SIDs map to MPLS labels for MPLS data plane

Need for BGP LS

- Segments are used to set up end-to-end paths (topological and services)
- Paths may span IGP areas, or even ASes
  - Segment information from one IGP area alone does not work



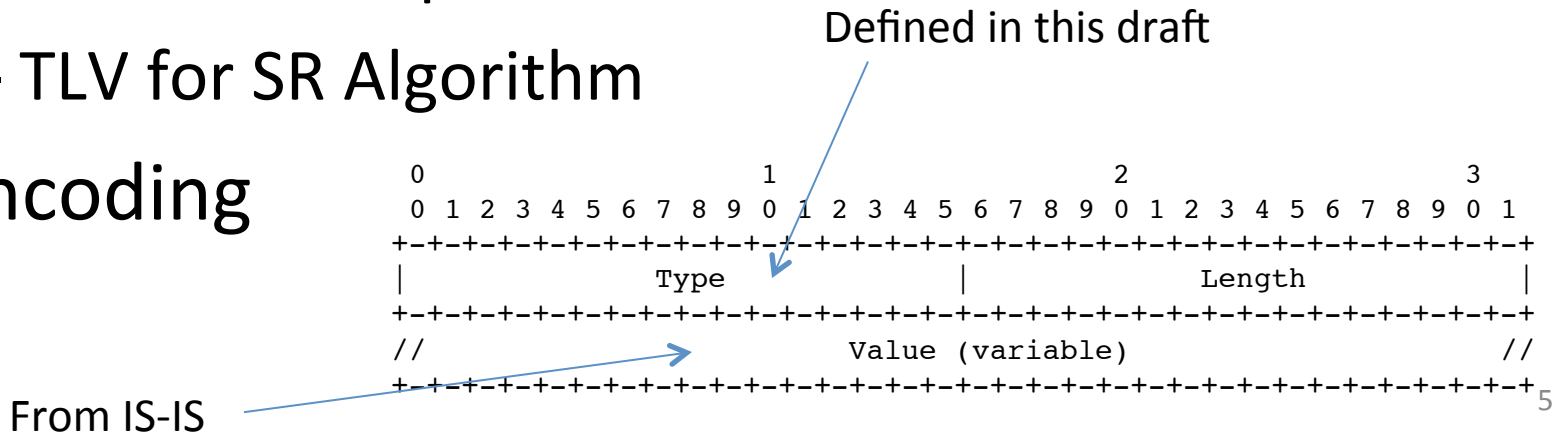
- BGP LS collects LSDB from all IGP areas
  - BGP LS provides visibility into segment information required for building end-to-end paths

BGP LS

- BGP LS models the IGP network as a collection of three types of objects: (i) Nodes, (ii) Links (ordered pair of nodes) and (iii) Prefixes
- Each object is encoded as BGP object
  - The "key" portion of the objects is the NLRI
  - The rest of the properties of the object are in the BGP-LS attribute
    - BGP-LS attribute is a set of TLVs; easily extended
- Approach: Add the segment information in the BGP-LS attribute of the corresponding object

Segment routing TLVs

- SR information TLVs are defined in I-D.previdi-isis-segment-routing-extensions
  - TLV for Prefix-SID
  - TLV for Adjacency-SID between two nodes as well as between nodes in a LAN
  - TLV for SID/Label binding for advertising paths from other protocols (and their optional ERO)
  - TLV for SR Capabilities
  - TLV for SR Algorithm
- Encoding



SR TLVs in Node Attribute

- The following SR TLVs are in the node attribute (BGP-LS attribute that is added to a node NLRI)

TLV Code Point	Description	Length	IS-IS SR TLV/sub-TLV
1033	SID/Label Binding	variable	149
1034	SR Capabilities	variable	2
1035	SR Algorithm	variable	15

SR TLVs in Link Attribute

- The following TLVs are added to a link attribute

TLV Code Point	Description	Length	IS-IS SR TLV/sub-TLV
1099	Adjacency Segment Identifier (Adj-SID) TLV	variable	31
1100	LAN Adjacency Segment Identifier (Adj-SID) TLV	variable	32

SR TLVs in Prefix Attribute

- The following TLVs are added to a Prefix attribute

TLV Code Point	Description	Length	IS-IS SR TLV/sub-TLV
1158	Prefix SID	variable	3

What next

- WG document
- Add more details on SID/label binding TLV
- Prototype implementations

Constrained Route Distribution with Multiple Address Families

Saikat Ray
Arjun Sreekantiah
Keyur Patel

Prologue

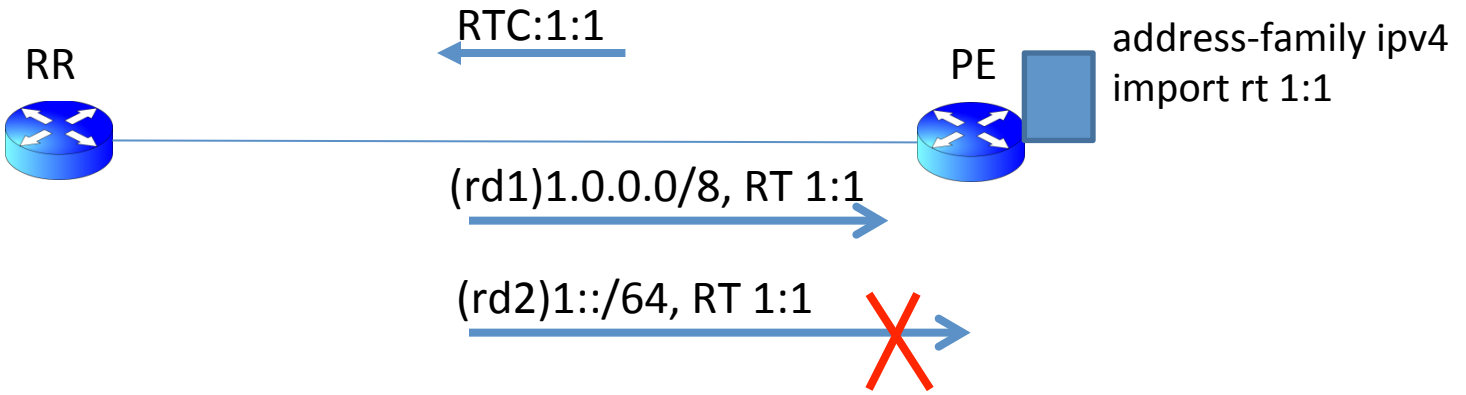
- draft-ray-idr-route-constrain-scope-00 presents two problems
- Problem 2 and suggested solution are presented here
  - We will make it a into a different draft
- Problem 1
  - Seeking input from WG about if the problem is useful to work on

Route target Constrain

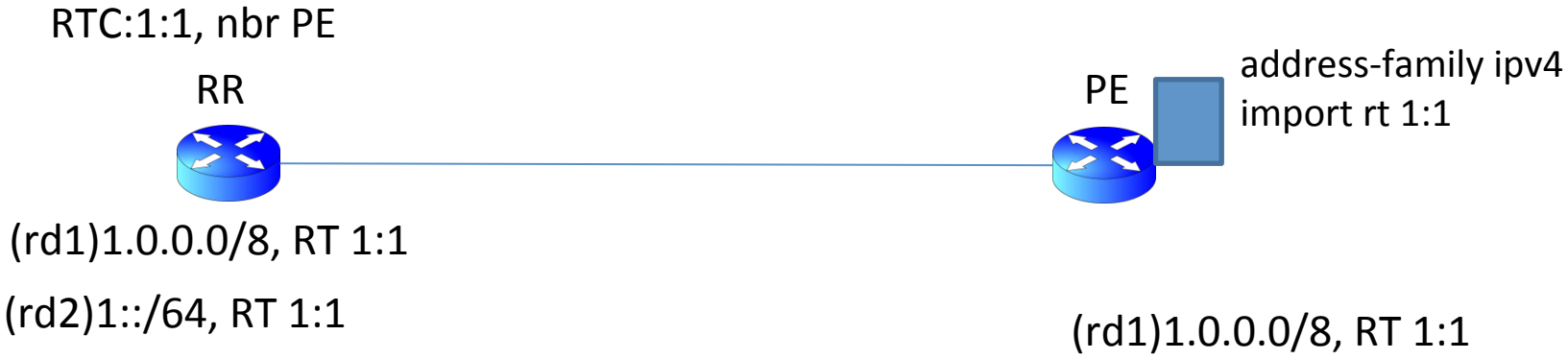
- RR sends all VPN routes to a PE for a negotiated address-family (say, 1/128)
- PE only keeps the routes whose RT is imported by at least one VRF
  - RR didn't have to send the other routes
- Optimization
  - Let PE tell RR which RTs the PE is "interested" in
  - RR sends only the matching routes to PE
- Address-family 1/132 is used for exchanging RTs of "interest"
  - NLRI encodes the RT of interest

RTC and Multiple VPN Address-family

- An RTC NLRI signals sender's interest in routes with the given RT from all (VPN) address-families
  - If the PE does not need the route, it drops them
  - E.g., no VRF with IPv6 address-family importing 1:1 (but there are VRFs with IPv6 address-family so that PE and RR negotiates VPNv4 Unicast address-family)

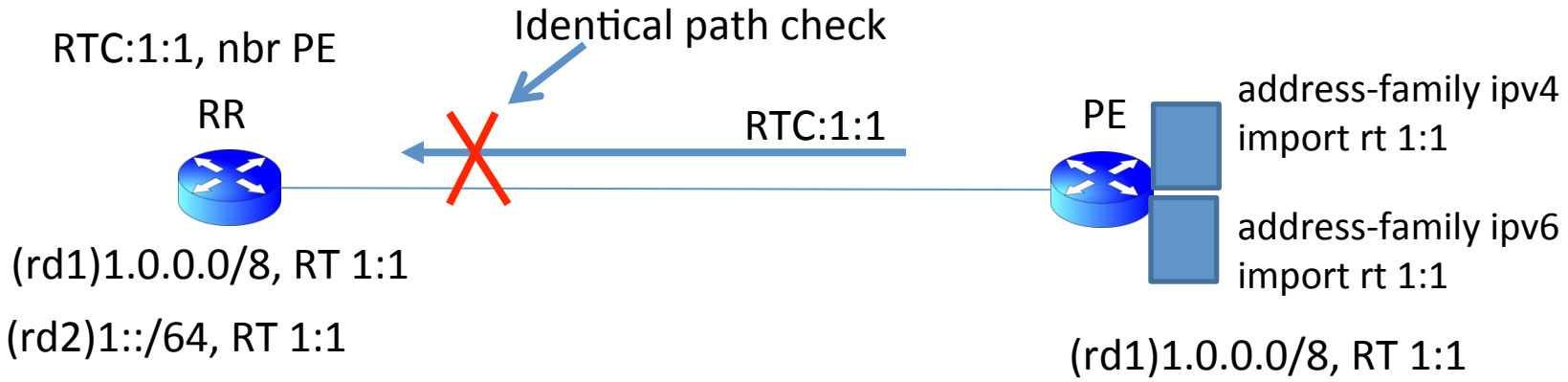


Problem: Incremental VRF addition



- Steady-state
  - PE has a VRF with IPv4 that imports RT:1:1, but no VRF with IPv6 that imports RT:1:1
  - RR has v4 and v6 routes with RT:1:1, and RTC:1:1 from PE
  - PE has v4 route with RT:1:1 from RR, but no v6 route with RT:1:1

Problem: Incremental VRF addition

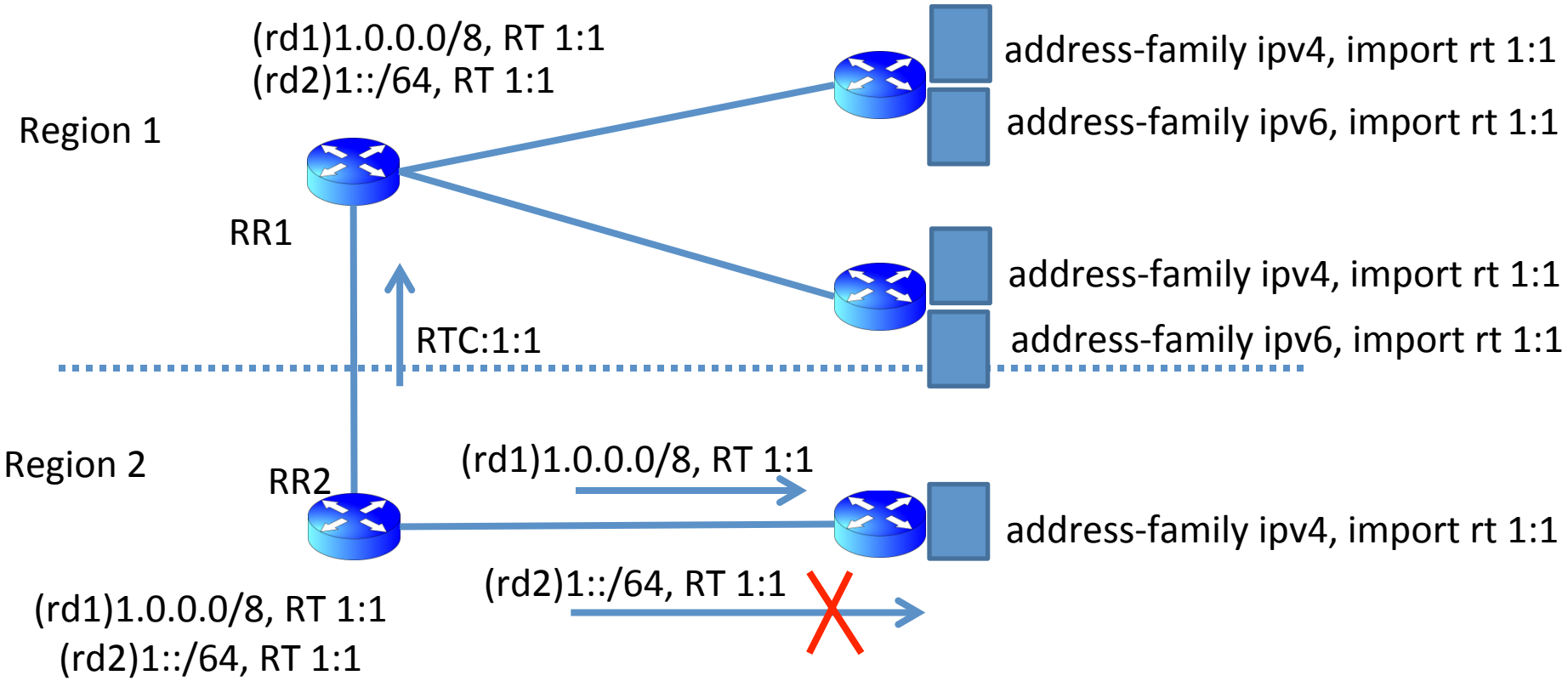


- VRF addition
  - A VRF with IPv6 that imports RT:1:1 is added to PE
  - PE sends RTC:1:1 to RR
  - RR drops it due to identical path check
- Work around
  - PE needs to send a route-refresh for 2/128 to RR
  - RR will send all VPNv6 routes to PE whose RTs match some RTC NLRI

Proposed Rules

- Rules
  - A BGP speaker that receives an identical RTC path from a neighbor must treat it as equivalent to a route-refresh request for the given RT for all (VPN) address-families.
  - If a new (VPN) address-family is negotiated between two BGP speakers without a session reset (e.g., using dynamic capability, or using multi-session feature), then existing RTC NLRIs' scope must be extended to the new address-family
- No changes to the protocol on the wire
- Why do we want to standardize this?
  - Prudent BGP implementations use identical path check; this would mandate a change in the behavior for RTC
  - The PE needs to know whether the RR supports this or not
    - A CLI knob on the RR
    - Indicate support in the Capability
      - Maybe we can use a reserve bit to indicate this for 1/132 MP capability?

Unnecessary route retention



- Region 1 RR has clients that require both IPv4 and IPv6 routes with RT:1:1
- Region 2 RR does not have any client that need IPv6 route with RT:1:1
- Region 2 RR still retains IPv6 routes with RT:1:1 and advertises them to its clients (who sent RTC:1:1)

Unnecessary route retention

- This is a problem if
  - Different address-families in different VPNs use the same RT
    - Not the usual operational practice
  - Not all sites have the same set of address-families
    - Regional differences - E.g., IPv6 in only one region
    - Transitions – e.g., IPv6 turned on "temporarily" on some sites
- Previous proposal to solve this add safi in NLRI
  - Not backward compatible
- Current proposal is to use extcomm to encode afi/safi scope in RTC path
  - Backward compatible, incrementally deployable, can be rolled back after transition period
- Question for the WG – is this an interesting enough problem for the WG?

AS Migration
draft-ietf-idr-as-migration

Wes George

Shane Amante

The problem

- BGP-speaking networks merge, acquire, split, reconfigure
  - this usually requires routers to change ASNs
  - Confederations not always a good solution
- Difficult for operators to coordinate ASN changes with eBGP peers
  - Each router moved to new ASN must have all eBGP peers reconfigure remote-as **simultaneously** or BGP sessions won't come up
  - doesn't scale to thousands of PE routers with hundreds of sessions each
- Mid-migration AS-Path lengthening creates undesirable traffic shifts

The Solution

- Vendors implemented BGP knobs that allow manipulation of ASN inside PE's BGP
  - Local-AS: Modify AS_PATH inbound to the Svc Provider's AS
  - Replace-AS: Modify AS_PATH outbound from the Svc Provider's AS
  - Internal BGP Alias: Seamlessly move iBGP sessions, e.g.: from PE's to RR's, from one ASN to a second ASN
  - Looks like normal spec-compliant eBGP session external to the router
- Requires no coordination/reconfiguration from eBGP peers
  - Remote-side (CE routers, esp. unmanaged) can still and do use legacy ASNs indefinitely
  - If it ain't broke, don't fix it

Why does IETF need to care?

- AS Migration tools documented in draft are:
  - Widely used by operators in Internet ASN migrations
  - Implemented to avoid BGP protocol errors (mismatching ASN's in OPEN and UPDATE PDUs)
- Need a stable reference to document these de facto standards and that they are in wide use
  - Changes that would break these capabilities (e.g SIDR BGPSec path validation) are a non-starter
  - draft-ietf-sidr-as-migration adds AS migration support in Path Validation

Draft progress

- Recently adopted by IDR, few changes from individual draft
- Several implementations have been deployed for over a decade.
- also presenting in GROW
- More reviewers?

Questions

- Document type: Info or BCP vs PS?
  - Currently: Informational
  - No interop needed (config and AS_PATH changes are only locally-significant)
  - Current draft points to 3 vendors' documentation, is 2119 text defining a single reference implementation necessary?

Performance-based BGP Routing Mechanism

draft-xu-idr-performance-routing-00

Xiaohu Xu (Huawei)

Hui Ni (Huawei)

Mohamed Boucadair (France Telecom)

Christian Jacquenet (France Telecom)

Ning So (Tata Communications)

Yongbing Fan (China Telecom)

IETF89, London

Motivation

- **Network latency is widely recognized as one of the major obstacles for migrating business applications to the cloud.**
  - **Cloud-based applications usually have very clearly defined and stringent network latency requirements.**
- **Service providers with global reach aim at delivering low-latency network connectivity services to their cloud service customers as a competitive advantage.**
- **Performance routing paradigm is meant to use network latency information as an input to the route selection process.**
- **It's expected that the performance routing paradigm could coexist with the vanilla routing paradigm.**
  - **Service providers could thus provide low-latency routing services while still offering the vanilla routing services depending on customers' requirements.**

Proposed Solution: Rationale

- **Enhance BGP with the ability to disseminate network latency information via a dedicated attribute and take that information as an input to the route selection process.**
- **The solution is designed to be backward compatible with existing BGP implementations and have no impact on the stability of the overall routing system.**
- **This document focuses exclusively on BGP matters.**
  - **All those BGP-irrelevant matters such as the mechanisms for measuring network latency are outside the scope of this document.**

Performance Routing Capability

- **"Performance" (low latency) routes SHOULD be exchanged by means of a specific SAFI (TBD) and also be carried as labeled routes.**
  - **Performance routes can then be looked as specific labeled routes associated with the network latency attribute.**
- **A MP-BGP speaker that advertises "performance" routes SHOULD use the Capabilities Optional Parameter [RFC5492] to inform its peers about the performance route computation capability.**
- **A MP-BGP speaker that implements the Performance Routing Capability MUST support the BGP Labeled Route Capability [RFC3107].**
  - **A BGP speaker that advertises the Performance Routing Capability to a peer using BGP Capabilities advertisement [RFC5492] does not have to advertise the BGP Labeled Route Capability to that peer.**

Performance Route Advertisement

- **Network latency metric is attached to the performance routes as NETWORK_LATENCY path attribute.**
- **Originating performance routes**
  - **A BGP speaker SHOULD be configurable to enable or disable the origination of performance routes.**
- **Distributing a performance route learnt from a BGP peer**
  - **If this BGP speaker has set itself as the NEXT_HOP of such route, the value of the NETWORK_LATENCY path attribute is increased by adding the network latency from itself to the previous NEXT_HOP of such route. Otherwise, the NETWORK_LATENCY path attribute of such route MUST NOT be modified.**
- **To keep performance routes stable enough, a BGP speaker SHOULD use a configurable threshold for network latency fluctuation to avoid sending any UPDATE which would otherwise be triggered by a minor network latency fluctuation below that threshold.**

Performance Route Selection

- **Performance route selection only requires the following modification to the tie-breaking procedures of the BGP route selection decision (phase 2, [RFC4271]):**
  - **Network latency metric comparison SHOULD be executed just ahead of the AS-Path Length comparison step.**
- **The Loc-RIB of performance routing paradigm is independent from that of the vanilla routing paradigm.**
  - **Accordingly, the performance routing table is independent from that of the vanilla routing table.**
  - **Whether performance routing or vanilla routing paradigms would be used for a given packet is a local policy issue which is outside the scope of the document.**

Deployment Considerations

- **It is strongly RECOMMENDED to deploy the performance-based BGP routing mechanism across multiple ASes which belong to a single administrative domain.**
- **Within each AS, it is RECOMMENTED to deliver a packet from a BGP speaker to the BGP NEXT_HOP via tunnels, typically TE LSP tunnels.**
  - **If a TE LSP is used between iBGP peers, it is RECOMMENDED to use the latency metric carried in Unidirectional Link Delay Sub-TLV [draft-ietf-ospf-te-metric-extensions] [draft-previdi-isis-te-metric-extensions] to calculate the cumulative link latency associated with the TE LSP and use that cumulative link latency to approximately represent the network latency. Thus, there is no need for frequent measurement of network latency between IBGP peers.**

Next Steps

- **Comments?**

IPv6 BGP Identifier Capability for BGP-4

draft-fan-idr-ipv6-bgp-id-00

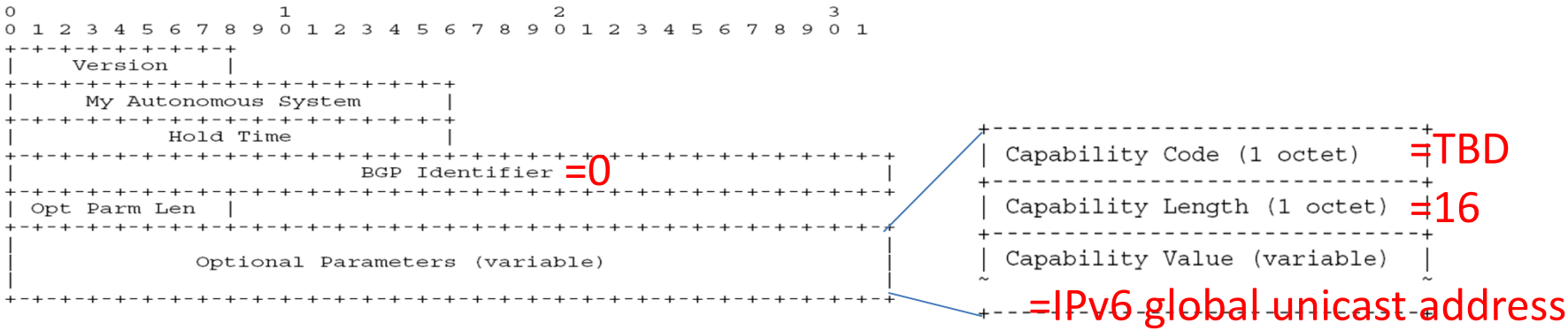
Peng Fan, Zhenqiang Li

Motivation

- The Identifier of a BGP speaker was specified as a valid IPv4 host address assigned to the BGP speaker in RFC4271; RFC6286 relaxed the definition to be a 4-octet, unsigned, non-zero and AS-wide unique integer.
- BGP Identifiers in a real network are often configured in the form of an IPv4 address to help network maintenance.
- The 4-octec integer Identifier in IPv6-only network requires additional configuration and planning consideration to guarantee uniqueness within the AS.
- This document extends BGP to allow a BGP Identifier to be a valid IPv6 global unicast address assigned to the BGP speaker.

Protocol Extension

- A new BGP capability code, “IPv6 BGP Identifier Capability”, is defined to indicate the support for IPv6 address as a BGP Identifier.
- OPEN message: the BGP Identifier field is set to zero, indicating the actual BGP Identifier is in the Capability Optional Parameter.
- IPv6 BGP Identifier Capability: The Capability Length field of the is set to 16, and the Capability value field is set to an IPv6 global unicast address.



- AGGREGATOR attribute: set accordingly; the BGP Identifier carried in the attribute is encoded as a 16-octet entity.

Operation

- Processing received OPEN messages:
  - If the BGP Identifier field is not zero: process in the way of the message that does not contain IPv6 BGP Identifier, and any IPv6 BGP Identifier Capability is ignored.
  - If the BGP Identifier field is zero, then check if any IPv6 BGP Identifier Capability is carried. If there is no IPv6 BGP ID Capability, or the capability value is not a valid IPv6 global unicast address, then a Notification message is generated, with Error Code set to 2 (OPEN Message Error) and Error subcode set to 3 (Bad BGP Identifier).
- Connection collision detection:
  - The BGP Identifiers of the peers involved in the collision are compared and only the connection initiated by the BGP speaker with the higher-valued BGP Identifier is retained.
- Route selection decision:
  - If a route is advertised by an IPv4 BGP speaker and an IPv6 BGP speaker respectively, then the route advertised by the IPv6 BGP speaker is selected.
  - If a route is advertised by two IPv6 BGP speakers respectively, then their IPv6 BGP IDs are compared, and the route advertised by the BGP speaker with the lower-valued BGP Identifier is selected.

Transition

- A BGP speaker supporting the IPv6 BGP Identifier must set a 128-bit Identifier.
- If the speaker is not aware of the capability of its peers, then a 32-bit Identifier is assigned for backup purpose.
- The speaker tries the 128-bit identifier first; if the peer does not support the new 128-bit ID capability, then a “bad bgp identifier” error message is generated (Identifier field of OPEN message received is zero).
- The speaker initiates a second connection using 32-bit identifier in the old way, and the connection falls back using 32-bit identifier.

Consideration

- Pretty much discussion on the list
- Advantages of extending the length of identifier
  - Help identify the location, e.g. for diagnosis and troubleshooting
  - Can be autoconfigured
- Other suggestion on the list:
  - Use other separated mapping system, e.g. DNS, text, v4-v6 addr mapping. (Extra record keeping adds more work for OAM)

- Do we update ID to convey more information or just keep the 32 random bits?

ADD-PATH for Route Servers
draft-francois-idr-rs-addpaths-00

Pierre Francois, IMDEA Networks
Camilo Cardona, IMDEA Networks
Adam Simpson, Alcatel-Lucent
Jeffrey Haas, Juniper Networks

Introduction

- Route servers facilitate operation in IXP
  - draft-ietf-idr-ix-bgp-route-server-04
- The use of RS can lead to path hiding
- ADD-PATH can provide a solution to path hiding
- The goal of this draft is to define the behavior of ADD-PATH in this environment
  - draft add path guidelines, for eBGP
  - Current focus: Route Servers

Operation of ADD-PATH for RS

- Propagation rules
  - Clients announce a single path per destination
  - Route Server announces all in-policy paths
- Error cases
- Preservation of resources for clients?
  - Another ADD-PATH mode (ADD-ALL-PATH <limit>)?
  - Proposition of the ADD-PATH limit capability?
    - draft-francois-idr-addpath-limit-00

Questions
ADD-PATH for Route Servers
draft-francois-idr-rs-addpaths-00

Pierre Francois, IMDEA Networks
Camilo Cardona, IMDEA Networks
Adam Simpson, Alcatel-Lucent
Jeffrey Haas, Juniper Networks

ADD-PATH limit capability
draft-francois-idr-addpath-limit-00

Pierre Francois, IMDEA Networks
Camilo Cardona, IMDEA Networks
Adam Simpson, Alcatel-Lucent
Jeffrey Haas, Juniper Networks

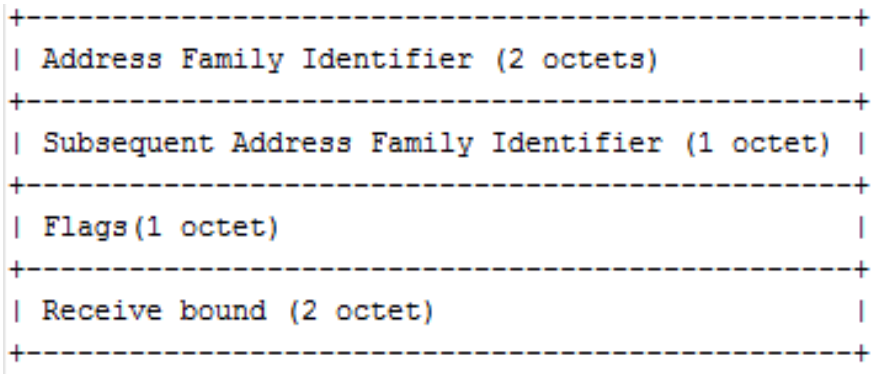
Introduction

- Context
  - Add-path / RS
  - eBGP

  (draft-francois-idr-rs-addpaths-00)
- Resource preservation for RS clients
- New capability
  - Maximum number of paths per NLRI that the receiver wants to receive
  - Currently: send / receive

Operation

- BGP capability:



  - Sender signals its ability to limit the number of paths that it can send to the receiver (1 bit in flags)
  - Receiver signals a limit on paths it wants to receive from the sender
- Error conditions

Comments received

- A capability makes it mandatory to restart the session if the path-limit changes
  - ORF as an alternative to a new capability?
  - Dynamic capability?
- How would the IXP select the paths?
  - currently left for the implementation
- Signal a limit on the total number of prefixes that a receiver wants to get?

Questions/Comments
ADD-PATH limit capability
draft-francois-idr-addpath-limit-00

Pierre Francois, IMDEA Networks
Camilo Cardona, IMDEA Networks
Adam Simpson, Alcatel-Lucent
Jeffrey Haas, Juniper Networks