IETF 89 IDR Meeting on Thursday, March 6, 2014
1300-1500 Afternoon Session I

Location: Blenheim

CHAIR(s): Susan Hares <shares@ndzh.com>
John Scudder <jgs@juniper.net>

o Administrivia
Chairs 10 minutes

- Note Well
- Scribe:
- Jabber Scribe: Wes George;
- Blue Sheets
- Document Status

Presentation on Document status:

- New documents are:
  - draft-ietf-idr-te-lsp-distribution
  - draft-ietf-idr-te-pm-bgp
  - draft-ietf-idr-mdcs-00
  - dratf-ietf-idr-mdrs-00
  - draft-ietf-idr-as-migration
- Pass WG LC
  - draft-ietf-idr-aigp
  - draft-ietf-idr-bgp-enhanced-route-refresh
  - draft-ietf-idr-last-as-reservation
  - draft-ietf-idr-ls-distribution (for early code point adoption)
- Heading for WG LC
  - draft-ietf-idr-ls-distribution (will begin on 3/12/14)
  - draft-as-migration
- IDR will be reviewing old IDR drafts
  - Asking the progress, if the draft is not necessary we will flush it from the queue.
- IDR co-chairs are looking for operator input on whether the current IDR RFCs should be updated due to reported errata.
  - Some Data Center operators are getting new implementations of BGP

Discussion:

- Jeff Haas: Has draft-haas-idr-flowspec-redirect-rt-bis-01 had a call for adoption?
- Sue Hares: I missed this. I will do the call for adoption from 3/6 to 3/20/14.

[added after Johnôs Error handling draft]

- John Scudder: In IDR we do WG call fairly late in the life-cycle of a document because we have a requirement for implementations. We in the IETF (for various reasons) have trained ourselves to comment at WG LC, and so a majority of review happens at last call. If youôve got an implementation requirement, review at last call is too late. We saw an example of this effect in AIGP. There were good comments at last call that would have resulted in a protocol change had they come earlier. However, because the suggestions came after implementations had been widely

fielded the answer was: ñNice change, but too lateò. Due to this fact, the chairs ask draft authors to indicate when they start implementation to poke the list. We should have a pre-WGLC review.

- Jeff: One of the tools we have is a WG Wiki that the IETF makes available. We should get our drafts listed there in table format and place a category for early review.
- John: We have talked about adding a Wiki. We will also use email for people who do not make a regular practice of looking at the wiki.
- Sue: (via chat): We plan to us the Wiki after this session.
- Alia: (via chat): We could have a status of ñwaiting for implementationò on the drafts on the main page.

o Revised Error Handling for BGP UPDATE Messages -06
http://www.ietf.org/id/draft-ietf-idr-error-handling-06.txt
Authors: Enke Chen, Pradosh Mohapatra, Keyur Patel
John Scudder 10 minutes

Presentation

- [John]: I plan for this be the last time to stand before you to discuss the error handling draft. There is a recording so you can hold me to this.
- [John]: The intent is to update the group that there were some non-trivial changes in version 6. We hope to move expeditiously to WG LC. To provide short summary of the error handling, Standard BGP resets the session anytime at any error. This handles errors, but it is problematic for operations. This draft (which is fairly mature) suggest to avoid resetting the session by treating every error possible as a ñwithdrawalò of prefixes covered by the error. If you cannot dig the prefixes out then you are trouble, the fall back to dropping the session reset.

- [John]: This has been implemented and deployed, and people think it is a good idea. However, at the last IDR meeting (IETF 88), Eric Rosen was presented AIGP, and he brought up the case where he observed that one piece of the error handling states that if the attribute flags are inconsistent with the attribute code, then just fix the flags.
  This was consistent with original content of error handling draft to avoid the session resets that RFC 4271 does. However, Eric Rosen pointed that the behavior of fixing of the attribute flags was unsafe. He convinced us that this change was ñtoo minimally disruptiveò and stepped over the line to ñcrazyò by allowing broken attributes to propagate. Therefore the authors back out this feature. This is will require a re-spin of implementations, and it will be backwardly compatible.

- [John]: As part of the review, we noticed that RFC 4271 mandates the NLRI must be syntax-checked, but it does not define what this means. We thought since error handling depends on defining what it means, we should be more proscriptive. So we added more text:
  - S.3.1: Error on length mismatch (overrun or underrun), Treat-as-withdraw
  - S.3.2: Lengths must be sane, Session reset if error
- John]: All of this are things that I hope implementations are already doing.
- [John]: Tony P. pointed out the following additions:
  - S.3.2: Plan to add two more error conditions for MP NLRI (bad attribute flags, and bad attribute length), Session reset if error
- [John]: This will result in version 7 of the draft. Please if you are interested in getting BGP right, please review this specification.
  - We should have a frank and open discussion.
  - I hope we have right, but we should ear.

Discussion:

- Keyur Patel: Should new drafts use this error handling as the base for their work (rather than RFC4271)?
- John Scudder: Yes they should. Until we get WG LC, it could change.
- Keyur Patel: This will avoid
- Sue Hares: Iôd like to put that in a stronger. You should put a section in your draft on how you align with the error handling draft.
  [discussion after error handling]

o BGP Link-State Information Distribution Implementation Report
http://www.ietf.org/internet-drafts/draft-gredler-idr-ls-distribution-impl-00.txt
co-authors: Hannes Gredler, Balaji Rajagopalan, Saikat Ray, Manish Bhardwaj
Hannes Gredler 10 minutes [13:17: 13:27]

Presentation:

- Hannes: Iôm talking about the implementation report viewed as many as a necessary evil. Rather than a necessary evil, it is a prudent step in deployment BGP code.
- [Slide 2 of slides-89-idr-9.pdf]
  - Are the LS TLVs generated the same between two implement?
  - The BGP LS is a superset of all the TE generated objects specified in the last 30 years. It is important that a route reflector can forward all of the TLVs.
- [Slide 3]: Build simple iBGP-eBGP-iBGP chain
- [Slide 4]: Build iBGP topology with Route Reflector
- [Slide 5]: Good news: Only one protocol point was unclear
  - Juniper and cisco had some ambiguity when generating a link (Unnumbered/Numbered) link. We need to update the specification to clarify this issue.
  - Majority of the issues were implementation issues. This shows it is important to have an interoperability testing of implementations.
  - AFI was coded as UINT8 - which caused early allocation BGP-LS AFI of 16338 to break things (including Route Refresh).
  - The Endian-ness for TEs for Bandwidth was broken,
  - Inconsistent keys were generated for OSPF ñPsuedonodesò/LAN abstractions (router-id + interface id), this has to be clarified in specification.
- [Slide 6]: draft-gredler-idr-ls-distribution-impl-00
  - Two implementations listed in the document.
  - Two other implementations were lists:
    - Telefonica route reflector, and
    - open-source Java BGP LS receiver.
  - We plan to update the document and publish it as a WG document.
- [Slide 8]: We plan to update the base specification, and ask for WG group adoption for the implementation report.
  - The implementation report needs to be updated to clarify all the places where things are unclear.

Discussion:

- J (?): Thank you. It is interesting to have feedback on the interoperability tests for this draft. Is there any message in slide with the white/black screens for Juniper/Cisco?
- Hannes: No message was intended, it was just two different screens on my linux desktop.
- J (?): On the implementation report, I felt section 7.3 was quite offending for implementations. You should consider dropping section 7.3.

- Hannes: This was boiler plate and a left-over, so we plan to drop it.
- J (?): Thank you.
- John: The adoption request is for the implementation report.
- Hannes: I will put the request on the list.

o Traffic Engineering Database dissemination for Hierarchical PCE scenarios
http://tools.ietf.org/html/draft-lopez-pce-hpce-ted-00
Co-authors: Victor Lopez, Daniel King, Stefano Previdi:
Oscar Gonzalez de Dios 10 minutes [13:27-13:37]

Presentation: [slides-89-idr-01.pdf]
- [Oscar]: This is work we are doing at the PCE WG that uses one of the BGP features defined by IDR. We were requested by the PCE chairs to present this work in IDR to get feedback from IDR.
- [slide 2]: Hierarchical PCE (H-PCE) selects sequence of domains in the inter-domain cases where there is not pre-defined sequence. The selection of domains is based on the topological information gathered by the parent PCE.
- [Slide 3]: Solution is to solve multi-domain path computation by cooperation via different PCE. The H-PCE architecture (RFC6805) has already been defined, and a solution draft published (draft-ietf-pce-hierarchy-extensions-00) using PCEP protocol and computation procedures.
    - However, the topology dissemination is an open issue in the drafts (draft-ietf-pce-questions).
    - Goal of this draft: analyze how topology dissemination may be usd to provide TE formation exchange between Parent and Child PCEs.
    - Non-goal: Solving this problem for all domains. This solution is for just a few domains.
- Slide 4: What and how to provide Topology dissemination
    - what: Inter-domain links, Edge-to-Edge Virtual TE links created form LSPs
    - how: Static, join IGP instance, PCE notification, Separate IGP, Northbound push to BGP-LS
        - The IGP would break the inter-domain boundaries,
        - The PCE notification puts this data along with other PCE data, and we think this is not the right approach.
        - A separate IGP instance could be create to pass information.
        - We think the BGP-LS is useful to pass the topology information.
- Slide 5: provide diagram of architecture
    - You can tune the information sent form the TED and PCE to the BGP-LS speaker. I want to send only specific area or inter-area routes. I may want to open the pipe a bit more to send some more abstracted information. You can send full description of the network if needed to calculate the interdomain

- Slide 6: Open issues:
    - BGP-LS: Shall we use for H-PCE distribution of topology data
    - OSPF-TE/IS-IS-TE mappings: These seem to be in good shape for packet world, but there is additional information needed for the optical world (links). H-PCE is used in multi-domain optical networks so there is more work. In some research projects we have been implementation BGP-LS for this use. This implementation was used to test Cisco, and is compliant with the draft.
    - Please ask

Discussion:
- Saikat Ray (Cisco): Just a clarification question, Is the Domain you are looking at a single AS or are you trying to get the links between the AS Domains, or between multiple IGP boundaries.

- Oscar: We need the links between the AS. You can also get abstracted information from the domains. You do not need to send the whole IGP of the domain, but you may wish to have edge-to-edge information.
- Saikat: BGP-LS can obtain does have other clients (other than BGP and ISIS) that can add that information that is already there. This open issue of the mapping between ISIS and TE, can you explain what is open here?
- Oscar: For optical, there is some extensions to OSPF-TE/ISIS-TE for the special traffic engineering information for optical paths that are not reflected there yet. There are some GMPLS attributes in BGP-LS, but no all attributes are added. The plan is that you get information from IGP and feed your TED database. From there, you pop the information into the BGP-LS and send it. Therefore, you need to map all the attributes.
- Saikat: Obviously, you want to add the information in ISIS or OSPF, we should add it to BGP-LS as well. BGP-LS also has an opaque TLV where you can send anything, and define within the opaque TLV you can add your TLV. We have defined during this start-up period to allow you to send information within the opaque TLV to get your project working.

o BGP Link-State extensions for Segment Routing
http://tools.ietf.org/html/draft-gredler-idr-bgp-ls-segment-routing-extension-01
co-authors: Hannes Gredler, Saikat Ray, Stefano Previdi, Clarence Filsfils,
Mach Chen, Jeff Tansura

Saikat Ray 15 minutes [13:37 -13:50]

Presentation: [sdlies-89-idr-6.pdf] ï BGP-LS extensions for segment routing
- [slide 2]
  - Segment: Flexible and scalable way of doing source routing
  - Segment: instructions impact path (ID) and services
  - IGPs advertise: <Segments, SID> with Igress Node adding SID stack to determine packet path
    - Per-Flow state only at ingress node;
    - SIDs map to MPLS label for MPLS data
- [Slide 3]: Need for BGP LS
  - Segments create end-to-end path across multiple areas
  - Could connect router to multiple IGP Areas,
  - BGP LS provides visibility into multiple IGP areas.
- [Slide 4]: BGP-LS part
  - Three types of objects: Nodes, links and prefix. Prefix in NLRI, and the rest of properties in BGP attributes.
  - add Segment information to BGP-LS
- [Slide 5]: Description of Segment Routing TLVs
  - 5 TLVS: Prefix-SID, Adjacency-SID (p2p/LAN), SID/Label binding, SR Capabilities, TLV for SR algorithm
- [Slide 6]: Mapping of BGL-LS SR TLV to IS-IS Node attribute (TLV/sub-TLV)
  - SID/label binding, SR capabilities, SR algorithm
- [Slide 7]: Mapping of BGL-LS SR TLV to IS-IS Link Attribute
  - Adjacency segment, LAN Adjacency Segement.
- [Slide 8]: Mapping of BGP-LS SR TLVs Prefix to IS-IS TLVs
  - New developments in the next version.
- [Slide 9]: We will ask this to be a WG draft.

Discussion:

Shane: I read the draft, I think I understand it. It is good to extend segment routing to include the multiple AS topologies. The question I have are around label stack depths. If I understand the draft correctly, the draft suggest that as you go inter-AS you will have a label stack per AS. When you return to your ingress PE, you will have a much greater Label stack that what exists today. Do I under the draft correctly?

Saikat: In most cases you do not have that big of a label stack. You could construct a case for a large label stack. However, in most cases you will follow the prefix path, and have one AS pathway. The question of label stack size is orthogonal to the decision to carry the information in BGP.

John: It is not completely orthogonal. BGP-LS may be used to build the bigger stacks. I think it relevant.

Shane: I think what you have written is understandable, but the operational deployment perspective. Inside one domain we can understand that there is limited label stacks. When you go beyond the AS boundary to additional ASes, this will put more pressure on the label stack.

Saikat: This is a valid point.

Moshi (?): Do we care if we only care about the top label? The Kompella discussion of entropy label depth aside.

Hannes: Trying to answer Shane question about labels, in segment routing we have flexibility. There is no strict model. At the ingress, we have a huge stack of labels, and then the label stack is monotonically shrinking. We can do a distributed stacking where nodes (further down the stack), may advertise a label representing several steps. In this case, you are pushing a set of labels (3-4) and you are reducing the labels as you go across a portion of the network. Once you get to the edge of this portion of the network (at a transit router), you push a new set for the next portion of the network. A second part of that question is hardware. I am talking to my hardware developers, and they ask how many labels do you need? My answer is the MTU/4. [chuckles in the room]

Jeff Tansura: You will not do label stack, you will just have one. We have already described how to deal with SR (source route) passing through LDP domains. You can deal with BGP the exact same way. You simply map one-to-one. There is no need to stack label, just map one domain to another.

John Scudder (co-chair): No more comments after this one.

Ilya Varlashkin: I am fine with the information in this document. Where is the interoperability with LDP in segment routing as such. Where does it fall? Does it need to be here? Should it be here? You might have some influence on which interoperability architecture you will use. You might have some influence on the BGP part. Does it come here or does it go to Spring?

Saikat: It sounds like a default string. If you want to get the label from somewhere else.

Illya Varlashkin: Are you planning to put the labels in BGP, and you hope that the SPRING group will design a good way to use it.

John Scudder: This sounds like a good Spring topic.

o Constrained Route Distribution with Multiple Address Families
http://www.ietf.org/id/draft-ray-idr-route-constrain-scope-00
co-authors: Arjun Sreekantiah, Keyur Patel
Saikat Ray: 10 minutes [13:50-14:02]

Presentation:
[slide 2]: Draft has 2 problems. Problem 2 and suggest solution are
presented in presentation. We are seeking feedback to Problem 1. This is not a new problem, but this is report from implementation experience.

[Slide 3:] Background: RR Sends all routes to the PE for a family, but PE only keeps routes imported into at least one VRF. Optimization is PE tells RR which RTs the PE wants to send only that route. There are multiple was to do this work (For Example ORF). In summary, the PE sends the RT of interest (For example, the AS + prefix) to the RR, and the RR knows the PE only wants these routes.

[Slide 4]: The RTC NLRI has the RT and the RT only. The community can standardize import/export of the VPN routes using the Route Targets. This means it is useful for EVPN, IPv4/IPv6 routes, and pretty much everything else. A simple example is in the slide. The RT is imported for rd 1 (RT 1:1 with prefix 1/8), and rd2 (RT 1:1 1::/64). rd2 is dropped due to the lack of any PE importing the IPv6 family.

[slide 5}: This creates two problems. The PE has only one VRF (RT 1:1 1/8) which it imports. RR has v4/v6 routes with RT 1:1, and RTC 1:1 form PE.

[Slide 6]: User adds VRF with IPv6 RT 1:1 to PE. PE Sends RTC 1:1 to RR. and it is dropped due to identical path check. A Work around is: PE sends route-refresh to 2/128 to RR, and RR will send all VPNv6 route to PE whose RTs match RTC 1:1. This is a larger set of transmission that just the routes that match RD2 (RT 1:1 1::/64).

[Slide 7]: Rule change BGP speaker that receives an identical RTC path from neighbor must treat as equivalent to Route-refresh for given RT for all (VPN) address-families.

[Slide 8]: We need to assure that when the BGP speaker sends the RTC path from the neighbor, you need to act upon it. Rule changes

- When an identical RTC path from a neighbor occurs, it should be treated as a equivalent to RR request for the RT only for all VPN address families. The PE will keep whatever is needed.
- If new VPN AFI is negotiated between two BGP speakers without a session reset (using dynamic capability or multi-session feature). Normally the session will flap. However, if you use dynamic capability or the multi-session draft or a different port so the original session does not flap. In which case, even though the VPN is coming later on the scope must include the new protocol.

- Why standardize since nothing on the wire? The RTC path acceptance is a change in process. The PE also needs to know whether the RR supports it this feature ï because the RR will either be ignored or be treated as a Route Refresh. Two ways to provide this is either a: CLI knob in the RR or a capability. We could utilize the reserve bits in the MP capability.

- We want feedback from the IDR WG on this problem.

[Slide 9/10]: Problem 2:

- Two RRs both in two different regions with the RTC 1:1,
- PEs in region 1 require both IPv4 and IPv6 address with RT 1:1
- PEs in region 2 only require IPv4, and the RR still retains the routes with IPv6.
- When does it occur:
    - Case a: Different address-families in different VPNs use the same RT (not the usual operational practice),
    - Case b: Not all sites have the same set of Address-families
- Previous proposal: Add the SAFI in NLRI
- Current proposal: Use extcomm in afi/safi in RTC

Questions: Is this a problem that is interested for the WG time?

Discussion:
Luyuan Fang (Microsoft): Small comment, you need to specify the RFC4684 in the drafts. Second comment, is a protocol issue. If you want to propose something that changes the protocol.

Saikat: We are trying to indicate that this change causes different behaviors in the protocol.

Luyuan: It would also be good to have a use case for the user.

o Autonomous System (AS) Migration Features and Their Effects on the BGP AS_PATH Attribute
http://www.ietf.org/id/draft-ietf-idr-as-migration-00.txt
co-author: Shane Amante
Wes George 10 minutes [14:02 ï 14:06] slides: [slides-89-idr-5.pdf]

Presentation:
[slide 1]: We have present this material in grow. We are going to breeze through material as it is in the draft.
[Slide 2/3]: Knobs allow us to do as migration by manipulating the path (merge, acquire, split, reconfigure) transparently to the EBGP peers. This has not been standardized in the protocol. It is a local issue, and does not require interoperability. We have run into a situation where numerous things in BGP specifications may break this so we want to document this feature.

[Slide 4]: Why care? The operator community makes heavy use of this feature.
Operators need stable reference to document these defacto standard in wide used. SIDR changes (BGPSec path validation) would break this feature (draft-ietf-sidr-as-migration) because it was meant to stop manipulation of SIDR path. You need a stable reference so you can fix a standardized behavior.

[Slide 5]: We adopted it by IDR. Multiple implementations so this is a rapid transition to WG LC.

[Slide 6]: Feedback for target as Info, BCP, or PS. We do not have any RFC2119 language in the draft. It has pointer to three vendor implementation due to documentation. Should we look at this as RFC2119 language?
It is not a cookbook for AS migration. It is a good

Discussion:

Ilya Varlashkin:: Currently it looks took much a cook book. I think you need to make these changes to emphasize what the features you are defining for the knobs, and not the use of the features. I will send the proposed text changes after the IETF.

Wes: Additional comments? No, thank you.

Sue: Please review this document for it will come

o Performance-based BGP Routing Mechanism
http://tools.ietf.org/html/draft-xu-idr-performance-routing-00
Xiaohu Xu 15 minutes [14:06 ï 14:22] [slides-89-idr-4.pdf]

Presentation:
[slide 2]: Network Latency is recognized is major obstacle to migrating business to Cloud. Service providers with global reach and low-latency consider this a competitive advantage. Performance routing is meant to use network latency as input to route selection process. BGP can use performance in parallel with vanilla routing paradigm.
[slide 3]: The solution is to use metrics to provide network latency. This solution is backward compatible.

[slide 4]: Performance routes should be exchanged by using specific SAFI, and carried as labeled routes (with associated latency). MP-BGP speakers announce a capability to support this type of BGP Labeled Route Capability.
[slide 5]: Latency is carried as a path attribute. This path attribute can be attached via configuration. If a BGP peer sets as Nexthop, the BGP latency should be increased by adding network latency to previous NextHop. The BGP speaker should use threshold to prevent fluctuation of routes based on latency changes.

[Slide 6]: Small Change to selection process. Latency should be compared ahead of path length comparison. This selection difference only occurs in the performance RIB, and is independent of the vanilla routing table.

[Slide 7]: Strongly recommended that this technology should only be in a ASes in single Administrative Domain. Within AS, recommended to use tunnel to go from BGP speaker to next-hop. If TE LSP used, use Unidirectional Link Delay Sub-TLV in ISIS/OSPF TE draft to calculate latency.

Discussion:
Saikat (cisco): The AIGP metric was intended to do the things you are doing. I understand that the AIGP metric is configured, and you want a measure link value. However, the AIGP metric does not have to be configured. Fundamentally you do not need another feature to pass this information. Another comment I want to make is a little subtle. If you do not have tunnels between the BGP speaker and the next-hop you might create a loop with this feature because latency will not be the same for all the BGP peer-next-hops.

John: Do you want to respond to this? The two points were why not AIGP, and forwarding loops.

Xiaohu: For the difference between this draft and the AIGP draft, I will comment on the mailing list.

Keyur Patel Regardless of using 1 SAFI or 2 SAFIs, or using a local decision on a BGP speaker ï if you do not use a tunneling mechanism (that is a hop-by-hop forwarding), you risk running into loops.

Wes George (TWC): Please donôt. (claps) Please stop helping.

- If this does proceed, and you put this selection before AS path length you need to determine how you will secure the path attribute going across EBGP boundaries. You need to get a trust mechanism within BGP. Otherwise, you risk problems. You did indicate using it with ASes administrative boundaries within a single provider. Unless the document makes some very explicit recommendations, that these features must be filtered at the AS boundaries between the management domains you run the risk of untrusted information impacted the decision tree in a way that can be really, really detrimental to your network. Or you run into the potential for a DOS Attack vector where I can set the latency at 1, and then this the best path for a certain set of things. There is a huge security set of considerations. Ultimately, this seems like a variant of F* put it in DNS, F* put in it BGP.

Xiaohu: Do you mean we should limit the implementation of this technology to multiple ASes belonging to a single provider?

John Scudder: I believe that Wes stated that since you are putting it in an inter-domain protocol you need to consider the security implications. That just because you say ñjust use in in a single domainò, this is not sufficient to guarantee it is not used in a single domain. You need to think about what security features you need to have if it is not.

Wes George (TWC) You can consider this as the transitive versus intransitive communities. It is useful in some cases, but when jump across AS boundaries there are a whole set of considerations about where this action is appropriate and where it is not. The jump across AS boundaries with certain features makes the feature ripe for abuse.

Xiaohu: Is your concern applicable to the AIGP draft as well as this draft?

Wes: I made this comment regarding the AIGP [scribe: soft). Now we would have two of these features.

Jared Maruch (NTT): When this draft came up on the list, I raised a number of concerns about the introduction of additional state and propagating this additional information. I still havenôt seen these addressed. Iôm surprised to seeing this discussed as a WG activity until these issues are addressed. This is not meant to attack the WG process or anything, but this is a draft should not move forward due to their being a number of other solutions that are already well-defined that are deployed. There is a lot more that goes into performance solutions beyond just latency. This solution is in many ways incomplete thought process that needs to be fully developed before any more consideration is given.

John Scudder: Those standing in the mike lines, please keep your comments brief. For those coming to the mike, the line is closed. No further people.

Jeff Haas (Juniper): Are you familiar with the Cost community? Why did you choose not use this instead.

Xiaohu: Since the cost community does not want to use a specific SAFI to identify the routes, that solution was not sufficient. If the Cost community can be attached to a specific SAFI then it would be sufficient. It is fine to the cost community or the AIGP attribute. This is just a path attribute to carry the latency attribute.

Sriganesh Kini (Ericsson): Before the draft enters into suggesting a solution, it should make the problem statement clear. What is the problem? Is the problem to find a path for minimum latency? Or is the problem to find a path that satisfies the latency constraint?

Xiaohu: Please send your question to the mailing list.

Sriganesh: Should I give a specific example? Are you treating latency like a metric that suggests I need the lowest latency path for this particular use case or client? Is the client saying give any path that is less than a particular latency (Eg. 100 ms latency)

Xiaohu: BGP is used to advertise the best path. The result is the best path.

Sriganesh: This is a BGP property. What is the problem statement for the draft?

John Scudder: I think the point you should take for your draft is that you have not clearly articulated a problem statement, and you should do so if you want to progress it forward.

Randy Bush (IIJ): I do not think this draft is worth further comments/attacks. I want to point out to Wes George and others, any statement that includes ñmust be filteredò is a trip to Disney land.

[chuckles].

John: Thank you Xiaohu:

o IPv6 BGP Identifier Capability for BGP-4
http://www.ietf.org/id/draft-fan-idr-ipv6-bgp-id-00
Peng Fan 10 minutes [14:22-14:30] slides-89-idr-10.pdf

Motivation:

- Identifier of BGP speaker was specified as a valid IPv4 host address in RFC4271. RFC6286 relaxed to 4-octet, unsigned, no-zero, AS-wide integer. 4 octet-integer identifier in IPv6-only network requires additional configuration to guarantee uniqueness within AS. This draft extends BGP identifier to be valid IPv6 global unicast address assigned to BGP speaker.

Solution:

- BGP capability code with IPv6 BGP identifier capability supports IPv6 address. Open message carries IPv6 identifier in Capability Optional Parameter. If this field exists, then process as IPv6. If not, use the original method. Connection collision of these two attributes will use higher-value BGP identifier in this Capability optional identifier.

- Transition: Each peer must support IPv6 BGP identifier must 128-bit identifier, and assign a 32-bit for back-up. If requires 128-bit identifier, send OPEN-message with bad BGP identifier.

- Question: Do we need think about this solution or stay with the 32 bit.

Discussion:

- Wes George: Please don't. More specifically, as the sunset-v4 chair-hati, there a bunch of protocols that use a 32 identifier by convention and not by protocol standard. IPv4 has been used for this identifier to define that 32 bit identifier. Do not solve the problem by making a 128 bit identifier. This feature is in multiple situations. The proper place for this work is IPv6-ops or sunset-v4 WG. This draft should stop where it is, and work on the sunset v4.

- Randy Bush (IIJ): What was unclear about the discussion on the list (that indicated you should not go forward with this solution)?

- Peng: I think the discussion on the list. When developing the RFC6286, how did you come to the conclusion of a 32 bit identifier? Was this concluded by operators rather than the protocol implementer?

- Randy: Just do me a favor John and Sue, not next time too.

- John: Your point is well-taken Randy.

- Jacob Hietz (Ericcson): In the draft you said you would set the identifier appropriately. If you put this through AS that do not support this feature, you would have to do something akin to the 4 byte AS. You would have to make a new aggregator. Another problem is the cluster ID uses a Router-ID, so this would have to be changed as well. You cannot do this by transition, but for the whole AS at one time. These are the two items off the top of my head.

- Sue: We are going to cut the line at this point.

- Keyur Patel (cisco): To expand on Jacobôs comment, It will be more than exchanging capabilities. Cluster-IDs and originator IDs get propagated to other guys, not just the guy who announced it. If you do not have capability exchanged with them, you have a larger problem.

o ADD-PATH limit capability
http://tools.ietf.org/html/draft-francois-idr-addpath-limit-00

co-authors: Pierre francois (IMDEA Networks), Adam Simpson (Acatel-Lucent, Jeff Haas (Juniper networks. [slides-89-idr-2.pdf]
Camilo Cardona (IMDEA Networks) 10 minutes [14:30-14:35]

presentation:
[Slide 2]:
- Camilo: Introduction of draft for using Add-path for Route Servers in E-BGP environment. Route server can create this problem call path hiding. The route-servers will send one of these path to a client, and the others will be hidden. This creates a problem.
- Add-path is the solution for this problem, but the add-path is an IBGP feature. This expands add-path to the route serve context. In the route server case, it does not make sense for the clients to send multiple paths to the route-server. We expand this feature to have multiple paths sent back from the route-server to the client. This tries to be non-disruptive when determining how to deal with each them. This feature looks at resource-preservation for the clients. The ISPs can have 600-700 route server clients in a heterogeneous environment where there are many types of routers. A route server client may not want all paths for a particular client, but 2-3 paths per destination. We can have an add-path mode where have a maximum number of paths per NLRI that the client (receiver) wants to accept.
- In this environment, there are two administrator (router-server and client administration) so the client may want to signal to the route-server the limits. This signaling will be covered in the next slot.

Discussion:
-John Scudder: What is the difference between ADD-Path-IN and ADD_path Limit
Camilo: In theory it should not be very different, but I think ADD-Path is where you select the best path, and then run the process again (as described in the next presentation). In the route server environment the

description of best is not really described. The best path could be any of the routes since it is the decision of the client which path is the best. If the route server doesnôt know what best path is?

John Scudder: Do I randomly pick ñNò routes?

Camilo: This is a solution, but not one I would implement long-term. We need more definition on this point. Initially, it is the solution that could be implemented now.

Thomas Mangin(Exa Networks): IT works today. If I would not have read the draft, I would have done it this way,

o ADD-PATH for Route Servers
http://tools.ietf.org/html/draft-francois-idr-rs-addpaths-00
Camilo Cardona 10 minutes [14:35-14:42)
co-authors: Pierre Francois (IMDEA Networks), Adam Simpson (Acatel-Lucent, Jeff Haas (Juniper networks. [slides-89-idr-3.pdf]

Presentation:

- We have route servers at IXP (exchange points) we want to be able to implement a maximum limit of routes.
  [draft-ietf-idr-ix-bgp-route-server-04]

- We have a capability that will indicate from the client to the route-server the number of paths that client wants to receive. In the document, we also mention some error conditions. We have received lots of comments for the draft.
- Robert R. indicated that if you have an ADD-Path limit capability, if you do change to the add-path then you need to reset the session. Alternative are ORFS and dynamic capabilities. I do not know the state of the dynamic capability.
- Other comments: You will fall into path hiding based on the paths chosen to be sent. John Scudder just mentioned this open issue.
- Comment: We had a comment to extend the message of the capability message extended to provide the max-prefix functionality. Why not use this message to see the max-prefix limit to the route-server to the sender. This should trigger a warning.

Discussion:

Jeff Haas: Robertôs comment on ORF was well-received. In the absence of Dynamic capabilities (which is my least favorite BGP feature), ORFs are not a bad mechanism for this feature. Iôd like to take a quick sense of the room. Whoôs read the draft? Of these people do you think you should keep capabilities versus ORF? The result is no strong opinion either way.

John Scudder: This is a question that needs to be asked on the list or in private email to the authors.

Keyur Patel (cisco): As we have exchanged private email, I would suggest that you do not put any number. Either announce all or announce just one path. Should you decide to put a number, it would be good to make it generic and apply it to I-BGP cases as well. This general application will make it apply to all of ADD path.

Luyuan Fang (Microsoft): I agree with what Keyur is saying. Second the add-path is a useful case for us. We would like to see it converge to the mechanism so it is more generic.

Saikat (Cisco): I understand the use case where you want to exchange the number to set the maximum pathway. One subtle point is that the route-server cannot change the next hop. The route-server may not be the best place for the client to affect this change. If you do not send all the paths, you may end up going someplace else that you do not want the packets to go.

John Scudder (co-chair): Contrast this to the current situation where the route-server sends exactly one path.

Saikat (Cisco): This is true. The point is that it is better to send all-paths so that you do have all information.

Rudeiger Volk: I can see the use of add-path on route-servers, but I would suggest instead make this a configuration parameter. I do not think that the case for negotiated limits to cut back the memory on the client. I do not think this is the dynamic nature of the protocol.

Thomas Mangin (Exa Networks): One thing to look at is the scenario of the current route server to add-path enable route-server. The route-server may have to do a partial calculation for routers which do not have the best-path enabled. This feature can be considered an extension to the document.