

BGP Autoconf Meeting on Tuesday 1/19 at 10-11 ET

Audio recording at: <https://youtu.be/NGwomqt9MF4>

Jeff: I started by looking at all of the drafts and documents?

Minutes notes: I think this included the unpublished documents I mentioned last week:

- draft-dt-idr-bgp-autoconf considerations (Jie Dong with Randy Bush credited)
- draft-ymbk-bgp-discovery-layers.xml (Randy Bush notes)

A partial list of items needed for the document (and likely more after I complete review):

- BGP Version needs to be in the discovery protocol - and potentially permit for multiple versions,
- GTSM details,
- BGP Identifier needs to be in the protocols along with implications on tie-breaking if more than one discovery protocol is in use at same time.
- What to do when discovery protocol goes "down".
- Routing implications to non-directly attached addresses. This includes BFD impacts.
- A partial list of items needed for the document (and likely more after I complete review):
- BGP Version needs to be in the discovery protocol - and potentially permit for multiple versions,
- GTSM details,
- BGP Identifier needs to be in the protocols, along with implications on tie-breaking if more than one discovery protocol is in use at same time.
- What to do when discovery protocol goes "down".
- Routing implications to non-directly attached addresses. This includes BFD impacts.

Kausik: What about Layer 7?

Robert: Some of these things are included in aft-ymbk-bgp-discovery-layers.xml.

Warren: Are these requirements complete? Were there any requirements listed that we missed?

Robert: I have a question/comment. What I hearing is that it is out of discovery that there is no "IP reachability" (e.g. peers on the same subnet)? Is this perfectly valid for part of the auto-discovery? The other part of auto-discovery is to work on BGP auto-discovery where there is connectivity within the domain. This other part of the work is the [subject] of the draft that I co-authored with Warren, and others. [draft-raszuk-idr-bgp-auto-discovery-06.txt].

Editor's note: Does Robert equate domain with AS?

Jeff: Your drafts are the drafts that I have not gotten around to re-reviewing again. Am I correctly understanding that your focus is “what you do” when you have IBGP peers and you do not know what to do?

Robert: The draft is about bootstrapping the BGP peers by passing a dedicate SAFI on the RR. [The function is what matters whether you] Call it an RR or Call it a controller. This is what Kausik was alluding as well. The reachability facets can be done using DNS for the name. DNS can just resolve the name (to an address). All the information regarding the BGP peers could be verified through a script (?against policy).

Warren: Maybe I misunderstood. I had thought the bgp-autoconf team was specifically focused on data center things.

Robert: This is a perfectly good question.

Warren: The other document (or documents) is sufficiently separated from the data-center environment that these documents can continue in parallel outside the design team. It is solving another problem. What do the chairs think?

Jeff: The easiest thing to determine if something is out of scope is to list it as “in scope”.

Sue: My earlier understanding is that this group was first focusing on bgp auto-configuration for data centers. Now, John was watching this group. Did I understand this correctly?

Sue: Whether the actual discussions went beyond that to IXP/ISP/WANs is important to me. I would like to see an initial draft on this topic from the design team or a group of authors to spur IDR to action.

Warren this information was in the slides from last week. First Data

Sue: My slides last week encourage this design team to get the data center requirements and considerations draft done by February 18th so that it can be considered in the IETF meeting in March. If I have someone who wants to work on the IXP/ISP/WAN documents as a DT document, I’d be thrilled. However, the first priority was getting the bgp auto-configuration document delivered to the working group. Is the goal clear this week Warren?

Warren: Yes, I think it is.

[Time 4:45]

Linda: For the data center and the WAN there will be different requirements. Do we assume for the data center that every node and every link as BGP enabled?

Warren: I do not think we assume that BGP autoconf exists on every node and every port. I think that the [nodes] things in a data center are much more likely to be IBGP peers. BGP peers are more likely to direct layer 2 connectivity such that any packet you send is likely to get there [over layer 2]. Things are more likely to be a single hop away. Of course, these statements are huge generalities since there are exceptions to these rules. BGP policies are likely to be less complex. What we build here may be

applicable to the WAN case, but this is not the focus of the first document. If it happens to if it happens to work across the Internet that would be good.

Linda: If these are the assumptions in the data center, the BGP auto-configuration will be a simpler than the WAN case. It would be nice that we could have assumptions for data centers and WAN assumptions in the front of the requirement document. This would help the DT have a better focus.

[Sue's Editing comments: I second the need for clear assumption on what DT and WAN topologies are.]

Jie: yes, this is a good idea.

Jeff: Where this overlaps a bunch of things. Why I am not trying to talk about the WAN or DC cases is that we will end up is here are the requirements for the elements [of data] that BGP needs to know to bring up its peering sessions. This set of data is going to result in an information model and a data model that lets us say "here is how to bring BGP up". And some of those things will be generally applicable to most scenarios and some of these things will be applicable to just a few cases. Obviously there are the transport requirements to do the discovery will be the piece that varies.

Jeff: What I'd like to do is to make sure that we have the information model put together for DC, I'd like to have done at least enough analysis of the other scenarios so that I can "these things are the same" and "these things are different". The motivation for that [the process] is the identification of the PDUs that need to be sent to get the BGP peers to talk to one another. How these [PDUs] get sent is the longer conversation regarding transport. The [identification of the PDUs] does have the impact [since it] determines whether the PDUs are a fixed set of PDUs or a flexible set of PDUs. Is this a generic set of PDUs that you tack on a few PDUs for each scenario? Those things have a radical impact on the design of the protocol. The transport discussion will fork.

Jeff: Does this make sense everyone? I do not want to do this as proof by assertion, but this is my thinking.

Acee: I think this is a good path.

Jie Dong: I think this is a good approach. We have agreed what was to be carried last year. We can conclude on that path first and then move on to the transport/encapsulation.

Acee: One of the things we had in the LLDP draft was whether you were going to use TCP-MD5 or TCP-AO authentication in the discovery path. I see you have GTSM, but you did not include the other authentications.

Jeff: I had not mentioned it, but as you recall from last Thursday's discussion is that security is massively under specified in every single proposal. Most people are leaving in a spot for security information.

Acee: People have encodings but we do not how this would bootstrap up and different cases.

Jeff: It is actually messier than this description. Let me walk you through the two distinct cases. What we are not actually covering inside of the discussion to this date is what information is being secured.

Jeff: What we have in L3DL is a counter example. It is modestly well covered that it is covering the discovery information in L3DL. This is great, but it is helping you build the trust relationship for that protocol layer. The auto-discovery is a thing [protocol] subject to security. The BGP session which you set up from the L3DL is a distinct element which needs to be secured. Where things get messy is the cases when you want more than one mechanism.

Jeff: For example, if you are not willing to accept “no authentication but are willing to use IP-Sec, AO-TCP or TCP-MD5. For each of these you will need key chain hints. Now, you have taken the problem to identifying a set of IP End points. Do you need to parallel v4 and v6? Do you need v6 over v4 or v4 over v6? There are all variation that draft-xu-idr-neighbor-autodiscovery-12.txt covers a little bit, but at least the coverage is better than other items.

Jeff: Once you’ve decided to accept a given set of transports endpoints set that you are willing to do, how do you decide on a set of security information that you are willing to transport information over. That’s not hard, it overlaps a little bit with what the IP-SEC calls security associations. However, we are looking for a set of things to try [in the sequence] of preferences. This is a big section all on its own.

Warren: We are discussing the requirements and not the full design of the security features of the protocol. Some of it can be punted to we believe we need authentication.

[12:44]

Sue: I will like to clearly see [specified in the document] the difference between requirements and the review of proposed [in the DT document]. This is one thing I liked in the structure Jie Dong had. Is this you feel comfortable in the design team document?

Jeff: Part of the point of noting that the authentication proposals are massively under specified is the [possible ?] requirement we end up needing is the requirement that states you [the node] needs to announce exactly one authentication that you are willing to do. In this case, you are precluding that the authentication could be different between the different parties engaged in the protocol.

Jeff: If there is some method by which you can advertise more than one authentication type, this has two impacts. First you must have an operational proposal as well as a design. Second, you must also state up front that any proposal has to allow for more than one.

Sue: I am not specifying a particular solution, because I am just monitoring this discussion. I am simply stating that I wish clarity.

[15:00]

Warren: I think that some of the proposals are not fully fleshed out because people do not want to specify the entire solution if it is not going to be used. People would say “we are not document the details here, but do not worry we can add extra TLVs here.” We are just not document here because it is a proposal and that might be a waste of time.

[15:20]

Jeff: This is the point of doing the requirements end of the specification. I know that Randy Bush would be rolling his eye balls. As much as Randy does not like over complicating protocols, these security aspects have things which must be thought through. We do not have specify in the requirements document what these things look like, but we need to specify that solutions will discuss these issues.

Warren: There are many people who are “ok” to do authentication, but happy to “not do authentication” as well. This is where it strays out the data center into other things. Many people the MD5-TCP as really strong checksum and not a security authentication. We can talk about this in the requirements.

Jeff: The document states up front that we are not intending to drive this document toward RFC. This document simply publishes to the IETF what the requirements are. This [status for the document] has a specific impact that if this tried to be an RFC, the Security ADs would require that all the details are specified. Because this document is not destined for RFC, this is acceptable [editorial paraphrase].

Kausik: Jeff are you stating that the part of the document should capture the security requirements as a section. The requirements of TCP-MD5, TCP-AO, and others should be captured.

Jeff: Using requirements language, we have to discuss:

- a) this is for security BGP,
- b) this is a discovery protocol,
- c) a proposal must discuss what the security mechanism is for securing the discovery protocol, and
- d) [since it is going to be carrying information for carrying the BGP session and we are providing end point information], there must be enough information for BGP to successful establish a transport connection over the authentication specified.

The authentication specified could be zero authentication.

The third set of axis is operational implications. Warren and I were discussing on Friday the trust domains can vary widely in the DC environment. If you are doing an L2 discovery protocol with something at the end of the link and you are willing to trust the entity because it is at the other end of the link, then “no authentication” may be an acceptable choice.

Or you can decide you do not trust the people who are managing you wires so must have this [level of security]. So the operational security section has to be done for each of the protocol proposals.

Warren: From an operational side, there is a trade-off between the many different axis [of design]. If do not carry about authentication because I trust the people plugging my wires, then I can almost [but not quite] unbox a switch, plug it in, and magic happens.

Warren: If I do not trust the people plugging things in, then I have to do a bunch of pre-configuration on the device before it get s plugged in and turned up. And so this has implications on how much of the auto-configuration stuff happens by itself. How much needs the orchestration system needs to do.

[20:00]

Warren: I wanted to understand that Jeff and I are editors on the document. This [editor status] means we take people’s review and comments and things. Something that is really important is that people need

to be reading the documents provide comments so we can be editorial. Rather than make stuff up and put it in whole cloth.

Jeff: An important detail is that Warren touched on is that this group is labeled bgp-autoconf. We are acting effectively as the bgp transport auto-conf group. A bigger solution such as having your box being automatically plugged and put itself into a switch fabric has other implications besides the BGP transport. The auto-configuration outside of transport has more information required. This document must consider if there are requirements beyond the bgp-conf transport that require things within the auto-configuration for the BGP transport to help the upper layers figure out important things.

Jeff: This is one of the reasons why role has showed up in some of the proposals. If I say I a transport endpoint and my role is a CLOS fabric level 1, then your device must be smart enough to use that information. Can we move this information from transport to BGP as part of the open messages? This is a possibility as well. This is where one the interesting scoping lines will need to be discussed in the document. How will you allow for the full auto-configuration of BGP ZTP (zero touch protocol) ?

[22:00]

Sue: Are you going to define where you are restricting your data information? Are you going to discuss this restriction in your original draft?

Jeff: The initial draft is going to try to point out for the information model the information that is required to get the transport job done [up and running]. We'll try to throw in first level discussion on "if you need help from the transport layer", it may be long here. Once you can form a transport session, the line blurs quickly. At that point, BGP open can do much of the work for you. The leaning of the DT was to do as much in the open as you can, but if you cannot get to the open – this is the type of requirements we are working on.

23:20

Sue: Any comments or questions? Jeff and Warren is there anything you need from the design team to an initial draft for people to poke at.

Warren: People need to read and re-read the existing set of documents so that they are familiar with the document. It is important to point out where we have missed things or where we have gotten incorrect. A comment might be "this proposal solves this thing" which I think is actually a requirement. Can the other protocols also do that? Is this really a requirement? It is important to determine what facilities a protocol provides so that you can consider requirements you might not have caught.

Sue: So then, I'm going to strong encourage everyone to do is to re-read the documents one more time. I read the documents last week, but I am going to re-read the documents again this week to be ready for next week. Jeff and Warren do you think have a few more bullet points or a document by next week.

Jeff: I am hoping that by the end of today (1/19) or tomorrow (1/20) that I'll have the information integrated so I can pass around I think we need to do.

Jeff: I need to re-read the github procedures. I should push out private branches. I lost track of what IETF says about github.

Warren: IETF github procedures are still up in the air so that you can just stick it in the main branch, and we do not push it as draft up to the IETF. My concern is that we have 27 different sub-branches.

Sue: My reading of the two RFCs was that github policy was WG or DT specific. If working in the main branch is easier for you, then please use the main branch.

Acee: I do not see any reason if you two are going to be editors and the rest of us are going to be reviewers, I do not see any reason why you could not publish it. Or do you not want people to read the first draft. With only two people, I do not

Warren: The RFCs on github are not IETF policy and procedures, but ideas on github. If you want to deviate from these procedures, then this is “fine” too.

Jeff: Github allows us to open issues and to track the resolution of the issues. The discussion does not happen in the system. If we get down to this issue must be addressed, we can use the issue tracker. But I do not have a strong opinion about doing this. I quite happy wrangling this [issue resolution] through email.

Warren: People use either type (issue tracker or email). I think email is fine. Some people feel that using the tracking in github is just fine. This

Jeff: We’ll have a better sense once I’ve done an integration pass and put all the things in one document. From there, we can decide what the actual work is.

Sue: All of that works for me.

Kausik: Do we get these slides or do we wait for Jeff to come up with something?

Sue: I’d like to have something we work against because that helps [focus] our work. The reason I asked Jeff what is timing was [to deliver the first document]. If the first draft is close, then this is a good way to start the discussion. In the meantime, there are a lot of documents to read and re-read.

Kausik: The last point I did not go through was the multiple discovery facilities.

Jeff: I’m not precluding any sort of discussion at this point. If you have opinions about something, please send these to the list.

Kausik: What is your thought on the multicast discovery?

30:00

Jeff: Very tersely, once we decide what goes inside the bgp auto-conf protocol we have multiple flavors of multicast involved. We have outright unicast (e.g. ARP discovery). For multicast, we have L2 multicast (e.g. LLDP) or L3DL form uses the IEEE 802 multicast (ethernet multicast, single switch or multiple switches). The IP multicast leverages those Ethernet MAC addresses to flood across the multicast domain (all things on this link/L2 switching domain). Each of these things has implications for this environment. For example, if you are link that is not really Ethernet but a point-to-point function (logically) the L3 multicast might tell you there is one thing there. However, there might be more than one thing. We need to decide what to do about that facet.

Jeff: We have the switch fabric case. You may want to decide to communicate with the direct opposite of the port. In which case, you may want to send to just that port. Since there are multiple domains, potentially these protocols are serving other purposes [besides bgp auto-conf]. it is necessary to decide how to do this.

Jeff: An example there is that [Juniper] has an internal implementation that is based on ARP and LLDP. If you have both, you need to decide what your precedence is if you have information from both.

Jeff: A different example is if you have something like L3DL. You have data forms for carrying BGP discovery, but that is not its primary purpose. What do with the L3DL when you have conflicting purposes. These are all discussion points.

Kausik: Ok.

Sue: From this week's work is that Jeff and Warren are going to get together and put a straw-man document out. The rest of us are going to read the documents again that were sent out last week. I will send the list again. We'll review and comment on the bgp-autoconf list. We'll get together next week to have any in-person discussion.

Sue: Jeff and Warren - It would be helpful if you as editors would suggest points that you need clarification on by Monday next week. Is this a good way to work?

Warren: It would be good as people think of things to send emails into the list as well. This way we can make sure we do not miss anything.

Jeff: No I think I've done

Sue: I will try to listen to the recording and send out notes. Thank you all for your time. I appreciate you re-engaging to get the data center requirements documents out.

that I believe we are going to end up.

can continue in parallel. It can be

The other part of