

MONDAY, July 17, 2017

0930-1200 Morning Session I

Congress Hall II RTG rtgwg Routing Area Working Group WG

RTG WG update

Jeff Tantsura and Chris Bowers 10 minutes

Meeting starting. Note well applies. Note Well has been updates since last meeting, please be familiar with it before you publish, present or disclose any information. IPR poll will be done before adopting and last call the documents, it will not progress unless we get each and every response on IPR. Agenda bashing.

JT: Friday's agenda is much shorter, likely will be shared with MPLS.

JT: We have one new RFC published since last meeting. A good example of how YANG model work could be done. Yang RIP and VRRP modes are ready to be published. Routing types and uloop drafts waiting for write up.

JT: Would like to see more discussion on the list on the active documents.

Bruno Decraene: what is the status of SPF backoff algorithm?

CB: I will ask Jeff speak as I am coauthor.

JT: We think it is ready to progress.

draft-ietf-rtgwg-routing-types

Acee Lindem 10 minute

A simple yang document. The intent is to put common types used for routing in a single place. it is corollary to ietf types and iana types yang documents covering the routing types.

JeffH: for the record, the last lingering item is for extensibility, we still need to resolve it.

Acee: It is covered in my presentation later.

Benoit: There are two different regex testing tools and I'd appreciate some testing on both of them. We put two outputs so be able to compare the two tools.

JT: Benoit, could you send email to rtgwg on this?

Acee: if we move the geo-location to a separate document, can it be a WG doc?

JT: this draft has been under LC planning for quite a long time. You can make the changes of removing the geolocation grouping.

CB: If you are asking whether geolocation is removed, can it go immediately to WGLC?

Acee: I'm asking whether it can be a WG doc.

JT: after the change, we'll reissue WGLB. There are other comments that came very late. After you resolve the changes we will issue the lc.

Jeff H: The target types are part of the module that is not maintained by iana. Is that intentional?

Acee: I don't want them to be myself. Do you think it should be?

JH: we should put them in iana so they actually have iana maintained. If you do not make it iana maintained then WG needs to pick a new document every time it is changed, it's up to the WG.

Acee: it is a generic extcommunity registry, this is different. It is not completely aligned with IANA. We do not have one type that does all extcommunities.

JH: We need to decide whether we touch it every time there is a change. That seems to be fundamental.

Loa Anderson: You said that you will take the discussion on label stacks that will be on Friday. That should be after the rtgwg meeting.

draft-ietf-rtgwg-ni-model and draft-ietf-rtgwg-lne-model

Lou Berger

20 minutes

Sue Hares: Asking you a question as IDR co-chair. We have a model that has been adopted for over 2 years, 15 implementations. Do we need to redo that too?

Lou: This is ADs message. I will ask the ADs in this room to respond to the question.

Alia: There is a reason it says should. This is a desirable thing, it is case by case, and BGP model falls into the discussion category.

Lou: the BGP model doesn't even compile now. From an IETF perspective I would expect some discussion in IDR to talk about what the plan is to take it forward in a way that is functional. Whether to make it NMDA compatible it's up to the IDR WG. From the DT perspective we have looked into that, it doesn't compile.

Sue: That is a historical moment here.

CB: There has been much of the uncertainty on this question for this for a long time and the fact that it does not compile now should not be seen as a criticizing.

Acee: it references types that's not IETF.

Chris B: Do not claim that it is a bad model if it did not keep up with the changes for a couple of months. Let them decide the NMDA.

Lou: things can be taken case by case, and this is a transition period.

Kent: the notifications are a generic mechanism, not specific to the routing model?

Lou: The notification name is find LNE failed, so the notification is pretty specific. Not sure what would be generic here as that is tied to notifications for routing uses.

Kent: right now the client would do a diff between the operational and the intended to figure out what actually got applied but there was a desire for there to be some sort of notification a generic mechanism. Is there a desire to have a generic mechanism for notifications?

Lou: we create a notification in case of failure. I think if there is a better way to do that in yang we are happy to change.

Kent: I don't know.

Chris H: Netconf processes things serially, it's a bigger generic problem. this is good if you need to process things out of order. There is no mechanism to report a failure on an async operation. This is applicable to routing. We need a generic way to resolve this.

Lou: do you think we should remove this?

Chris H: I think we should remove this and yes a more generic construct but that needs to be discussed with netconf first.

Lou: for the notification issue, do we want to resolve it before LC?

Jeff T: We would like to get this resolved before the LC as otherwise that will be a long process.

Lou: Please discuss this on the mailing list, you may not be the only one having questions about this. Any discussion that helps clarifying the things helps to move those two documents forward.

Guidelines for YANG module authors on using the new Network Management

Datastore Architecture (NMDA) (draft-dsdt-nmda-guidelines)

Robert Wilton 15 minutes

Robert presenting.

Sue Hares: Thank you for the excellent bgp example. I would encourage you to present this in IDR. And hear what the operators have to say. Did you ask operators what does it mean for them to change the model of BGP now?

Robert: Personally I do not know that. At the moment there is hesitation towards the IETF models as they are not considered to be standard.

Sue: A correction: It is adopted regardless of the status of the standard level.

Robert: You have more of openconfig model in the IDR.

Sue: That is incorrect, but let's continue. You think there is a good reason for operators to make this change?

Robert: yes.

Sue: This is a rewrite of the model. The summary is that there is less space use and more consistency of the model are the summarized benefits?

Robert: generally yes. More consistencies in many ways, in terms of implementations. One of the issues that Openconfig models has today is that they are not strictly complied with Yang. An example of system created interface, technically you're not allowed to have it in the config model because you're creating a config node to represent a state. They bend the rules a bit with creating the leaves in the config. They can't use choice or presence container because they don't work in OC style. So it's not just being cleaner.

Sue: The tool chains would be easier to support given this better usage of YANG.

Acee: Just a data point, we did a survey and we did not have any implementation that cared whether we changed the model (this is for ospf). For OSPF were not many cases where intended and applied delivered different values. The size of the model was more than a third reduced with NMDA.

John Messenger: Another advantage of this is that there is less duplication and it is easier to maintain.

Benoit: I was expecting that you will explain the tooling and how to better use it to move to NMDA, can you say a few words about it?

Robert: The easy approach is to create an extra state tree. The other approach is to take the existing IETF split style model and generate the combined tree. It's not perfect yet. I need to check how that works with grouping.

Benoit: It would be great if you could document this and post to the list.

Robert: yes.

Alia: When we are talking about this there are many reasons for doing this. The meat of the model is abstraction. This is the presentation layer on top of the model. It is a presentation style difference, it is not content style difference, and that is the important. The models may look that they change a lot but that is presentation layer, not the content.

Igor B/Huawei: We have discovered that openconfig style is quite awkward when you use groupings. For example, if you take a piece of tree and use it as RPC input and output, having state everywhere becomes really awkward, and NMDA style solves it very nicely.

Chris Hopps: I know there is work being done in netconf to support this new rpc. Is it also be worked on for restconf?

Rob: Yes.

Xufeng: It seems that there are quite a large number of cases, it is not a few only.

draft-bashandy-rtgwg-segment-routing-ti-lfa and draft-bashandy-rtgwg-segment-routing-uloop

Ahmed Bashandy
Ahmed presenting.

15 minutes

Chris Bowers: Why does the draft restrict SRLG protection to local SRLGS?

Ahmed: The algorithms that we have work on local SRLF, but the idea of constructing a label stack can work on any SRLG, local or remote.

Chris B: Are you describing the algorithms in the detail that limits the applicability to local SRLGs? Why wouldn't the document discuss remote SRLG as well?

Ahmed: For future versions that could be just added in. I have to scope it to something,

Chris B: I would suggest adding in remote SRLGs too.

Ahmed: The algorithms are very different.

Chris B: I am aware. I just suggest to add remote SRLGs.

Ahmed: The algorithms are different and that would make the scope of the draft much broader. Maybe in the next versions we can decide on that.

Chris B: so you specifically don't want to cover it.

Ahmed: not in this version of the draft. Maybe later.

JT: I would keep the IPv6 discussion out, there is 6man wg for that.

Stewart: in the draft text on abandoning the mechanism when the constraints aren't met.

Ahmed: that's implementation details.

Stewart: That is not an implementation detail, you need to signal to all the routers.

Ahmed: no, I have to configure my router to do link protection only. Why do I need to tell a router to do a link protection for the prefix only?

Stewart: remember we're talking about when the conditions are not met. You may not be able to push the right number of labels as an example. Other routers are trying to do post convergence and loop avoidance.

Ahmed: I do not see the relationship here. That is a local behavior, other router knowing it or not is irrelevant.

Stewart: It is LF convergence process that starts afterwards is interesting. It may escalate as you discover more information. In that case you need to abandon any LF and move to best effort convergence.

Ahmed: If we started with link protection and later discovered that we need to do node protection. This is implementation detail.

Stewart: I do not think this is implementation detail. You have to describe the process you're going to run. Not everyone will work out the subtleties. This is about what you want people to do. The second point - you go into loop avoidance and that takes time, and you need to abandon loop avoidance which other people help you with.

Ahmed: Loop avoidance is a local behavior. It's irrespective of what happens downstream, it works whether downstream decides to do TILFA.

Alia/no hats: I've thought a little bit about faster out of course. if you have a worse failure so that your failure assumption doesn't hold you get loops, and the rest of the network if you're trying to do particularly for anything like this which is more sophisticated than LFA, you want to also do micro loop avoidance, right? and that assumes that the node at the point of local failure is doing a local repair and continues to hold that repair in while the reconvergence happens elsewhere, and if you're trying to do the micro loop forwarding avoidance, and if the point of local repair is not holding a repair because a worse failure happened, then you also want to terminate the micro loop forwarding prevention and go back to let's just converge as fast as possible and therefore it is not local behavior.

Ahmed: first of all the assumption that micro loop avoidance requires the node attached to the failure to do something is wrong. I do not care about what the node attached to the failure do whether it does the TILFA or not it works.

Chris B: you are presenting one particular uloop avoidance technique, but in general other uloop avoidance techniques will require it, so Alia's point about worst failure case always holds.

Ahmed: This is for SR TILFA.

Stewart: We need more text in the document to explain this.

Chris B: you cannot simply say the uloop avoidance draft forms a part of the architecture that it works. If that is the case you need to state that.

Ahmed: it's for SR micro loop avoidance. I don't know how to be more specific about this.

Stewart: Even SR has partially strict routing.

Ahmed: The title mentions SR.

Alia: We are asking for text that interacts with other uloop avoidance techniques and when do you need to bail out to best effort. One reason it doesn't have to be local, second we are trying to get the explanation to people who are not here and who plan to make implementations. If you get it wrong you loop traffic.

Ahmed: the implementation of this technique is to certain failure.

Stewart: We disagree.

Stewart: Another question - a lot of traffic will not go near the PLR after the failure. How does this work in the real network vs hypothetical one?

Ahmed: What is the concern here?

Stewart: It may have bandwidth availability implications?

Ahmed: I will only force traffic if it comes to me over the post converges path from me to others if it doesn't come to me I cannot do anything to it if it comes over there. I still don't understand the concern.

Jeff T: Summarizing - there is definitely clarity needed in the draft and WG members asked for the clarification text.

Ahmed: I will add SR in many places in the draft.

Jeff T: That is not what is asked for.

Chris B: Add text so that the experts on the topic of the FRR can understand what you are proposing.

Ahmed: Ok, for me it is quite obvious.

Chris B: You mentioned the IPR disclosure, you need to do that. The general rule is to ask before the WG adoption.

Ahmed: There was one for this version.

Chris B: Have you already disclosed all IPR?

Ahmed: Probably. I'll check.

Chris B: We will ask that question to the WG.

Alia: There are new IPR disclosures on this draft, please read the note well. The point is to disclose IPR at the earliest point.

Alia: for the link down case, how do you tell what kind of the failure has happened to know what the post convergence path is?

Ahmed: again this is configuration driven. You come and tell me do link protection, okay so I will do link production. You come and tell me do node protection I will do node protection. It is not that that I know that through some funny mechanism because it's really impossible. It's the other way around.

Alia: I was thinking more about local SRLG so what you're actually going to is the post convergence path if the worst case failure were to happen.

Ahmed: no I would say this is configuration or policy driven. You tell me what you want.

Alia: What you are doing is that you are finding the post convergence path for the configured worst case failure. What if the failure that happened is actually different?

A: That is an operator aspect, they configured it.

Chris B: do you take into account SRLG in this case?

Ahmed: For SRLG we do not take care of loop avoidance in this version of the draft.

Chris B: Remote SRLGs cause problems and they were left out - you should explicitly state why you are leaving it out.

Chris B: No. Node protection itself is a form of SRLG. Once you are in the protection topic, you need to explain why you are restringing to a subset of SRLGS.

Ahmed: I am restricting it to a certain problems. I don't need to explain why I'm not addressing other problems.

Stewart: it depends on what the discussion on the list is. Now I'd like to bring up another sort of topic, which is something that's always puzzled me when we do fast reroute in a segment routing environment. That is presumably we routed the packets through a number of hops for a particular reason, so presumably we need to apply the same policies that was applied to those selections in the selection of the alternate route. For example if the original policy was set up to avoid going over a particular link for example of security a reason, you want to make sure that your fast reroute doesn't avoid those policies. So I think there needs to be some discussion about policy which wouldn't normally apply in a pure best effort Network.

Ahmed: this draft protects normal shortest paths as calculated by IGP as well as SIDs.

Hannes Gredler: You are overselling here, you are saying topology independent but that is not topology independent, but carefully escaping the difficult cases here.

Ahmed: It does not matter how your network is stitched together this will provide you a protection.

Hannes G: You are assuming an overly idealistic model here that does not correspond to reality.

Ahmed: Topology is connectivity. There is a failure scenario when two links fail at the same time. I am not covering that. I'm restricting the protection to certain failure scenarios on any topology.

Hannes: That is a very realistic scenario.

Jeff T: You need to provide clarifications before we can progress.

draft-przygienda-rift

Tony Przygienda 15 minutes
Tony presenting.

Greg Mirsky: Do you refer to single hop BFD here?

Tony: Yes, there is no application for multihop RIFT.

Greg: You make this explicit that this is single hop BFD?

Tony: No, but we can clarify that.

Greg: I am interested as I had discussions with DC architects that you do not need BFD as the hardware can detect things very fast.

Tony: We know from experience of decades that Loss of light is not a reliable fault detection mechanisms, otherwise BFD wouldn't exist. BFD does an excellent job.

Acee: For the southbound TIEs you always flood self-originated. What happens that at any level you are injecting prefix TIE for your loopback, or only when the node looks the connectivity?

Tony: We recommend that you inject t always, which makes you always reachable from southbound. The problem is real today.

Acee: You do not use E-W links in the failure if you lose the link. Does that happen automatically all the time unless there is a failure?

Tony: Yes. That's why it's built that way.

draft-white-openfabric

Russ White 10 minutes
Russ presenting.

Acee: the it's really area information and you're repurposing scoping it at the link level to limit flooding it, seems like kind of a strange mechanism.

Russ: Yes. But it seems to be the simplest solution. the main problem you run into with isis is there's no header bits remaining to do anything with, so you can do different mac addresses which Lou Berger suggested it on lists. Or you could do a new LSP type which is what link local flooding basically does, or there's other games you can play. But essentially since there's no header bits to play, there are very limited options here.

Acee: but you still have to use it. you use the information at an area scope even though.

Russ: the it's actually a link-local only. You have to treat the received LSP as if it's a normal LSP but you simply don't set this SRM bits so you don't reflood it.

Acee: in long term there might be a better solution. I don't have it.

Russ: yes, if anybody has suggestion please send it to me.

Xiaohu/Huawei: For each level of the flooders, have you considered to have a backup one?

Russ: Yes. That's in the draft. The goal is every IS only receive one copy of the lsp. however if you are concerned with weird timing problems, where you don't then you can always choose a

backup and make it where everybody gets two copies. That's in the draft. The other thing is that in the draft it talks about something Les convinced me to do. There's actually a follow up CSPF so that you know that everybody's still synchronized, because that's actually simpler in ipsis to do. You can actually see the failure to flood. It's actually pretty simple mechanism. So there are two ways of handling that in the draft right now.

Les: You're talking about CSNPs.

Russ: yes.

Jeff T: done for today. See you on Friday.

FRIDAY, July 21, 2017
0930-1130 Morning Session I
Congress Hall III RTG rtgwg Routing Area Working Group WG

Chris B: Meeting starting with Alvaro giving an update on DC routing,

Update on Routing in the Datacenter
Alvaro Retana 5 minutes

Alvaro: we've been discussing here in this working group for the last maybe a couple of IETF cycles some proposals and other things around routing in the data center. We had some presentations on Monday, so there's been a little confusion of what we're gonna do. I think a plan is going to be going forward, I will send an email probably next week to the working group so that everyone actually knows. I think we all realize that there are specific requirements or circumstances inside of data Center, what we have to do with the topology, so what I have to do with the way of potential routing protocol surrounding solutions to behave inside the data center. So that could lead us or has led us already to define solutions for the data center that are not just the regular routing protocols that we have today. So this also means at least in my mind that we may have more than one solution, because not all the topologies are the same, not all the data centers for the same size, not everyone have all that same speed etc. So the plan was if you remember from last time for someone to go off and come up with requirements, and then we would discuss it. Well, that can happen.

Jeff T: it's happening but slow.

Alvaro: so Jeff says it is happening but slow, which to me it means it didn't happen. So what we want to make happen for next IETF is to hold a nonworking group forming BoF in Singapore to do two things: one discuss some of these special circumstances and characteristics that we think we need, second to go over the solutions for some of the potential solutions. This is important not to have a beauty contest or a winner necessarily because we may have more than one potential solution. So the idea is not to say solution one is better than two, because it may be better in some circumstances but not in others. If we do that and if we identify interesting solutions that for which we can also identify interest in the IETF to work on them, then we can figure out what the next steps are. Meaning we're going to work on something specific, that we charter a working group or two, whatever the conclusions at that point. Then we can go forward and figure something out, so that doesn't necessarily mean and that's why I think that there should be non-working group forming that we are going to end up with a working group. it doesn't mean we're going to end up with anything. We not only need to identify whether the work is interesting, also the work should be in IETF and of course there is interest. as much as any solution that any one of you may push, if you are the only one to want to work on that then there's not a lot of point in creating anything inside IETF at all. so Jeff is going to help coordinate so the proponents around so that we can have some discussion at the beginning of that BoF, if you are interested in presenting something in the next few weeks hopefully identify chairs and we'll go forward from there. Obviously all of this is subject to discussion.

Rick Taylor: clarifying question, from my understanding the purpose of the IETF is really around interoperability particularly with routing protocols. This protocol work on my box and yours. Within the data center surely you own everything in that data center, it's owned by whoever is running it. Is there interoperability there?

Alvaro: so I'm going to say yes, there is interpretability. They obviously probably sold the same solution to different people, so from that point of view I think there's well as a valuable point of users of having interoperable protocols like you have in your internal network like OSPF.

Stewart: so I accept the need for doing special routing protocols in the data center absolutely, but I noticed it's part of a trend for domain specific routing protocols, and I wonder whether we might want to go up a level and create a toolkit so that you can build domain specific routing protocols, and then the data center one might be a profile.

Alvaro: sure, there may be something that we can do. One of the things that I want to do is to create a mailing list so that we can specifically discuss that, and take that off the plate of this working group. So we can have clarity and focus on that, so hopefully I'll get that done in the next couple weeks, and we can start having this questions like that. That's another potential option.

Les Ginsberg: there are various drafts in protocol working groups at the moment, what does anything what you propose have to do with how they progress?

Alvaro: I'm gonna say nothing. So I think that that you know there's a class of solutions that of course are around enhancements of routing protocols. There is already work that has been done and published as RFCs that I personally believe could be used to enhance running protocols for data center. I don't want, it's not my intent, to slow anything down. As I said the beginning, one of the important things here is to gauge interest and enthusiasm. if there is work extending specific protocols, and the extensions happen to be around the working group, I'm fine with that.

Jeff T: a few more comments. So there's some new features coming up, new silicon that would enable us to build non classic topologies and we need to start looking into protocols that could address non classical data center topologies, this number one. Number two we see data center getting out of data center, so leaves being the edge which requires quite different set of features in the routing protocol. So long term we need to address all of this, not just BGP and DC which works perfectly fine and hundreds of people have deployed. So it's natural evolution of routing, and long term this is where we would like to take.

Alvaro: thank you.

VPNplus, and enhanced VPN based on Segment Routing
draft-bryant-rtgwg-enhanced-vpn
Stewart Bryant 30 minutes

Stewart presenting.

[discussion]

Chris B: please don't use N word since it is confusing.

Stewart: it was network slicing I was talking about.

Chris B: Since the topic of NS was discussed - this has some relevance to it. I want to make clear that the process of the BoF took on Monday, we are not going to have a discussion of this technology relevance to NS that is done elsewhere.

Stewart: It has a more universal applicability in construction of networks.

Chris B: Not to get overflow from bof.

Uma: this is the closest work for the ns in the transport domain. Question - you said that to create a vpn a sid list is created and packet selection criteria is applied. What is packet selection criteria?

Stewart: we will be expanding the document in more detail. how you select the packets is part of the dialogue between the owner of this packet and the tenant, we can use ACL or netflow or whatever.

Uma: I can take this offline. Specific requirement in one domain is to get packets the GTP packets, and this traffic needs to be mapped to tunnels.

Chris B: Please take offline.

Jeff T: there is a number of extensions to BGP LS etc.

Uma: I am asking about tunnel to fec mapping on the router.

Stewart: we will figure it out when we get to that level of detail.

Adrian Farrel: Two pints - i was one of the chairs of NS bof, the strong feeling was not to wait for an overarching architecture to do what is useful for operation networks. If there is value then do not sit there, do it.

Jeff T: We need to provide a way to consume transport network

Adrian: Is that a statement or an opinion?

Jeff T: both

Adrian: 2nd opinion, I want to pick on the representation of vpn+. as the customer having some influence on the network.

Stewart: No, they don't.

Adrian: What came over in your slides was the user was given a tight coupling with the way how the network was operated and build. I think it should be between the controller and the network.

Chris B: out of time, cutting the line.

Adrian: We may have a radical agreement, a controller based system where a controller has knowledge of the system and operates on the variety of tools to achieve that. I do not think that you are doing more than telling that there are many tools that can do this already.

Stewart: I agree.

Lou Berger: your statement on putting things together - a bunch of us are happy to hear that it makes sense. It is not clear of the details. We have ways of doing te, actn, models and frameworks for transport sdn network. When we have that cookbook that will show how the pieces fit together,

Stewart: I do not want to invent anything new, including detnet.

Lou Berger: You highlighted a definition that we need to be focused more on the area and that is SR-TE.

Wim: I agree this is a good document to put pieces together. There were documents in SPRING that were in the beginning but we were told that those documents do not have value to go forward. Putting together pieces that we have together - but I do not see what new needs to be done to address the use cases.

Stewart: We have two document - the unified SR, and a better version of SFC.

Chris B: cutting the discussion.

draft-arkko-arch-low-latency

Jari Arkko 10 minutes

Jari presenting,

[discussion]

Ning Zhou/huawei: I like the idea of seeing this from the architecture view. It is difficult to achieve by using single technology only for different use cases. Building an architecture view and have gap analysis of existing technologies in ietf is a good idea. Maybe some potential gaps need to be analyzed. BBF has a similar project for the user service. It's a good way to go and we have similar ideas.

Chris B: we do not have much time.

Jari: Thank you. This discussion is partially similar to netslicing. We aim to have a big picture view.

Lou: one of the things that we identified in the existing QoS work - we have a gap and how do we get an information from the application through the transport protocol down to the qos aware transport network and also in the other direction. Maybe architecture side could attack that.

Introducing the Path Aware

Networking (PAN) proposed RG
Jen Linkova 20 minutes

Jen presentin on PANRG.

[discussion]

JT: The recording of the PANRG session is available on YouTube.

CB: end of RTGWWG meeting.