Table of Contents

Table of Contents

CT WG LC comments

CT-WGLC-Q1: Ketant Talulikar comments (7/22/2023) –

Link: https://mailarchive.ietf.org/arch/msg/idr/q1PBTKAnlsuyjYwd4fwpAgCEssI/

Github: https://github.com/ietf-wg-idr/draft-ietf-idr-bgp-ct/issues/30

6 issues below:

1. The document has improved with the multiple recent versions; thanks to the authors. I hope the comments and suggestions below help improve the document further.
2. I have provided comments to cross-check and fix the inconsistent and confusing use of terminologies (e.g., mapping community, color vs. transport class, tunnel vs. encapsulation, etc.) and the use of some new terminologies while we have some existing well-known ones.
3. The document can be trimmed significantly by removing the text related to functionality that are provided by other individual drafts (e.g., MNH, MPLS private labels, SRv6 inter-domain, etc.) to their respective drafts. While those are being referred to as informative references, that is incorrect since the functionality described cannot be realized without those features – therefore normative. Not to mention, it is always better to provide a precise document that focuses on the specification (even if experimental) to the IESG.
4. The draft does not cover the use of BGP CT with SRv6 on its own without dependencies on features in other individual drafts. There are also some details missing. One option may be to remove SRv6 at this point and have a separate document for it when ready.
5. There are some technical issues identified which need to be fixed.
6. I've also provided some suggestions and minor comments or questions.
7. The document has many warnings and some errors as reported by IDnits - these should be easy to fix. There are also spelling and grammatical errors which can be identified and fixed. I've focused only on the technical aspects.

Action items:

1. Terminology clarity
2. Move unneeded functionality to other drafts,
3. SRv6 solution needs to be a complete solution in the draft
4. Editorial issues and nits mentioned by Ketan

CT-WGLC-Q2: Jeff Haas – WG LC CT review (July 17)

[Link mail list: https://mailarchive.ietf.org/arch/msg/idr/vxnJDv5gUnExRctD1tYuy8Zy0Do/]
[Link Github] https://github.com/ietf-wg-idr/draft-ietf-idr-bgp-ct/issues/27

Original Text:

Sue asked me to spend some time reviewing -ct to review its current state. Here's that status. A similar effort will be coming for the -car document soon.

A high level concern (repeated in my stream of review comments below) is that features such as multinexthop, mpls namespaces, rt-constrain extensions and cpr are used in a normative sense in places in the document. While the intent seems to be more in the lines of "if this feature was deployed, here's how it'd behave", that pushes it more towards normative than informative in many of the cases. My expectation is that during IESG review the same point will be made. I suspect the best answer is to split those cases out of the core CT document, or target doing so for final publication. (Note that we're in the midst of WGLC, so that's probably NOW.)

Note that I also haven't carried out the document cross-check of the text vs. the opened github issues. I believe Sue is tracking that.

-- Jeff

I spent some energy reviewing the draft in its current state with a focus on content rather than generating a grammar diff.

Things in need of some work:

- Some terms get use without definition. Service Node/Border node, for example.
- Over-use of commas in many places.
- Parenthetical remarks made that are part of the main sentence and shouldn't be parenthetical.
- Small things like the latin needs commas after its use. "i.e.," rather than "i.e." Removing "For" prior to "e.g.".
- inter-as option A/B/C are used without reference.

Many of these will get cleaned up by the rfc editor.

Figure 1, the "architecture" isn't really an architecture - it's an example topology leveraging multiple of the components.

"import processing" is used in a non-RFC 4271 normative fashion. What's intended here is processing of a route received by a bgp speaker and its interaction with import policy. I'd suggest that rather than do bgp-speak everywhere, "import processing" should be defined in the terminology section.

Issue, Section 5:
"The first community on the route that matches a Mapping Community". "First community" isn't well defined in community processing procedures. Implementations tend to locally canonicalize the received set of communities and may internally re-order them without regard for the order of the routes as received in the PDU. What I believe is intended here is the first matching community according to the locally configured policy.

If there are multiple mapping communities on a route for whatever reason, this order may cause different nodes to take different mapping actions as part of their processing. Such behavior may lead to inconsistent routing within a domain. The document does state, "If a route contains more than one Mapping Community, it indicates that the route considers these distinct Mapping Communities as equivalent in Intent." There likely needs to be a caveat that intent equivalence needs be wary of consistent route selection across intents to avoid forwarding loops.

Issue, Section 6:
"Note that the length will always be the sum of 20 (number of bits in Label field), plus 3 (number of bits in Rsrv field), plus 1 (number of bits in S field), plus the length in bits of the
Prefix (RD:IP prefix)."

The text above already referred to the multiple-label behavior as covered in RFC 8277. The related math within the description of the Length field needs to account for the fact that more than one label may be present.

Issue, Section 6:
"When the length of Next hop Address field is 16 (or 32) the next hop address is of type IPv6 address (potentially followed by the link-local IPv6 address of the next hop)." This needs a reference to RFC 2545.

Issue, Section 7.3:
"If the resolution process does not find a matching route in any of the associated TRDBs, the received BGP CT route MUST be considered unusable for forwarding purpose and be withdrawn."

While the intent of "and be withdrawn" is intended to say the route isn't usable and should be removed from the Adj-Ribs-Out resulting in a previously advertised path no longer being advertised, it'd be better to simply end this as "MUST be considered unresolvable. (See RFC 4271, Section 9.1.2.1)"

"The received BGP CT route MUST be added to the TRDB corresponding to the Transport Class "C1". So that service routes can resolve over this BGP CT ingress route. RD is stripped by the ingress node from the BGP CT NLRI prefix when a BGP CT route is added to a TRDB."

This could probably be clearer, especially since the text in prior sections drew the analogy to RFC 8277 rather than RFC 4364 L3VPN procedures.

Similarly, it's not clear if the "MUST be added" is covered by the case where the route is unresolvable.

Please consider re-working these paragraphs.

Issues, Section 7.6:
"When multiple BNs exist such that they advertise a "RD:EP" prefix to Route Reflectors (RRs)"

This is the first occurrence of RD:EP. Personally, I think "endpoint" is good terminology for describing the ip prefix contained in the CT route, however it's not consistently used in sections prior to this point. It'd be useful, for example, in the section describing resolvability. It'd also be useful in the section covering installing the routes in a specific instance of the TRDB as the component that exists after the RD is stripped.

The scenario described here also says you SHOULD use add-paths. I'd suggest restructuring this section to lay out motivation first: If it's the case that multiple instances of a given RD:EP exist with different forwarding characteristics, then add-paths is helpful.

Issue, Section 7.7:
This section motivates the scenario where an ABR route reflector resets the nexthop locally. This isn't a common scenario, but it does happen and it's good to discuss it.

However, the mitigation as a "SHOULD provide a way to alter the tie breaking". Doing this inconsistently is likely to cause the problem described when executing this scenario. The remainder of the section offers operational mitigations.

My recommendation is if this scenario is expected to be common, it is necessary to discuss this as a MUST. Further the section should be called out as "changes to the BGP decision process" in its section name.

Issues, Section 7.8:
Section 7.8 repeats the "withdrawn" comment above and needs similar changes.

The discussion of fallback schemes perhaps deserves some discussion about consistency of scheme configuration within a deployment. Inconsistent resolution due to lack of consistency in the various schemes or their various tables can result in forwarding loops.

Issue, Section 7.11:
The precedence mechanism discusses multinexthop as a feature. That feature is currently listed as informative rather than normative. I suspect the IESG will catch this as well.

The authors may need to make a choice as to whether to exclude multinexthop from the draft or remove normative use cases of it in the draft. The easiest option would be to exclude it from this draft and include the precedence as a discussion the multinexthop draft.

Issue, Section 7.12:
Informative use for redirect-ip here may also wish to be reconsidered.

Issue, Section 8.2:
"An egress SN MAY advertise BGP CT route for RD:eSN with two Route Targets: transport-target:0: and a RT carrying :. Where TC is the Transport Class identifier, and eSN is the IP-address used by SN as BGP next hop in its service route advertisements."

RFC4684 and the eSN in question can accomplish IPv4 nexthops via RT-Constrain behavior.

IPv6 route targets in RT-Constrain is work IDR has adopted, but hasn't advanced - and the existing proposal is a matter of some controversy.

The citation of draft-zzhang-idr-bgp-rt-constrains-extension similarly bends the informative vs. normative discussion here.

Issue, Section 12.1:
The topology ASCII diagram appears to be broken.

Sections 12 and 13 have good information in in them, but may be better suited for the Appendices since they show configuration rather than describe protocol procedure.

Issue, Section 14.2:
It's reasonable to want to describe the reservation for the non-transitive extended community to catch its code point. Please make a point that this code point is not currently used in this draft and that operations against it are reserved for future documents.

This also impacts the name, the "Transport Class". What you really have are the "Transitive Transport Class" and the "non-Transitive Transport Class". The document thus far refers universally to the transitive one. You can simplify your life by labeling them "Transport Class" and "Non-Transitive Transport Class" to make them fully clear.

The IANA Name fields likely should receive a similar update.

Similarly, this impacts the "Route Target" that is named in both of the new registries.

I'm calling out this point because of the large extended community reorg work done in the past several years. We don't want to repeat the same headache later for these registrations.

Section 14.4:
If you're going to call out the reserved value 0 as having a name, that's good. However, note that the remaining values are left for arbitrary use by the operator.

While it seems dumb to have to point this out, we've had people try to reallocate such things after the fact for protocol purposes. See large bgp communities for example.

Section 15:
Congratulations on a very nice security considerations section. The security reviewers will assuredly have something negative to say, but that's how they work.

Appendix A.1.1 again makes normative comments based on the multinexthop feature.

Appendix C.1:
The references to the performance results may not be sufficiently stable for publication as an RFC.

Appendix D:
Mpls-namespaces perhaps introduce the same normative/informative issues that multinexthop does. It may be more appropriate to move the discussion of the interaction with CT to an appendix of that document.

Appendix E.2, similar consideration for CPR for informative/normative references. Perhaps one different is that CPR is an operational recommendation that doesn't otherwise (at the time of its last revision) alter protocol procedure.

Jeff's High level issues:

- A high level concern (repeated in my stream of review comments below) is that features such as multinexthop, mpls namespaces, rt-constrain extensions and cpr are used in a normative sense in places in the document. While the intent seems to be more in the lines of "if this feature was deployed, here's how it'd behave", that pushes it more towards normative than informative in many of the cases. My expectation is that during IESG review the same point will be made. I suspect the best answer is to split those cases out of the core CT document, or target doing so for final publication. (Note that we're in the midst of WGLC, so that's probably NOW.)

Jeff's Editorial Issues:

- Some terms get used without definition. Service Node/Border node, for example.
- Overuse of commas in many places.
- Parenthetical remarks made that are part of the main sentence and shouldn't be parenthetical.
- Small things like the latin needs commas after its use. "i.e.," rather than "i.e." Removing "For" prior to "e.g.".
- inter-as option A/B/C are used without reference.

CT-WGLC-RTG-DIR Review: Open Issues from RTG-DIR review (Boucadiar Review)

Github link: https://github.com/ietf-wg-idr/draft-ietf-idr-bgp-ct/issues/25

Text in github: https://github.com/ietf-wg-idr/draft-ietf-idr-bgp-ct/blob/0cda9b95957203599f9418b0781d5575eab75557/draft-ietf-idr-bgp-ct.txt#L925C1-L926C27

This is not aligned with this part:

Reserved: A 2-octet reserved bits.
That MUST be set to zero on transmission.
This field SHOULD be ignored on reception and left unaltered.

@boucadair, please let us know if we can retain the SHOULD/MUST clause for "Reserved" field in TC Route Target Extended Community the same based on the above clarification from @jhaas-pfrc?

Issue: Resolve zero setting in section 6 for Reserved bits.

CT-WGLC-OPS-DIR Review

Link: https://mailarchive.ietf.org/arch/msg/idr/ny5WOEcJwsqpI4YqUExBECKXfMQ/

Github: https://github.com/ietf-wg-idr/draft-ietf-idr-bgp-ct/issues/31

Author: Bo Wu

Status: Not Ready
Version: draft-ietf-idr-bgp-ct-12
Bo Wu:

Thanks for the draft. I am the assigned Ops reviewer to conduct an "early" review of this draft.

This draft proposes a new BGP address family ((AFI/SAFIs, 1/76, 2/76)) and new Transport Class Route Target extended community, that enables the advertisement of underlay routes with identified classes. And the new CT address family
follows the RFC 8277 mechanism to allocate MPLS labels to the underlay routes.

Here are my comments:

1. It is suggested to define a section on Operations and Manageability Considerations to facilitate understanding on this aspect. 1） Management Consideration as a sub-section: It seems that sections 7.9 to 7.10 explains the management of name space. E.g., the management of RT, RD, and Transport Class name space in sections 7.9 and 7.10 are described. And Section 13 Deployment Consideration seems also mentions the transport class/color namespace. 2） Interoperability Consideration as a sub-section: 7.11 and 7.12 explains interoperability with the existing BGP protocol, which does not seem to be the focus of the BGP CT protocol procedure 3） Sub-section that applies: Section 8. Scaling Considerations, Section 9. OAM Considerations, Section 13. Deployment Considerations.

2. Section 10 Applicability to Network Slicing
It is recommended to clarify which specific section of TEAS-NS is related to Transport Class? In addition, it is recommended that the term be used with the TEAS-NS, for example, a TSC is defined as a Network Slice Controller in the TEAS-NS.

3. Section 11 SRv6 Support
Section 13.4. Applicability to IPv6 also mentions SRv6. It is recommended that these two sections be combined. Given that SRv6 is defined separately, should Sections 3 and 12 also indicate that it applies only to MPLS?

4. “Community” usage confusion
Overall impression. Transport Class Route Target extended community is introduced in this draft, but mapping community and color community are also mentioned in many places. It is better to explain why color community is not reused? Also to clarify that Transport Class Route Target extended community not limited to BGP CT family?

Thanks,
Bo Wu

Response from Kaliraj to OPS-DIR:

Link: https://mailarchive.ietf.org/arch/msg/idr/_v1TrCzsnihuUxxdL2lub4KiZcI/

Hi Bo Wu,

Apologies for the delay in response.

We have gone thru your comments, and reorganized few of the sections. Please take a look at draft version 14.

https://www.ietf.org/archive/id/draft-ietf-idr-bgp-ct-14.html

Further please find some responses inline. KV>

In below response,
secX.Y_v12 means: section X.Y in draft version 12
secX.Y_v14 means: section X.Y in draft version 14

Thanks
Kaliraj

Hi,

Thanks for the draft.

I am the assigned Ops reviewer to conduct an "early" review of this draft.

This draft proposes a new BGP address family ((AFI/SAFIs, 1/76, 2/76)) and new Transport Class Route Target extended community, that enables the advertisement of underlay routes with identified classes. And the new CT address family follows the RFC 8277 mechanism to allocate MPLS labels to the underlay routes.

Here are my comments:

1. It is suggested to define a section on Operations and Manageability Considerations to facilitate understanding on this aspect. 1） Management Consideration as a sub-section:

> **[Kaliraj-R1]:** We created section sec10_v14<https://www.ietf.org/archive/id/draft-ietf-idr-bgp-ct-14.html#section-10> as suggested and collected the subsections relating to RD management, label allocation mode etc. under it.

It seems that sections 7.9 to 7.10 explains the management of name space. E.g., the management of RT, RD, and Transport Class name space in sections 7.9 and 7.10 are described.

> **[Kaliraj-R1]** As suggested, moved sec7.9_v12<https://www.ietf.org/archive/id/draft-ietf-idr-bgp-ct-12.html#section-7.9> subsection to sec11_v14<https://www.ietf.org/archive/id/draft-ietf-idr-bgp-ct-14.html#section-11> (Deployment considerations).
> **[Kaliraj-R1]** sec7.10_v12<https://www.ietf.org/archive/id/draft-ietf-idr-bgp-ct-12.html#section-7.10> actually talks about how best-effort routes can be carried in BGP CT family. As part of that, it talks about color management of the best-effort transport-class. So I think it is part of core procedures.

And Section 13 Deployment
Consideration seems also mentions the transport class/color namespace.

> **[Kaliraj-R1]** I thought about whether "Operation and Management considerations" and "Deployment consideration" sections need to be merged. It appears better to keep them separate. Sec 10 and 11 in draft-v14. Please take a look what you think.

2) Interoperability Consideration as a sub-section: 7.11 and 7.12 explains interoperability with the existing BGP protocol, which does not seem to be the focus of the BGP CT protocol procedure

> **[Kaliraj-R1]** I think, sec7.11_v12<https://www.ietf.org/archive/id/draft-ietf-idr-bgp-ct-12.html#section-7.11> is part of core procedures, because it describes how to unambiguously identify the effective-color of a nexthop, when color can exist at various places on a BGP route. Sec7.12_v12<https://www.ietf.org/archive/id/draft-ietf-idr-bgp-ct-12.html#section-7.12> is also important to state here because it is a new procedure of how flowspec routes can redirect traffic. So left these in core procedures.

3） Sub-section that applies:
Section 8. Scaling Considerations, Section 9. OAM Considerations, Section 13. Deployment Considerations.

> **[Kalira-R1j]** OAM is moved into sec10_v14<https://www.ietf.org/archive/id/draft-ietf-idr-bgp-ct-14.html#section-10> as suggested. But left 'Scaling consisderations' as separate section, since I think that's different from Operations.

2. Section 10 Applicability to Network Slicing
It is recommended to clarify which specific section of TEAS-NS is related to Transport Class? In addition, it is recommended that the term be used with the TEAS-NS, for example, a TSC is defined as a Network Slice Controller in the TEAS-NS.

> **[Kaliraj-R1]** the hyperlink pointed to sec 4 in TEAS-NS. Clarified text to mention that. Pls see sec12_v14<https://www.ietf.org/archive/id/draft-ietf-idr-bgp-ct-14.html#section-12> . Also, changed TSC to NSC as suggested. Thanks for pointing this out.

3. Section 11 SRv6 Support
Section 13.4. Applicability to IPv6 also mentions SRv6. It is recommended that these two sections be combined. Given that SRv6 is defined separately, should Sections 3 and 12 also indicate that it applies only to MPLS?

> **[Kaliraj-R1]** IPv6 section is different from SRv6 section. This was added based on pre-review comments from rtg-dir. This section describes how BGP CT procedures work in a network with IPv6 control plane and IPv6 data traffic. To avoid confusion, I removed reference of SRv6 from sec7.12_v14<https://www.ietf.org/archive/id/draft-ietf-idr-bgp-ct-14.html#section-7.12> and confined it to MPLS-forwarding, since sec7.13_v14<https://www.ietf.org/archive/id/draft-ietf-idr-bgp-ct-14.html#section-7.13> talks about SRv6-forwarding. Also, moved these two sections to protocol procedures, since they specify how core-procedures work for these scenarios.

4. "Community" usage confusion
Overall impression. Transport Class Route Target extended community is introduced in this draft, but mapping community and color community are also mentioned in many places. It is better to explain why color community is not reused? Also to clarify that Transport Class Route Target extended community

not limited to BGP CT family?

> [Kaliraj-R1] Color community is indeed re-used. Mapping Community is just a "role", not a new type of community. Both Color-community and Transport-RT community play the role of a mapping-community. Any BGP community can play this role actually. Clarified this in sec sec5.1_v14<https://www.ietf.org/archive/id/draft-ietf-idr-bgp-ct-14.html#section-5.1>.
>
> [Kalirja-R1] Please let me know if a call would help to discuss any remaining concerns. Thanks.

Shepherd's action items:

- Confirm resolution of OPS-DIR review in the following 4 sections:
  - Manageability
  - Section 10
  - Section 11
- Confirm that SRv6 issues and community confusion has been resolve.

WG Adoption issues – Reviews

CT Adoption Call issues

Open adoption call issues:

- F3-CT-Issue-5, F3-CT-Issue-6,
- F3-WG-Issue-3, F3-WG-Issue-4, F3-WG-Issue-6

Closed adoption issues:

- F3-CT-Issue-1, F3-CT-Issue-2, F3-CT-Issue-3, F3-CT-Issue-4,
- F3-WG-Issue-1, F3-WG-Issue-2, F3-WG-Issue-5, F3-WG-Issue-7, F3-WG-Issue-8

F3-CT-Issue1: SAFI 76 only in Option C

Person: Robert Raszuk
Link: (TBD)
WG LC Response status: none
Response to post-WG LC query: TBD
Github: https://github.com/ietf-wg-idr/draft-ietf-idr-bgp-ct/issues/1
Status: closed

F3-CT-Issue-2: Sizing

Person: Ketan Talaulikar (ketan.ietf@gmail.com)
Response to WG LC: https://mailarchive.ietf.org/arch/msg/idr/q1PBTKAnlsuyjYwd4fwpAgCEssI/
Github: https://github.com/ietf-wg-idr/draft-ietf-idr-bgp-ct/issues/2
Status: closed, additional comments in WG LC comment from Ketan

F3-CT-Issue-3: BGP and RTC

Person: Jeffrey Zhang
Pre-WG LC Response: Ok
Github Link: https://github.com/ietf-wg-idr/draft-ietf-idr-bgp-ct/issues/3
Status: closed

F3-CT-Issue-4: Clarification by Bruno on BGP-CT NLRI

Person: Bruno Decreane
Status: No desire to review, status closed
Link: https://mailarchive.ietf.org/arch/msg/idr/YfGESa7kbByBAAKWI9TR_ukFVoQ/
Github link: https://github.com/ietf-wg-idr/draft-ietf-idr-bgp-ct/issues/4

F3-CT-Issue-5: Perpetuating Know issues in MPLS label filed in SAFI

Person: Ketan Talaulikar (ketan.ietf@gmail.com)
Response to WG LC: https://mailarchive.ietf.org/arch/msg/idr/q1PBTKAnlsuyjYwd4fwpAgCEssI/

Status: https://github.com/ietf-wg-idr/draft-ietf-idr-bgp-ct/issues/5

F3-CT-Issue-6: Unique RD usage and caveats (related to F3-CT-Issue-6)

Github issue: https://github.com/ietf-wg-idr/draft-ietf-idr-bgp-ct/issues/6
Status: open, pending resolution

Commenter: Swadesh Agrawal (swaagraw@cisco.com)

Links to comments:

- Swadesh: https://mailarchive.ietf.org/arch/msg/idr/IOYt6PiIHafrLjk4W6kqjWTGuSA/
- Kaliraj-R1: https://mailarchive.ietf.org/arch/msg/idr/kRJSPbyxrcbMStqy8C0bgFigVrI/
- Swadesh-R1: https://mailarchive.ietf.org/arch/msg/idr/_YmF94svUk2_VmObob--9WlYtQY/
- Kaliraj-R2: https://mailarchive.ietf.org/arch/msg/idr/WNsd391q4ELPIcXhvvZVWyusd3Y/
- Swadesh-R2: https://mailarchive.ietf.org/arch/msg/idr/B73Vppd8h8fqGyyz_ozNP1b1nmk/
- Kaliraj-R3: https://mailarchive.ietf.org/arch/msg/idr/a3zJ4y7eumYTU9Fc-xUJn9shtLU/

Swadesh: It can be seen from Figure 13 rows 4 and 6, failure of an originator (such as ABR) will result in slow convergence as LSP is end to end and failure of originator needs to be propagated to ingress PE to converge.

> **Kaliraj-R1:** And from rows 1,3,5,7,9,11, that failure is not propagated until ingress-PE. This table is an exhaustive list of all possible combinations.
>
> **[Swadesh-R1]:** CT draft recommends use of unique RD. Hence, for the recommended case (i.e rows 2,4,6,8,10 and 12), convergence is slow as LSP is end to end and failure of originator needs to be propagated to ingress PE to converge. Further as per my understanding, control plane churn of CT routes is not localized for rows 1,3,5,7,9 and 11 as well. Please see [the] next comment with explanation.
>
> **[Kaliraj-R2]: [As]** stated in Figure-13 [https://datatracker.ietf.org/doc/html/draft-ietf-idr-bgp-ct-12#figure-13], CT provides a sliding control that operators can use to control route-visibility vs route-scale at ingress PE. [It] is an exhaustive list of combinations. Unique RDs improve troubleshooting and route visibility, with an increase in ingress route-scale. Duplicate RDs can be used to reduce ingress routes, if required, with limited egress-PEs visibility.
>
> > **[Swadesh-R2]:** [I believe there is a] serious limitation of end-to-end slow convergence due to unique RD as a route visibility and troubleshooting choice. Moreover, the workaround of stripping RD at BRs does not reduce the control plane scale and churn either. In fact it exposes the problem of carrying [the] same forwarding LSP information in redundant BGP routes from a device. There is no logical reason to incur this complexity and overhead.
> >
> > [**Swadesh-R2]:** This is not a “sliding control” but a design problem with the CT NLRI data model. The CT draft should really capture the limitations for each of the rows of figure 13 as discussed in this thread, to allow operators to make the appropriate choices.

To avoid it "RD stripping" or “TC,EP” label allocation procedures at BNs is stated as an option in section 7.4. But even with that option, the control plane churn is still not kept within local domain as CT control plane is signaling redundant routes that carries same label.

> **Kaliraj-R1:** About the “churn not kept local” claim, not true. The advertised CT label does not change when local failure events happen and nexthop changes from one nexthop to another. Because of “TC, EP” label allocation mode. So Churn is not propagated further than local BN. IOW, no new updates are sent and ingress-PE does not see this failure event.
>
> **[Swadesh-R2]** Here is my understanding of CT procedures. Lets take example of figure topology 12 and figure 13 row 5 case. 4 PEs (PE11-PE14) have RDs PE11 to PE14 respectively as its unique RD case.
>
> - Anycast address is 1.1.1.1 across 4 PEs.
> - CT routes (RD:IP) learnt on ASBR11 from originator PEs are PE11:1.1.1.1, PE12:1.1.1.1, PE13:1.1.1.1 and PE14:1.1.1.1 with same TC GOLD.
> - ASBR 11 allocate label 16001 for (GOLD, 1.1.1.1) and advertise 4 routes PE11:1.1.1.1, PE12:1.1.1.1, PE13:1.1.1.1 and PE14:1.1.1.1 with label 16001 to PE31.
> - Issue 1: Above 4 routes carry exactly same label 16001 to PE31. This unnecessary control plane scale with same forwarding information.
> - Issue 2: Now if PE11 goes down, ASBR11 need to withdraw (PE11:1.1.1.1) CT route from ingress PE31. So local domain control plane churn is being propagated to PE31.
>
> [On procedures and Issue-1]
> **[Kaliraj-R2]:** That’s correct understanding of the behavior. but not an issue per-se. It provides visibility into how many egress-PEs are currently serving the Anycast-service. Some of our customers see that as a good feature.
>
> > **[Swadesh-R2]:** This is not a feature but a design limitation with the CT NLRI. It results in unnecessary redundant BGP routes that increases end to end scale and churn without any functionality. This need to be fixed or called out clearly in the draft.
>
> **[Kaliraj-R2**]: This is also regular BGP behavior. Not an issue. “Egress-PE down” case will be sent as BGP withdrawal to ingress-PE in regular option-C (LU), except if you use MPLS-namespaces (https://datatracker.ietf.org/doc/html/draft-ietf-idr-bgp-ct-12#name-context-protocol-nexthop-ad). I was talking about the cases where the churn is kept local for events like link-down or nexthop/label-changes. Because of per(EP, TC) label allocation mode, a new label is not advertised.
>
> > **[Swadesh-R2]:** Discussion is for row 5 of figure 13 where redundant routers originate the given endpoint. In such failures, BGP withdrawals need not be sent end to end but can be contained within the local domain because path is still available from redundant originator. The CT handling above is not regular BGP behavior

Row 5 shows for 16 routes there are only 2 labels advertised. Multiple redundant routes are advertised with same forwarding information and increases control plane state. This was the issue raised as a problem created by RD in NLRI. The impact aggravates as number of anycast originators increases.

> **[Kaliraj-R1]:** Please pay attention to rows 7-12 also. Which has only 2 or 4 routes advertised, with 2 unique labels. Operators have the flexibility to choose the desired visibility at ingress-PE,

with the desired scaling characteristics. This table is an exhaustive list of all combinations. That helps operators to choose which mode fits their needs the best.

[Swadesh-R1] I did pay attention to 7-12 rows as well. Rows 7,8 and 11,12 are for the same RD case. This case defeats the draft's stated purpose of using RD in NLRI. Rows 6 and 10 suffer from end to end slow convergence. Row 5 exposes the redundant route problem (16 routes for 2 forwarding state to ingress PE) and aggravates with increase in number of SNs. Same is with row 9 and aggravates as number of BNs increases. (TC,EP) allocation scheme is not containing control plane churn within the domain as claimed in section 7.4; as I stated in my previous email.

[Kaliraj-R2]: Same as above (see previous Kaliraj-R2 comment)

[Swadesh-R2]: Yes. And as stated in previous comment it's an issue and needs to captured in draft for each row of figure 13.

The updates to the draft do not address the raised issue. However, it states (in sec 7.4) that route churn is avoided, and is proportional to number of labels but that is not the case as explained above.

As a related observation, there was a solution for above issue proposed by authors on the list to use local RD of BN node when "Stripping RD". However, it looks like that solution has been discarded as its not discussed in the draft.

BGP-CT-UPDATE-PACKING-TEST results included are for an unrealistic scenario in practice; and also do not cover relevant deployment cases :

For example it captures 1.9 million BGP CT MPLS routes packed in 7851 update messages. That means about 250 routes sharing attributes and packing every update message completely. It seems test is done with all routes (around 400k) for a given color having exactly same attributes. This is not a practical example. A more practical case would be to have a packing ratio, for example 5-6 routes to a set of attributes.

Kaliraj-R1: The goal of the experiment is to see the impact of carrying 'Color as an Attribute' as against 'Color in NLRI'. The issue raised was that, carrying color as attribute will affect packing. So this experiment demonstrates that the observed convergence time is in accepted limits, even when color is carried as an attribute. In any controlled experiment, we want to vary one variable to observe the result.

[Swadesh-R-1] I am not sure of such [a] discussion. The observed issue was for label index and SRv6 SID that [is] per-prefix information with RFC 8277 style encoding that carries such information in attribute and breaks BGP update packing. In any case, it will be helpful to have such [an] analysis of BGP CT for the WG.

More importantly, the test results do not include or analyze impact of label index, SRv6 SID etc. that are per-prefix information.

Kaliraj-R1: This experiment provides actual benchmarking test results for one case (color as an attribute), that can be extrapolated for other cases as well where SID(label-index/SRv6) is carried as an attribute, just like the Color.

[Swadesh-R1]: Just to reiterate, Color was not the discussion point for update packing. With just 5 colors across 1.9 millions routes, nobody sees update packing as a concern.

[Kaliraj-R2]: OK. Btw, these scale requirements are from https://datatracker.ietf.org/doc/html/draft-hr-spring-intentaware-routing-using-color-01#section-6.3.2.

[Swadesh-R2]: It is not a good choice for a new BGP SAFI to be designed such that every prefix needs to be carried in separate update message when using label index(SR MPLS) and SRv6. It increases BGP control plane data size by multiple fold. Problems will be seen in scaled networks.

[Swadesh-R2]: Colors carried in attribute was never a problem from update packing point of view and not an issue called out in adoption call. The issue was raised for label-index and SRv6 SID that are per prefix information. Current results provided is of no practical use.

[Swadesh-R1] The concern arises for label index and SRv6 SID that is per prefix information carried in attribute in BGP CT encoding. This breaks packing and has an impact.

[Kaliraj-R2]: This experiment shows the numbers with color carried as attribute. Results will be comparable when label-index/SRv6-SID/TEA are carried as attribute as-well. IMHO, it may not be worth carrying all those in the NLRI, in an attempt to micro-optimize this further.

[Swadesh] Non deterministic usage of IMPLICIT NULL:

- Implicit NULL is a valid MPLS label and indicates no label to push by receiver. Label path to BGP nexthop is still valid/expected.

  Kaliraj-R1: Intra-domain tunneled path to the BGP nexthop may or may-not be labeled. Implicit-Null label carried in BGP-LU/BGP-CT route doesn't claim anything about the intra-domain tunnel. It just says no BGP-LU/BGP-CT label needs to be pushed in forwarding.

  [Swadesh-R1]: Thanks for clarification on procedure. But when I read draft, it indicates towards new meaning of IMPLICT NULL. Quoting exact text in draft "R4 will carry the special MPLS Label with value 3 (Implicit-NULL) in RFC 8277 encoding, which tells R1 not to push any MPLS label towards R4". It will be better to update your response text in the draft.
  Section 13.2.2.1 is extending implicit NULL label presence to indicate that originator does not support MPLS. This is not possible as the two cases cannot be distinguished.

  [Kaliraj-R2]: Sure, will clarify the text to say, "Implicit-Null label carried in BGP-LU/BGP-CT route indicates that no BGP-LU/BGP-CT label is pushed in forwarding".

  [Swadesh-R2]: Thanks. But mis delivery of traffic is possible if an MPLS tunnel exists to next hop with this procedure. This should be captured in the draft.

Kaliraj-R-1: Section 13.2.2.1 is extending [the] implicit NULL label presence to indicate that originator does not support MPLS. This is not possible as the two cases cannot be distinguished. so, there is no ambiguity. Implicit-NULL is only saying no BGP-LU/BGP-CT label needs to be pushed in forwarding.

[Swadesh-R-1]: Same Response as the previous point.

- Section 13.2.2.1 is extending implicit NULL label presence to indicate that originator does not support MPLS. This is not possible as the two cases cannot be distinguished. For example in figure 11 and 12 not sure why R3 won't send MPLS traffic to R4 as stated in last paragraph. Similar is the problem with section 13.2.2.2.

[Kaliraj-R-1] as shown in these figures, R4 does not support MPLS. So there can be no MPLS-tunnel from R3->R4. [So] why would R3 send MPLS traffic to R4? When R3 tries to resolve PNH==R4, it will find no matching MPLS tunnel, and the route will remain Unusable.

[Swadesh-R-1]: It's an operational burden to make sure that no router has MPLS path to R4 (MPLS path can be for other purposes). Otherwise there can be mis-forwarding with IMPLICIT-NULL in 8277 style encoding for non MPLS encapsulation signaling (SRv6, UDP) in BGP CT. It should be captured in the draft.

[**Kaliraj-R2]:** R4 does not support MPLS. So there can be no MPLS path towards it. There is no operational burden. Thanks for the comments.

[Swadesh-R2] Previous point response applies here as well.

[Swadesh-R2]: (at top)

There were 3 issues called-out; neither of which have been addressed.

1. The first one is a significant design limitation that the use of unique RDs results in:
    a. Lack of multipath and localized fast convergence within a domain for originator failures (even though an operator has deployed redundant routers)
    b. To achieve multipath, 'stripping RD' and 'TC, IP' allocation results advertisement of duplicate/redundant BGP routes with the same forwarding label
    c. which in turn increases control plane state on all routers upstream across multiple domains, and exposes the failure churn outside the originating domain all the way to ingress PEs in other domains

    This is not a new issue, it has existed since day-1 of CT and still remains, in spite of all the revisions of the draft.

    This is not just an editorial issue. It is a significant deviation from [the] currently deployed BGP-LU, which does not have these duplicate route/churn issues and provides multipath/active-backup within each local domain. It is a manifestation of the wrong data model of signaling RD in NLRI for BGP hop-by-hop routes.

    The limitations need to be captured clearly in the draft as impact of the respective options, if they are not going to be addressed.

[Kaliraj-R3]: BGP CT allows operators full flexibility of achieving what a deployment needs.

2. The second issue is that of the inefficiency caused by the choice of the CT NLRI which only supports MPLS labels in the NLRI. Any use other than MPLS, such as SR prefix-SID (label-index) or SRv6 SID means every route needs to be sent in a separate BGP update message with no packing possible. The scale/performance test data completely ignores this issue and shows data for a non-existent problem.

   [Kaliraj-R3]: The update packing test results show numbers for a scenario where "Update packing does not happen". It serves good for any reason why Update packing may not happen (e.g. dissimilar aigp-attributes, bgp communities, loc-pref, SIDs, Color carrying communities). The test result can be extrapolated to the other cases too, because all of them break update packing. Everything cannot be carried in NLRI to micro-optimize update packing.

3. The 3rd issue is that the draft introduces non-deterministic usage of IMPLICT NULL. It can result in mis-delivery of traffic and is an operational burden to make sure no MPLS path exists to next hop. This is again a result of CT mandating signaling label in NLRI even for non-MPLS encapsulation.

   [Kaliraj-R3]: In Fig 10 (https://www.ietf.org/archive/id/draft-ietf-idr-bgp-ct-12.html#figure-10), if you have an ingress-node R1 that assumes to have successfully signaled a MPLS transport tunnel to device R4 that does not support MPLS forwarding at all, it is a bug in R1 outside the scope of BGP CT.
   - [For example] e.g., if that ingress node R1 receives a AFI/SAFI:1/1 route with R4 as next hop, that will also result in mpls pkt 'attempted' to be sent towards R4.
   - That is not a problem in AFI/SAFI:1,1 procedures. Just a bug in R1 implementation.

   [Kaliraj-R3]: Like stated already, Implicit-Null in a BGP route only says no MPLS-label need to be pushed "at that BGP-route's layer". It does not make any assumptions about the transport-tunnel that the route resolves over.

Issues raised:

- Clarity in design choices in the usage of unique RDs
- Clarity in design choices in using RFC8277 TLV encodings versus other types of TLV encodings for MPLS
- Clarity in implicit-NULL usage (section 13.2.2.1 and section 13.2.2.2)
- Clarity in scope of test parameters and scope

Adoption Call WG issues

F3-WG-Issue-1: New Address Families [Shunwan Zhuang]

- IDR mail thread: https://mailarchive.ietf.org/arch/msg/idr/4T3-b4_ckpGu3BwjwuESqpYsoFk/
- Query for closure: TBD
- Github link: https://github.com/ietf-wg-idr/draft-ietf-idr-bgp-ct/issues/7
- Status: closed

F3-WG-Issue-2: Support for SR-v6 (Jingrong Xie) (xiejingron@huawei.com)

[Author] Jingrong Xie **[xiejingrong@huawei.com]**
IDR mail link: https://mailarchive.ietf.org/arch/msg/idr/7C7dlvIgZuNNx3rLorC6S24Ta50/
IDR mail link: https://mailarchive.ietf.org/arch/msg/idr/7C7dlvIgZuNNx3rLorC6S24Ta50/
CT: Provide an illustration of SRv6 data plane (e.g., E2E SRv6 & intra-domain SRv6) based on a sample topology?:

- Status: closed.
- Github link: https://github.com/ietf-wg-idr/draft-ietf-idr-bgp-ct/issues/8

F3-WG-Issue-3: Key Operational Differences between CAR and CT drafts (Bruno Decraene)

People: Bruno Decraene and Jeff Haas
IDR mail thread: https://mailarchive.ietf.org/arch/msg/idr/-N9CncTl8JtwDLmGZEJ1RLqzMSM/

- Shepherd's review of -02: (TBD)
- WG LC review of -02:
  - Bruno Decraene: Did not want to comment
  - Jeff Haas questions: See CT review
    - [Link mail list: https://mailarchive.ietf.org/arch/msg/idr/vxnJDv5gUnExRctD1tYuy8Zy0Do/]
- Github link: https://github.com/ietf-wg-idr/draft-ietf-idr-bgp-ct/issues/9
- Status: merge into CT-WGLC-Q2: Jeff Haas

F3-WG-Issue-4: Intent at Service level [Ketan Talaulikar]

Originator: Ketan Talaulikar:
IDR thread link: https://mailarchive.ietf.org/arch/msg/idr/hHto6CYV6zWeTju7gHWLH1qRsOA/

- Ketan Response during WG LC: Link: https://mailarchive.ietf.org/arch/msg/idr/q1PBTKAnlsuyjYwd4fwpAgCEssI/
- Action item:
  - CT: Add a section to discuss how color is implement in the VPN service layer.
  - CT: Add a definition of intent that aligns with Spring and other IETF/IRTF WGs
- Github original issue: https://github.com/ietf-wg-idr/draft-ietf-idr-bgp-ct/issues/10
- Github for WG LC issue: https://github.com/ietf-wg-idr/draft-ietf-idr-bgp-ct/issues/30
- Status: merged into CT-WG LC -01

F3-WG-Issue-5: Technology BGP-CT and CAR are based [upon] and implications [Jeffrey Zhang]

[Jeffrey Zhang]: Whether BGP-CT/CAR are based on VPN or BGP-LU and which one is better to go forward.

[text]: Following section explains the relationship and distinction between SAFI 76 and SAFI 4, SAFI 128.
https://datatracker.ietf.org/doc/html/draft-kaliraj-idr-bgp-classful-transport-planes-17#section-9
Email: https://mailarchive.ietf.org/arch/msg/idr/fLSx_Qh9BZJweQd0AQSNx1q7nOk/

Github original issue: https://github.com/ietf-wg-idr/draft-ietf-idr-bgp-ct/issues/11
Status: closed

- CT: Clarify the interaction with RDs and RTCs by discussing how CT handles RDs, RTCs, labels and other VPN signaling information that sent to domain with CT.
- CT discuss how efficient CT is in domains which do not handle SR-MPLS or VPNs.
- Sections changed: (TBD)
- WG LC query: (TBD)

F3-WG-Issue-6: Benefits of Route Targets [Swadesh Agrawal]

Thread: https://mailarchive.ietf.org/arch/msg/idr/v9f1wKjalFFOBq-3NmtN1Cg_2eQ/
Github: https://github.com/ietf-wg-idr/draft-ietf-idr-bgp-ct/issues/12
Status: open with comments in WG LC

[author]: Swadesh Agrawal
Action items:

- CT: Provide normative text and examples for non-agreeing color domains with examples on how transport class is used. This example should include the example in F3-WG-issue-6 – which includes attaching multiple RTs to be used in different color domains is not practical

Post-WG-LC Resolution:

- Github issue: https://github.com/ietf-wg-idr/draft-ietf-idr-bgp-ct/issues/12

F3-WG-Issue-7: Compatibility of BGP-CT and BGP-CAR to SR-PCE (Shraddha Hegde)

IDR mail thread: https://mailarchive.ietf.org/arch/msg/idr/zWqlGvaL3zS2NqTsDAAk9L0iH-Q/

Issue author: Shraddha Hegde <shraddha@juniper.net> Wed, 27 July 2022

Action items:

- CT: Consider how CT implements or interoperates with all the constructs in RFC 9256 and RFC9252. CT: Provide a short section in your document regarding support.
- CT: Describe the limits of any community, extended community, wide-community regarding color when interacting with CT's mapping community.

Pre-WG LC:

- Github issue: https://github.com/ietf-wg-idr/draft-ietf-idr-bgp-ct/issues/13
- Status: closed

F3-WG-Issue-8: Scaling and Expected Route Size

[Robert Raszuk]: posts the following as follow-on to Jeff message sizes, but it is a different thread.
[IDR message thread: https://mailarchive.ietf.org/arch/msg/idr/v8kkDGmr3ViPIR4UEmOPJbJ8B44/]

- CT: Discuss in scale section how CAR scales to:
  - limits in draft-hr-spring-intentaware-routing-using-color-00
  - Jeff Haas' rough route calculation: 1.5 million routes, given 10K update, about 2.5 minutes of convergence
  - Robert's use case: transient route problems every 5-10 sections every 50 seconds

Pre-WG LC:

- Github issue: https://github.com/ietf-wg-idr/draft-ietf-idr-bgp-ct/issues/14
- Status: closed