SendHoldTimer and you

Ben Cartwright-Cox
NONOG-6 / NIX 2024

HoldTimer

- BGP is TCP
- TCP does not have built in connection timeouts
- BGP peers send "KEEPALIVE" frames typically every 30 seconds between each other to reset a hold timer of (normally) 90 seconds
- You can lower this time if you want, but if the other router gets very busy you might "hold timer expired" when you didn't want to

- You can disable the HoldTimer, but I think you are insane if you do

The problem

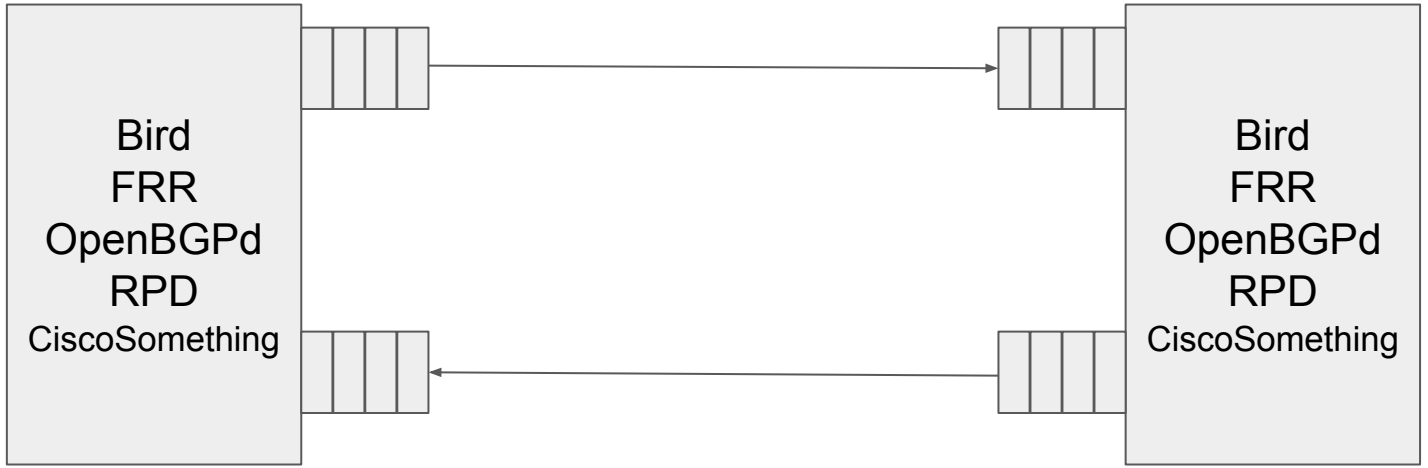
- BGP is TCP
- TCP is an underspecified two-node consensus algorithm

The problem

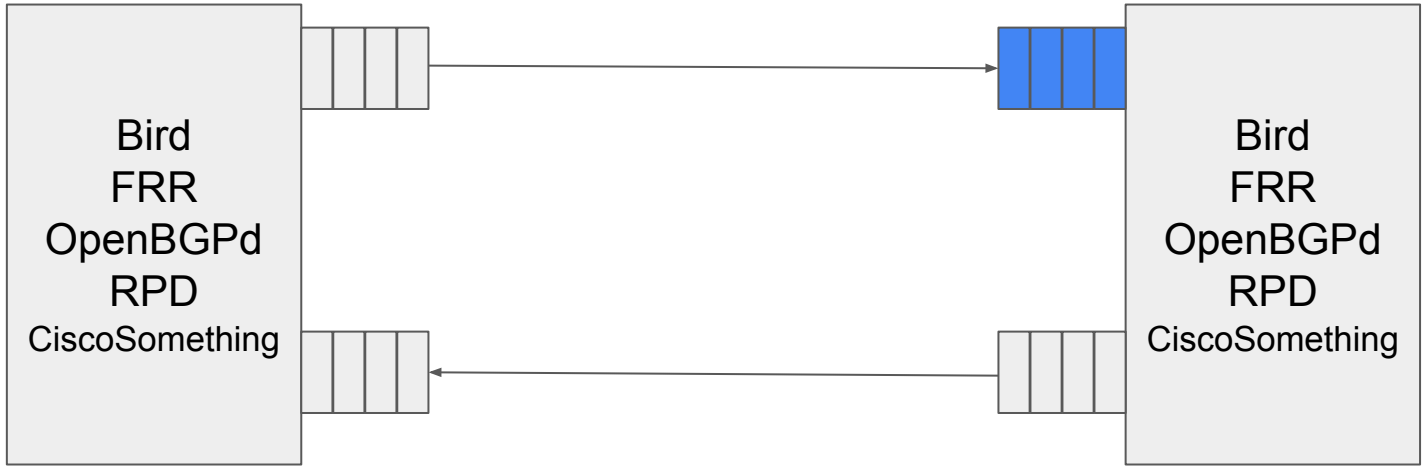
- BGP is TCP
- TCP is an underspecified two-node consensus algorithm

```
root@orange-ns:~# ss -plan
Netid State      Recv-Q Send-Q                                  Local Address:Port
...
tcp   ESTAB      0      50                                   193.251.143.108:1021
...
```

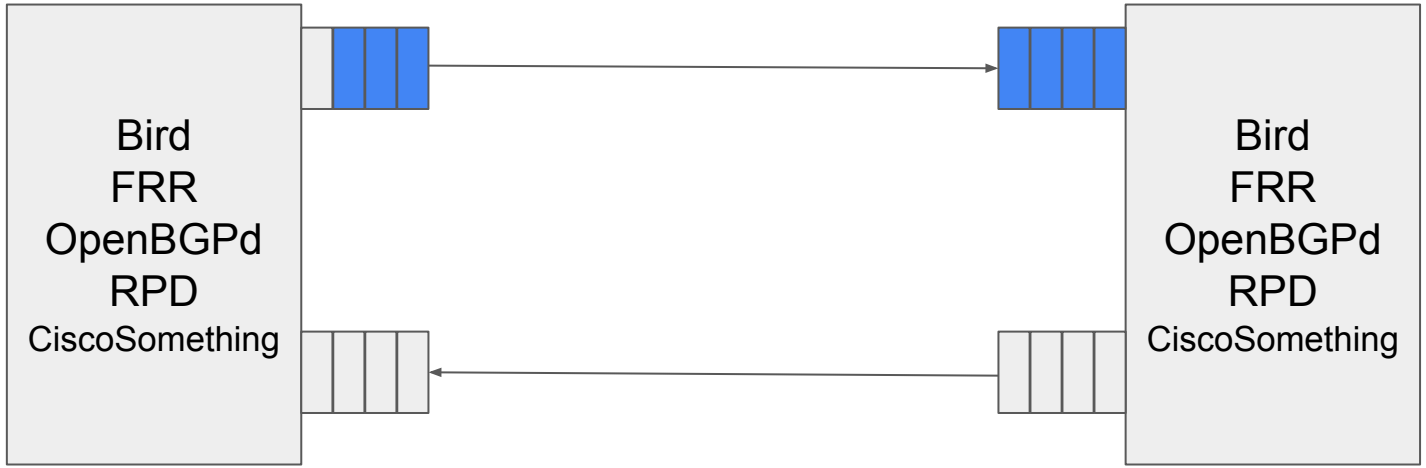
The problem


Bird
FRR
OpenBGPd
RPD
CiscoSomething
Bird
FRR
OpenBGPd
RPD
CiscoSomething

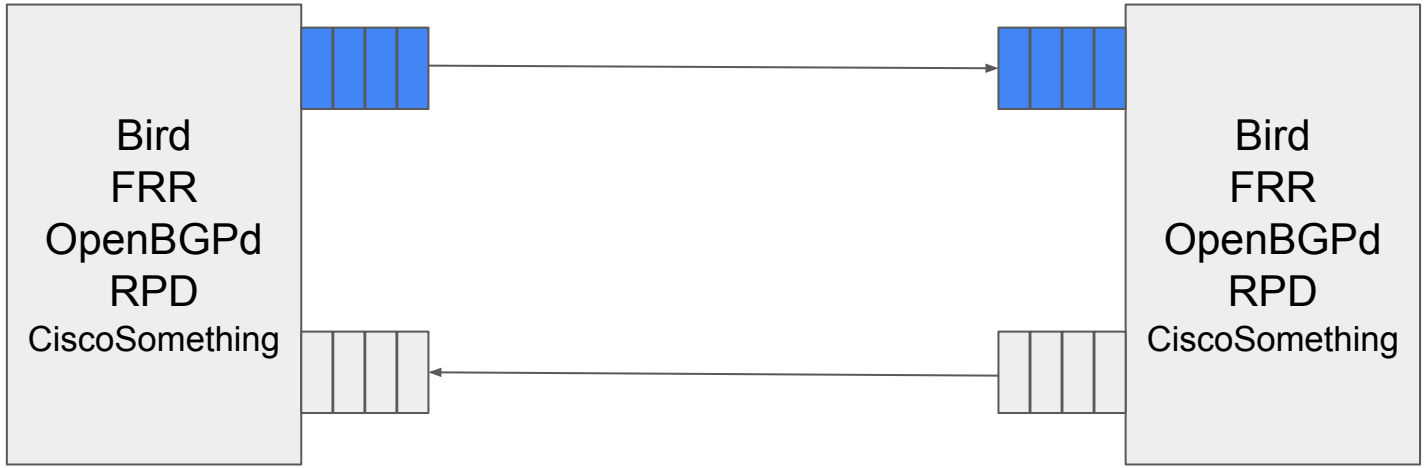
The problem


Bird
FRR
OpenBGPd
RPD
CiscoSomething
Bird
FRR
OpenBGPd
RPD
CiscoSomething

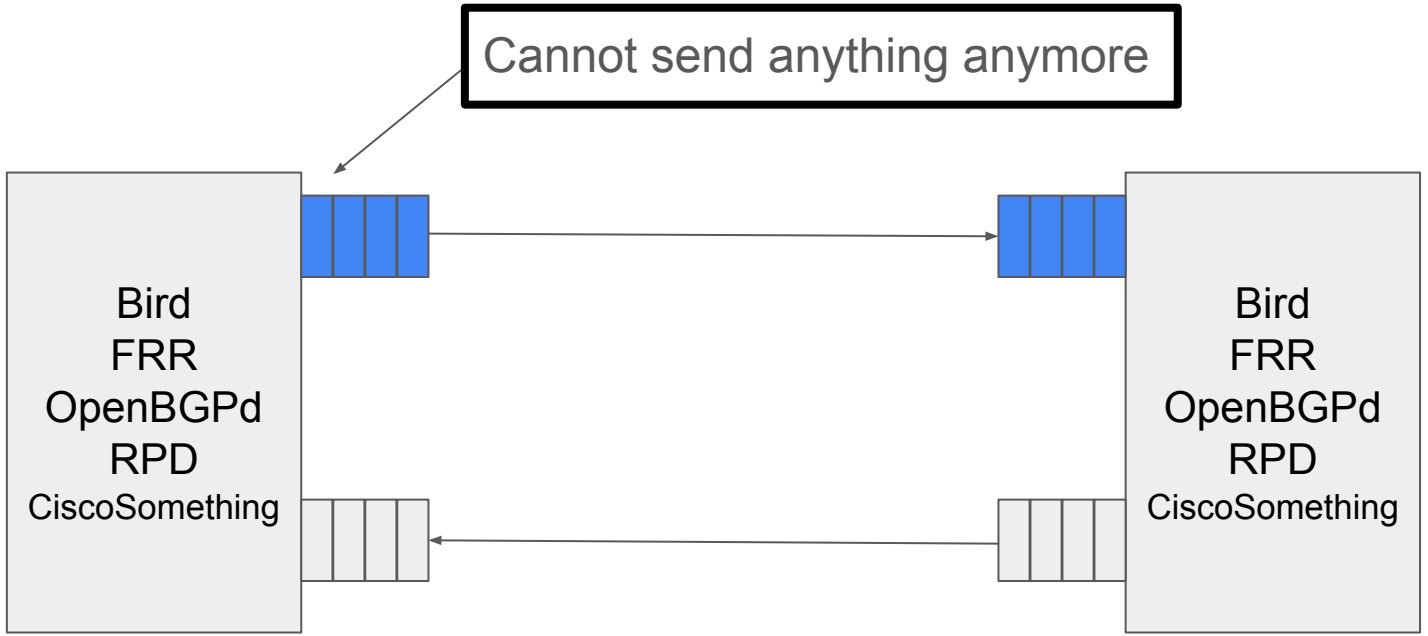
The problem


Bird
FRR
OpenBGPd
RPD
CiscoSomething
Bird
FRR
OpenBGPd
RPD
CiscoSomething

The problem


Bird
FRR
OpenBGPd
RPD
CiscoSomething
Bird
FRR
OpenBGPd
RPD
CiscoSomething

The problem


Cannot send anything anymore
Bird
FRR
OpenBGPd
RPD
CiscoSomething
Bird
FRR
OpenBGPd
RPD
CiscoSomething

The problem

- Being unable to send is actually quite normal
- When you are sending a full table to a customer, there is a high chance your bgpd is in this situation that it wants to send more, but it cannot
- So you cannot just disconnect all peers that end up in this situation otherwise BGP stops working as a whole

The problem

- Being unable to send is actually quite normal
- When you are sending a full table to a customer, there is a high chance your bgpd is in this situation that it wants to send more, but it cannot
- So you cannot just disconnect all peers that end up in this situation otherwise BGP stops working as a whole
- But if you just hope that the socket will soon become unblocked. You may be stuck forever
- Typically the HoldTimer catches this because the other peer has died and wont send KEEPALIVEs anymore

The problem

- Sometimes BGPd's can stop reading their socket, but keep sending KEEPALIVEs… Forever.
- Nuclear worst case, You cannot at this point send new/withdraw route updates and the session remains up forever
- But because the hold timer on your end is refreshed. Everything keeps working

The Symptoms

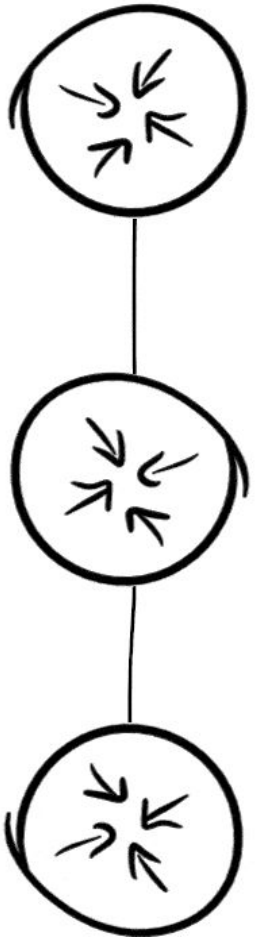
The Symptoms

- Tables go "stale" with out of date routing data
- This impacts downstreams as well

The Symptoms

- Tables go "stale" with out of date routing data
- This impacts downstreams as well

- This surfaces often as "stuck routes"



Why you should care about this?

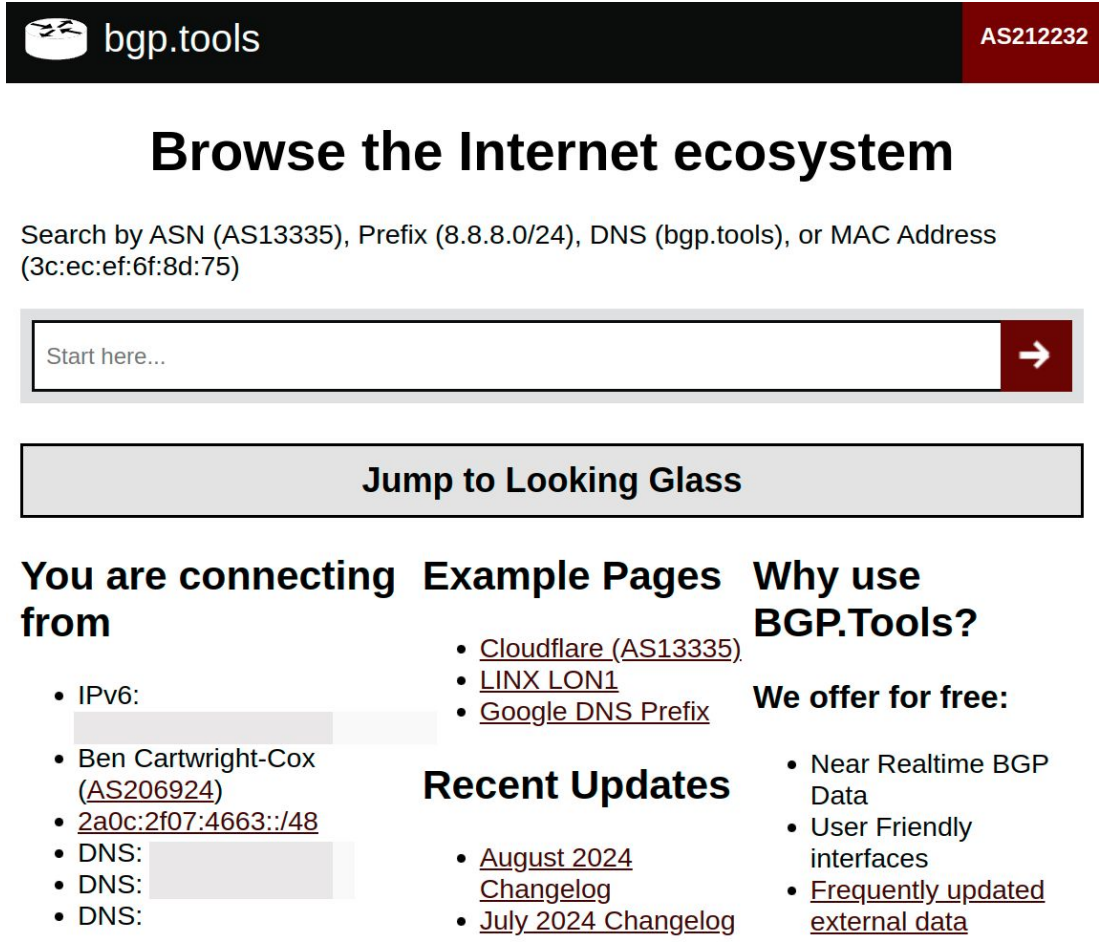
- Providers would generally prefer that the AS-PATH is accurate, and that networks don't send route for prefixes they cannot get to

- Your own prefixes may get "stuck" in providers backbones and be either blackhole'd, or send down stupid paths that you cannot easily fix without contacting the offending network
  - Who may be a ex-supplier, possibly extra spicy

Why do I care about this?

- I run bgp.tools
- bgp.tools is at the core, a BGP route collector
- "Stuck" BGP routing data not only makes the site worse, but also messes with my own monitoring and annoys my customers

- I am incentivised to fix this for my own sanity



The AS204318 BGP Clock

- I put a /34 of IPv6 to work, and got a AS204318 to announce two prefixes
- First, the whole /34
  - Otherwise other networks IRR filter generation does not work in time
- Second, a /48 the represents the current hour of the year, in hexadecimal in the IPv6 prefix
- Right now, there are 13 "visible" prefixes according to bgp.tools for AS204318
  - Remember there should only be 2

Benjamin Cartwright-Cox

AS Number **204318**

Overview | Prefixes | Connectivity | Whois

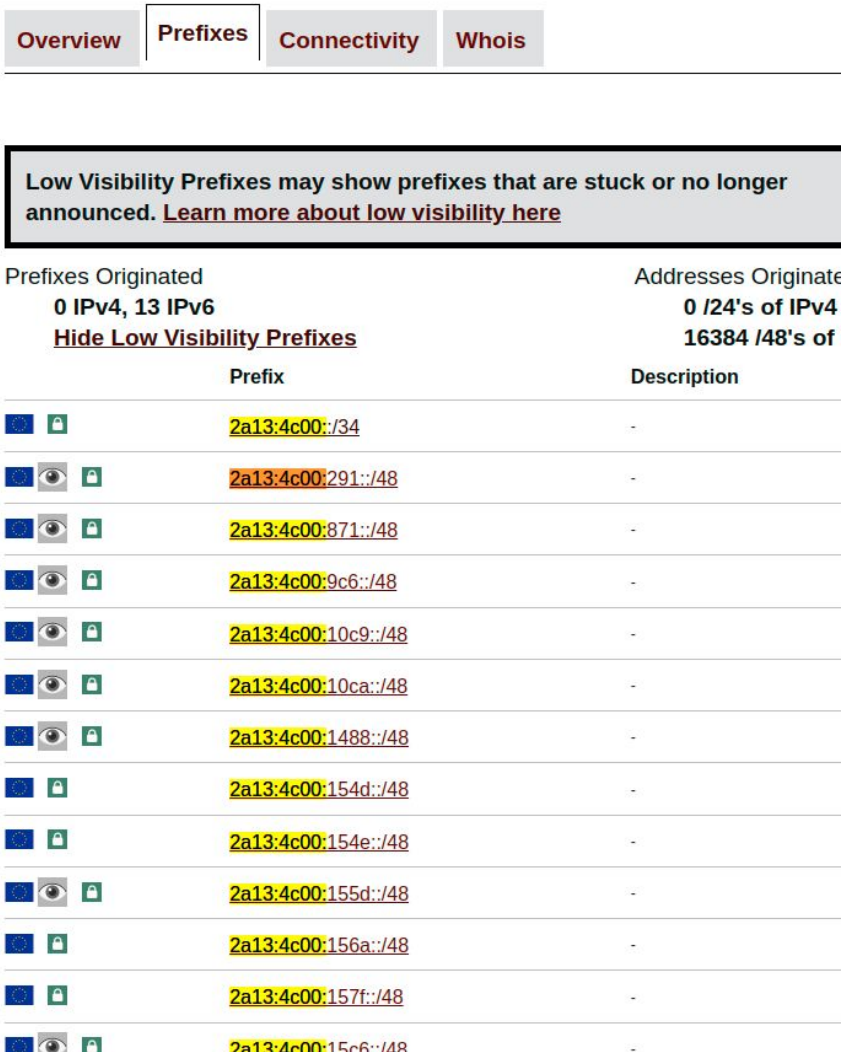
Prefixes Originated
0 IPv4, 13 IPv6
Show Low Visibility Prefixes

Addresses Originated
0 /24's of IPv4
16384 /48's of IPv6

Prefix	Description
2a13:4c00::/34	-
2a13:4c00:154d::/48	-
2a13:4c00:154e::/48	-
2a13:4c00:156a::/48	-
2a13:4c00:157f::/48	-
2a13:4c00:15dd::/48	-
2a13:4c00:15de::/48	-
2a13:4c00:15df::/48	-
2a13:4c00:15e1::/48	-
2a13:4c00:15e2::/48	-
2a13:4c00:166e::/48	-
2a13:4c00:166f::/48	-
2a13:4c00:17e1::/48	-

The AS204318 BGP Clock (2)

- Actually I lied, bgp.tools sees 40 different prefixes for AS204318, but hides them because they have too low visibility
- Low visibility means that less than 30 "vantage points" (IE: BGP sessions) can see it
  - To be honest, this might need to be higher anyway
- But having 38 stuck prefixes on this setup alone, and only on IPv6 is pretty bad!
  - This setup would be incredibly expensive to run on IPv4
  - But the problem does exist on IPv4, just harder to measure



Ok, cool, but how do you plan to fix this

-

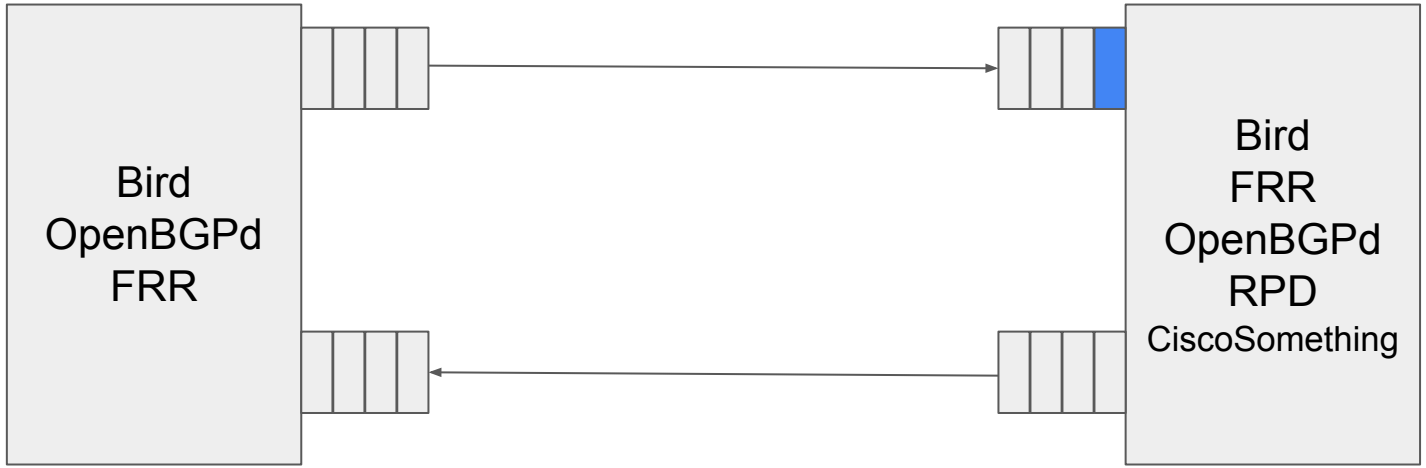
```
BLHIX_RS1_4 BGP         ---         up     2024-09-08    Established
  BGP state:          Established
    Neighbor address: 195.208.208.100
    Neighbor AS:      8631
    Local AS:         212232
    Neighbor ID:      195.208.208.100
[...]
    Session:          external AS4
    Source address:   195.208.208.202
    Hold timer:       77.776/90
    Keepalive timer:  25.632/30
```

Ok, cool, but how do you plan to fix this

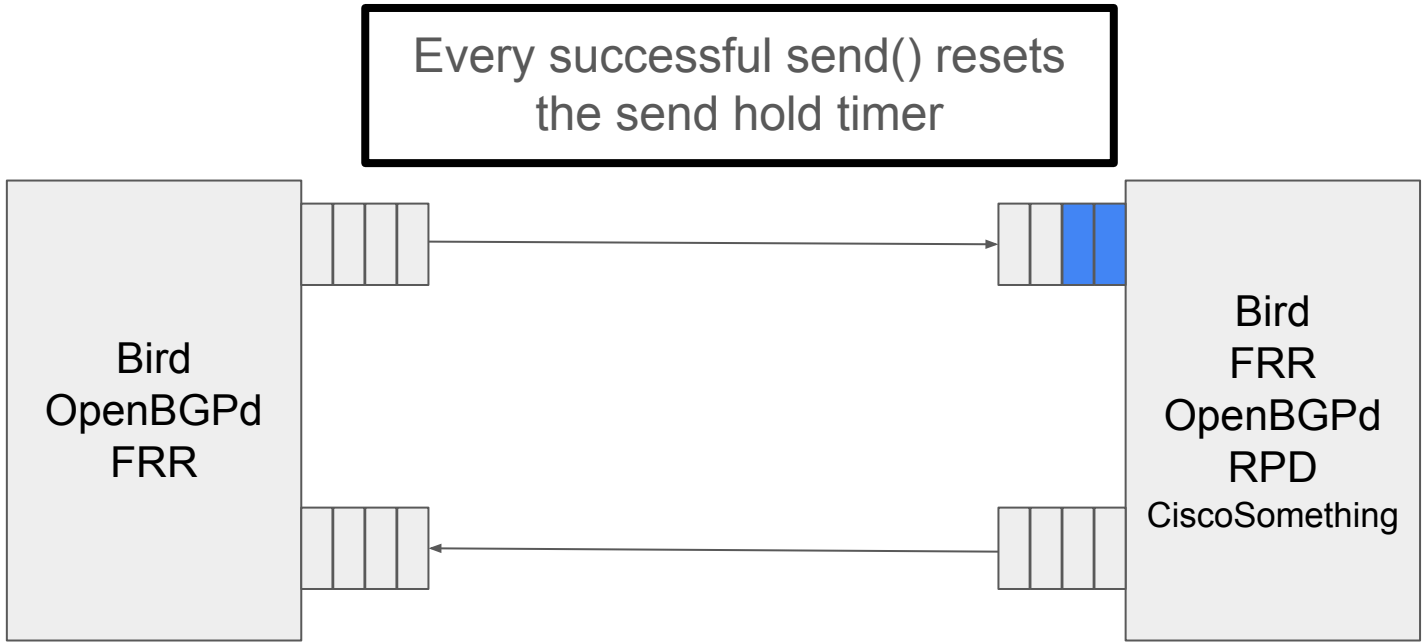
- Make a new hold timer!

```
rrsflow    BGP        ---        up     2024-06-25    Established
  BGP state:          Established
    Neighbor address: 192.168.190.253
    Neighbor AS:      206924
    Local AS:         206924
    Neighbor ID:      192.168.180.5
[...]
    Session:          internal multihop route-reflector AS4
    Source address:   185.230.223.3
    Hold timer:       115.525/240
    Keepalive timer:  32.070/80
    Send hold timer:  436.949/480
```

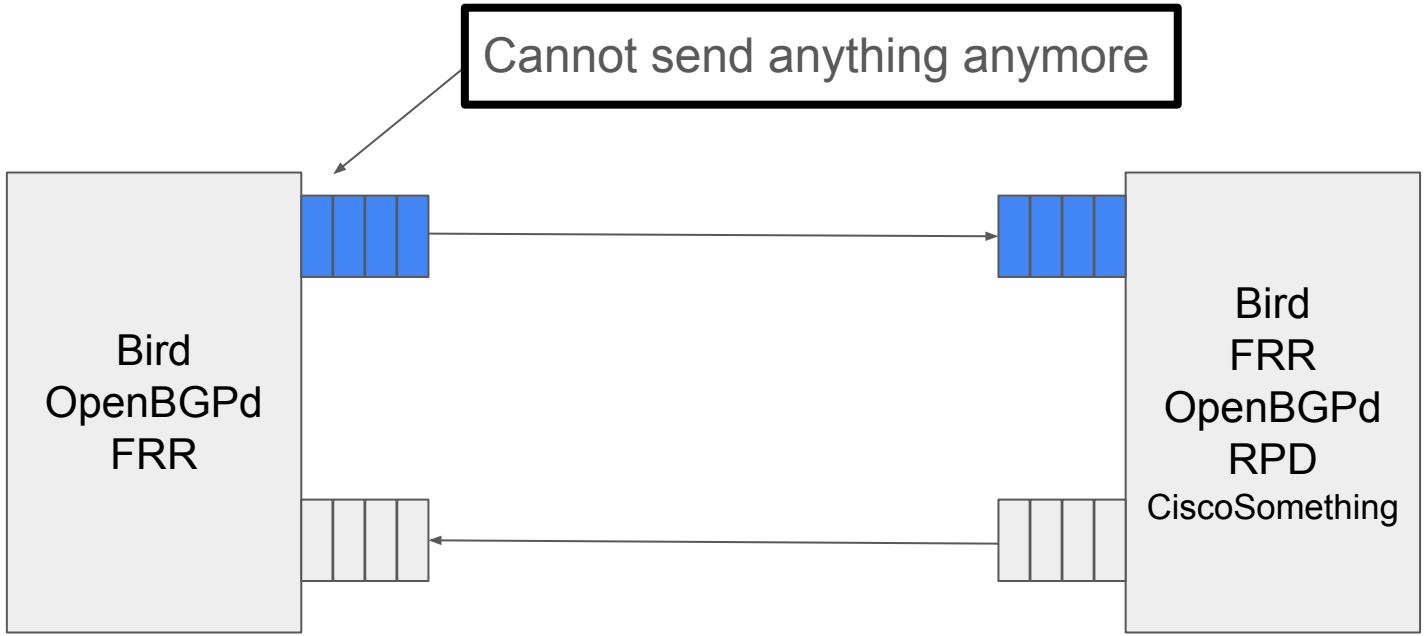
The Mechanics



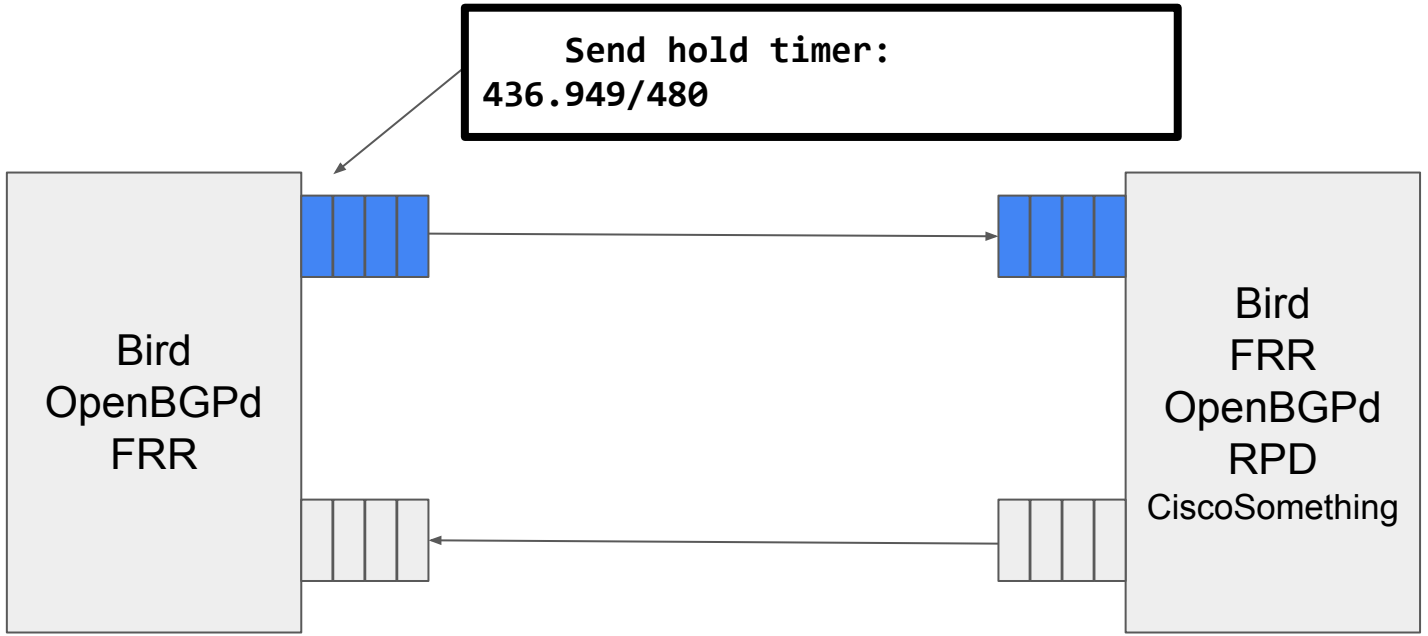
The Mechanics



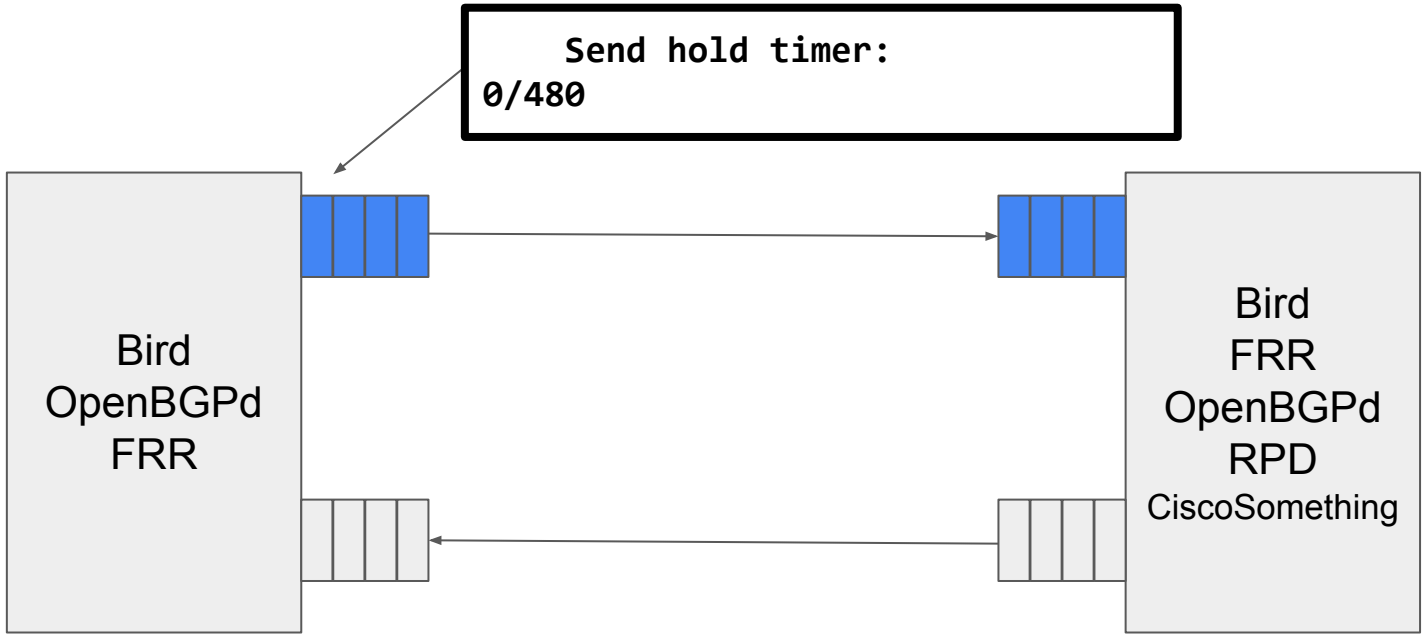
The Mechanics



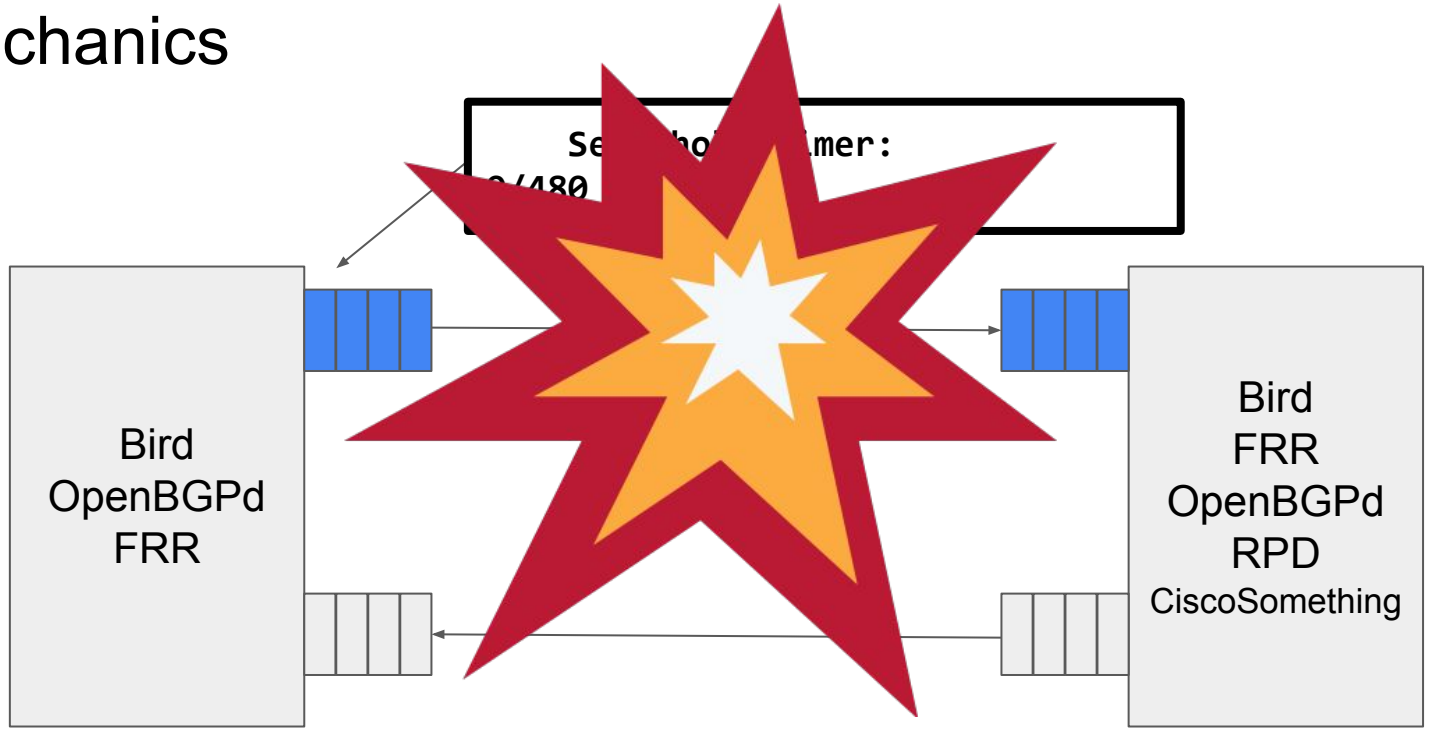
The Mechanics



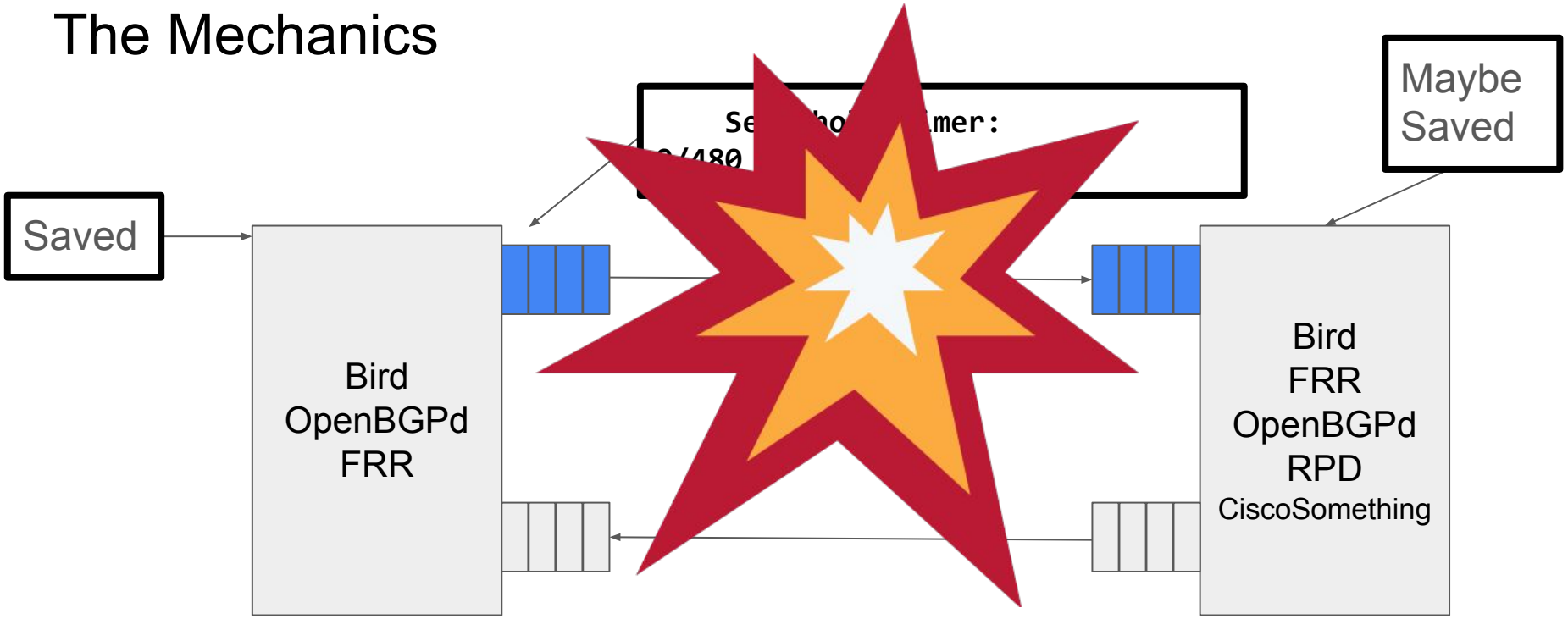
The Mechanics



The Mechanics

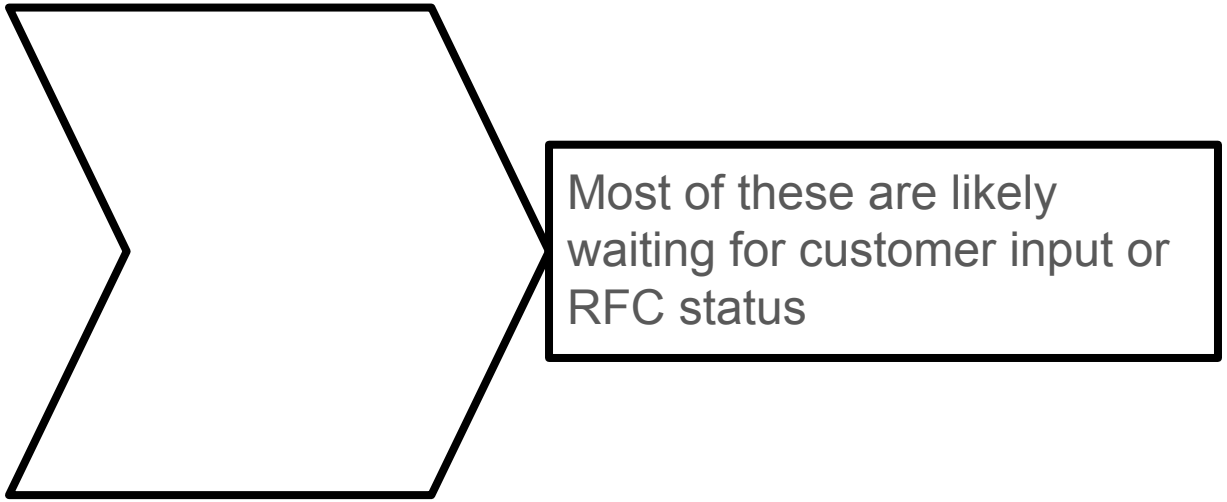


The Mechanics



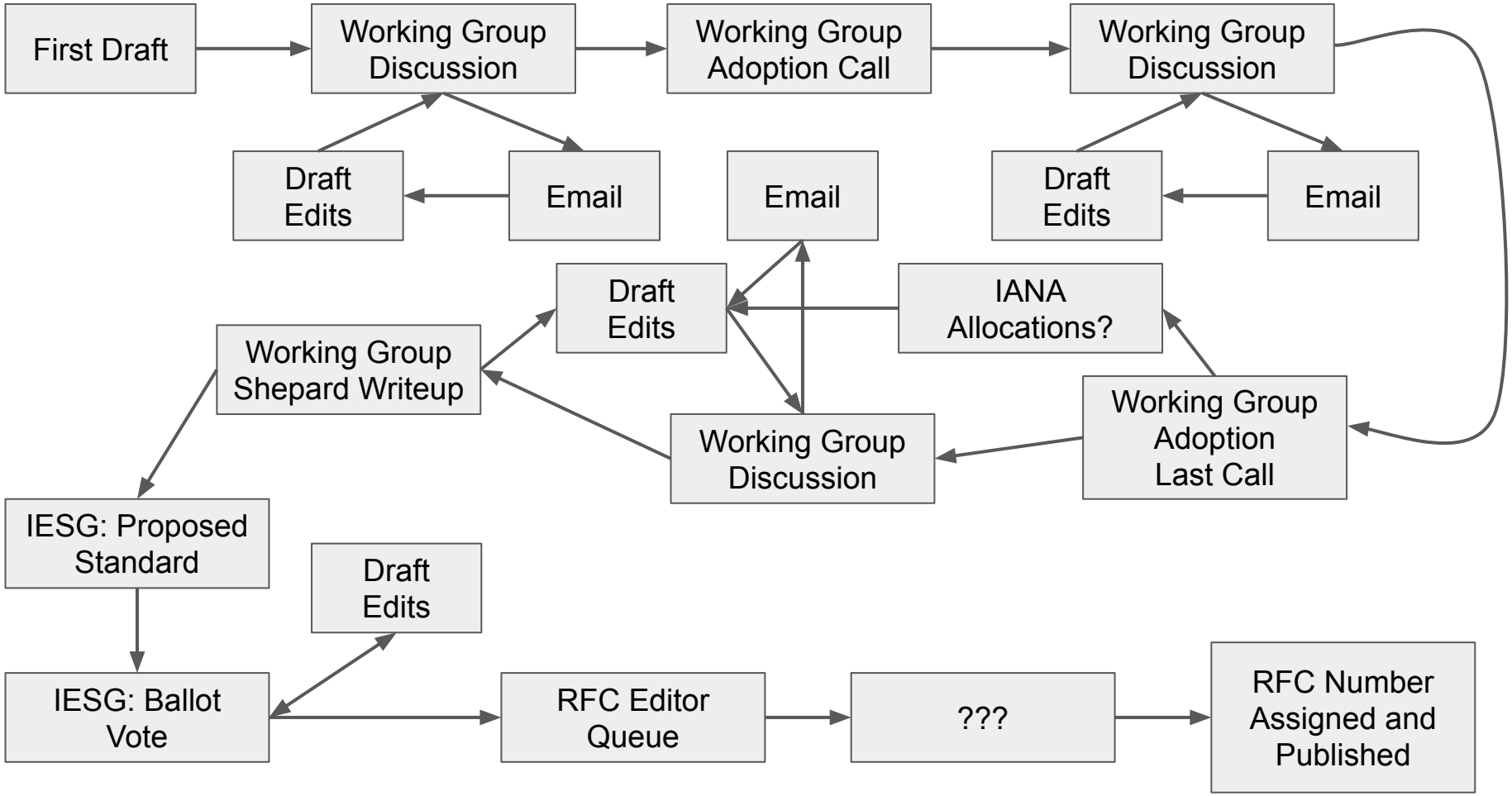
Software Availability

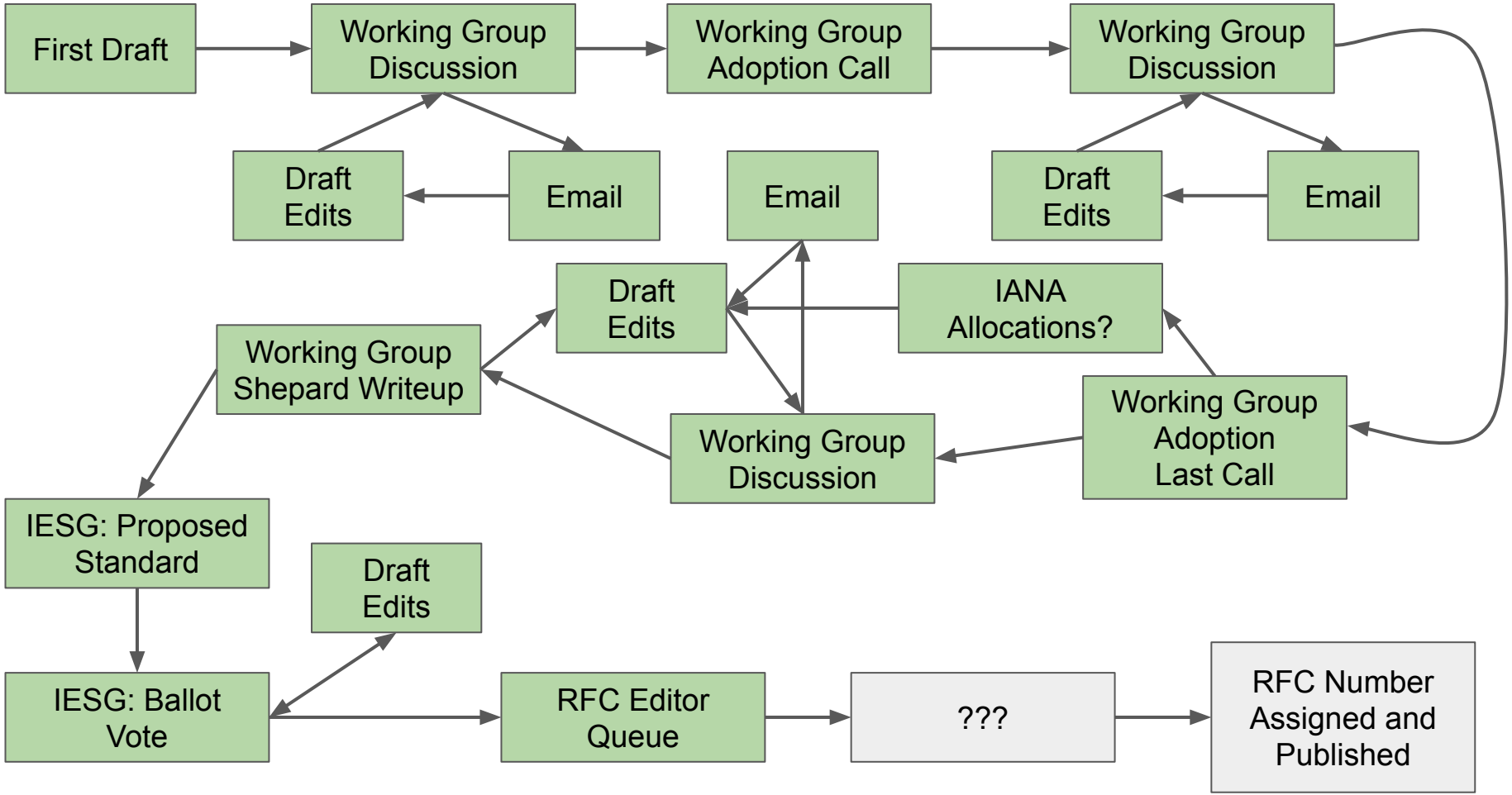
- ✅ BIRD - 2.15+
- ✅ FRR - 8.3+
- ✅ OpenBGPd - 6.9+
- ❌ Cisco
  - ❌ NX-OS
  - ❌ iOS-XR
  - ❌ iOS-XE
  - 🪦 iOS-"Classic"
- ❌ MikroTik
- ❌ EXOS
- ❌ JunOS
- ❌ EOS
- ❌ SR-OS



How hard can it be to get a RFC?

Hard, Maybe, Sometimes


First Draft
Working Group Discussion
Working Group Adoption Call
Working Group Discussion
Draft Edits
Email
Email
Draft Edits
Email
Draft Edits
IANA Allocations?
Working Group Shepard Writeup
Working Group Discussion
Working Group Adoption Last Call
IESG: Proposed Standard
Draft Edits
IESG: Ballot Vote
RFC Editor Queue
???
RFC Number Assigned and Published


First Draft
Working Group Discussion
Working Group Adoption Call
Working Group Discussion
Draft Edits
Email
Email
Draft Edits
Email
Draft Edits
IANA Allocations?
Working Group Shepard Writeup
Working Group Discussion
Working Group Adoption Last Call
IESG: Proposed Standard
Draft Edits
IESG: Ballot Vote
RFC Editor Queue
???
RFC Number Assigned and Published

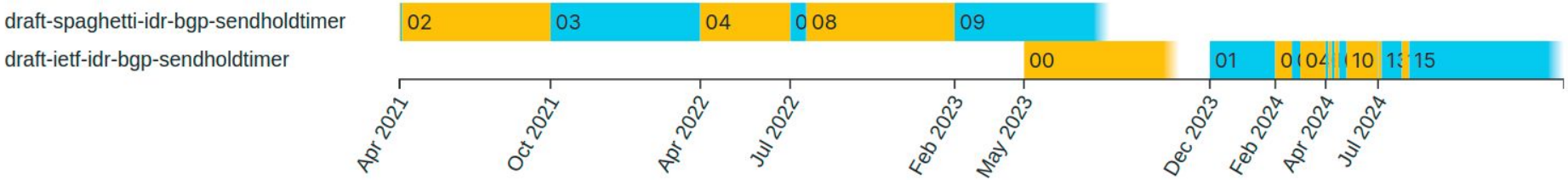
Border Gateway Protocol 4 (BGP-4) Send Hold Timer

draft-ietf-idr-bgp-sendholdtimer-15

Status | IESG evaluation record | IESG writeups | Email expansions | History

Versions:

00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15



Document	**Type**	Active Internet-Draft (idr WG)
	Authors	Job Snijders ✉, Ben Cartwright-Cox ✉, Yingzhen Qu ✉
	Last updated	2024-08-30 (Latest revision 2024-08-06)
	Replaces	draft-spaghetti-idr-bgp-sendholdtimer
	RFC stream	Internet Engineering Task Force (IETF)

IETF Timeline

- Apr 2021: "Hunting down the stuck BGP routes" and "draft-**spaghetti**-idr-bgp-sendholdtimer" is published
  - Spaghetti because you can put "anything" there before it is "adopted", so spaghetti it is.

xml2rfc

draft-example.xml U ● | Makefile

draft-example.xml

```xml
<?xml version="1.0" encoding="UTF-8"?>

<?rfc sortrefs="yes"?>
<?rfc subcompact="no"?>
<?rfc symrefs="yes"?>
<?rfc toc="yes"?>
<?rfc tocdepth="3"?>
<?rfc compact="yes"?>
<?rfc subcompact="no"?>

<rfc category="std" docName="draft-aaa-idr-skibidi-toilet-0" ipr="trust200902" submissionType="IETF" consensus="true">
<front>
  <title abbrev="Skibidi Toilet">Skibidi Toilet</title>
  <author fullname="Ben Cartwright-Cox" initials="B." surname="Cartwright-Cox">
    <organization abbrev="Port 179">Port 179 Ltd</organization>
    <address>
      <postal><street/><city>London</city><code /><country>United Kingdom</country></postal>
      <email>ben@benjojo.co.uk</email>
    </address>
  </author>
  <keyword>meme</keyword>
  <abstract>
    <t>Skibidi toilet</t>
  </abstract>
</front>

<middle>
  <section title="Introduction">
    <t>
Skibidi Toilet is a machinima web series released through YouTube videos and shorts, created by Alexey Gerasimov and uploaded on his YouTube channel
DaFuq!?Boom!. Produced using Source Filmmaker, the series follows a fictional war between human-headed toilets and humanoid characters with electroni
```

xml2rfc

```
root@ietf-containment-zone:~# pip3 install xml2rfc
[...]
root@ietf-containment-zone:~# xml2rfc draft.xml  --html --text
 Created file draft.txt
 Created file draft.html
root@ietf-containment-zone:~#
```

xml2rfc

```
# cat draft.txt




Network Working Group                                  B. Cartwright-Cox
Internet-Draft                                                 Port 179
Intended status: Standards Track                    10 September 2024
Expires: 14 March 2025


                            Skibidi Toilet
                    draft-aaa-idr-skibidi-toilet-0

Abstract

   Skibidi toilet
```

xml2rfc

- It nicely renders txt files and HTML/PDF etc for you!

Internet-Draft Skibidi Toilet September 2024

Table of Contents

1. Introduction

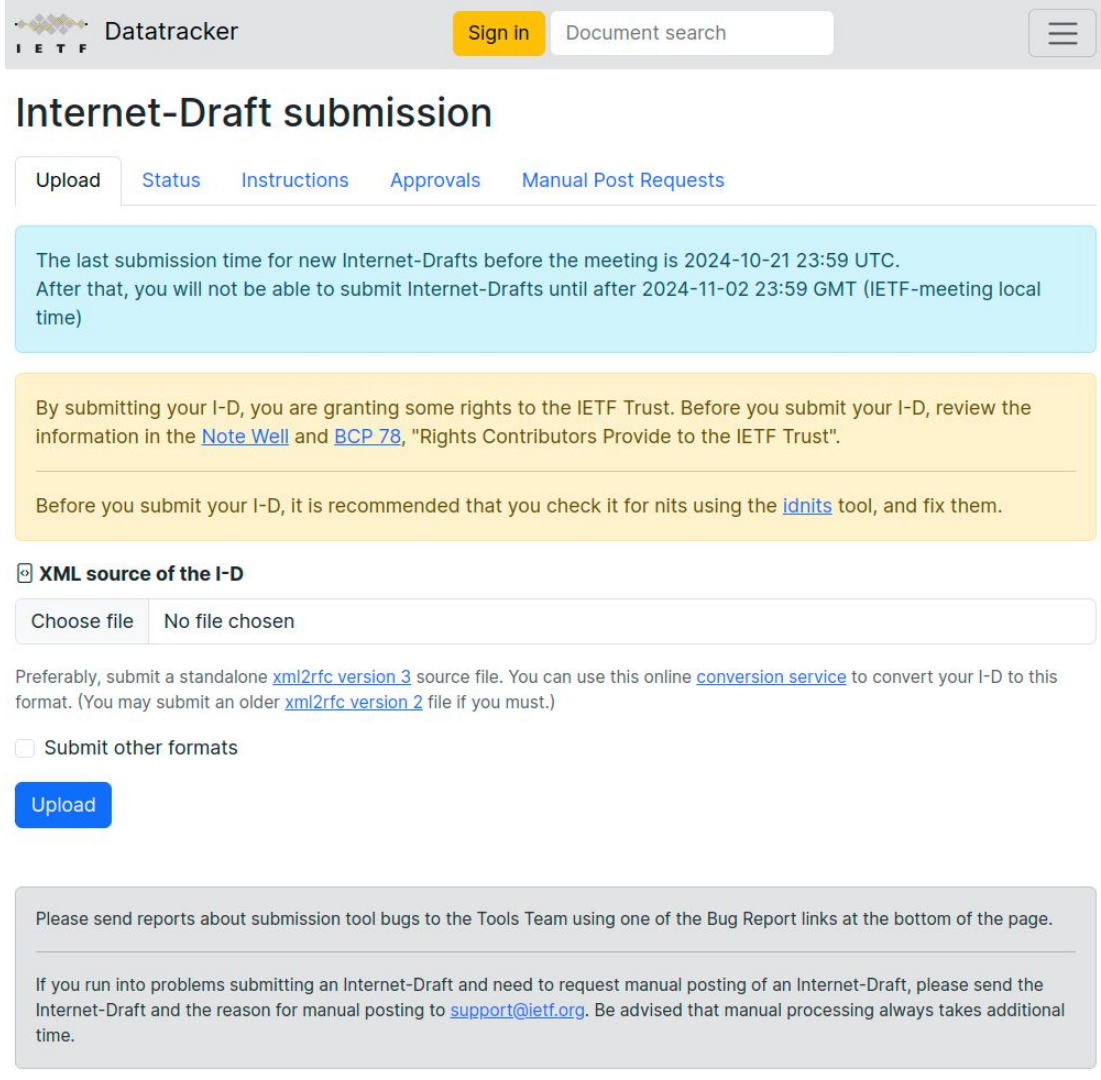
Skibidi Toilet is a machinima web series released through YouTube videos and shorts, created by Alexey Gerasimov and uploaded on his YouTube channel DaFuq!?Boom!. Produced using Source Filmmaker, the series follows a fictional war between human-headed toilets and humanoid characters with electronic devices for heads.

Since the first short was posted in February 2023, Skibidi Toilet has become viral as an internet meme across various social media platforms, particularly among Generation Alpha. Many commentators saw their embracement of the series as Generation Alpha's first

IETF Datatracker

- Consumes this XML and (after Author email verification) publishes it to the site for others to see

- Everything onwards is basically done inside datatracker submissions, or other people do stuff on datatracker to your draft



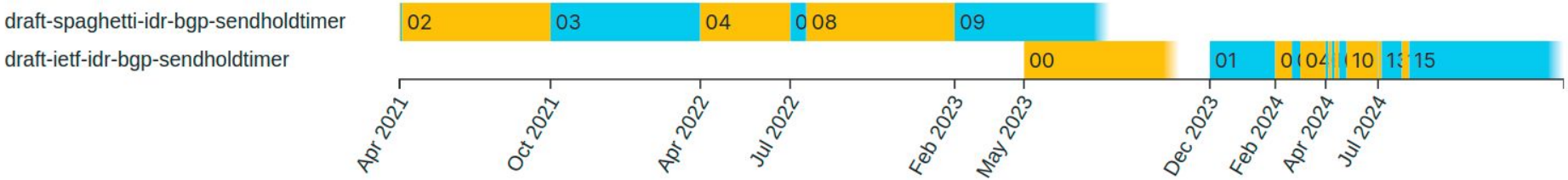
Border Gateway Protocol 4 (BGP-4) Send Hold Timer

draft-ietf-idr-bgp-sendholdtimer-15

Status | IESG evaluation record | IESG writeups | Email expansions | History

Versions:

00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15



Document	**Type**	Active Internet-Draft (idr WG)
	Authors	Job Snijders ✉, Ben Cartwright-Cox ✉, Yingzhen Qu ✉
	Last updated	2024-08-30 (Latest revision 2024-08-06)
	Replaces	draft-spaghetti-idr-bgp-sendholdtimer
	RFC stream	Internet Engineering Task Force (IETF)

IETF Timeline

- Apr 2021: "Hunting down the stuck BGP routes" and "draft-**spaghetti**-idr-bgp-sendholdtimer" is published
  - Spaghetti because you can put "anything" there before it is "adopted", so spaghetti it is.

IETF Timeline

- Apr 2021: "Hunting down the stuck BGP routes" and "draft-**spaghetti**-idr-bgp-sendholdtimer" is published
  - Spaghetti because you can put "anything" there before it is "adopted", so spaghetti it is.
- Apr 2021: *Lively* initial mailing list discussion
  - About 40 emails exchanged with various people

IETF Discussion

- The IETF is a interesting standards body, as it requires no monetary investment to participate in, Unlike:
  - IEEE SA
  - 3GPP
  - MPEG
  - GSM
  - ETSI etc
- This means that all types of people can chime in on mailing lists
  - This can be a good thing, more eyeballs can point out more flaws
  - This can be a bad thing, because people can be very annoying or potentially participate in bad faith

IETF Timeline

- Apr 2021: "Hunting down the stuck BGP routes" and "draft-**spaghetti**-idr-bgp-sendholdtimer" is published
  - Spaghetti because you can put "anything" there before it is "adopted", so spaghetti it is.
- Apr 2021: *Lively* initial mailing list discussion
  - About 40 emails exchanged with various people
- Jul 2022: Request adoption to the working group
  - ***Lively*** email discussion, About 90+ emails exchanged with a smaller group of people
  - Eventually Adopted in May 2023

IETF Working groups / Adoptions / etc

- The IETF is split into different WG (Working Groups) for various common subjects/areas of work

IETF Working groups

asap
asdf
calext
cbor
cellar
dispatch
dmarc
ecrit
emailcore
extra
jmap
mailmaint
mediaman
mimi
mlcodec
regext
satp
sipcore
sml
stir
vcon
wimse
gendispatch
6lo
6man
add
bpf
deleg
dhc
dmm
dnssd
dprive
drip
dtn
intarea
madinas
ntp
schc
snac
tictoc
anima
bmwg
dnsop
grow
iotops
ippm
ivy
mboned
mops
netconf
netmod
nmop
opsawg
sidrops
srv6ops
v6ops
babel
bess
bfd
bier
cats
ccamp
detnet
idr
lisp
lsr
lsvr
manet
mpls
nvo3
pals
pce
pim
rift
roll
rtgwg
savnet
spring
teas
tvr
ace
acme
cose
dance
dult
emu
gnap
ipsecme
jose
keytrans
kitten
lake
lamps
mls
oauth
ohai
openpgp
ppm
pquip
privacypass
radext
rats
scim
scitt
secdispatch
spice
suit
teep
tls
uta
avtcore
ccwg
cdni
core
httpapi
httpbis
masque
moq
nfsv4
quic
taps
tcpm
tsvwg
webtrans
wish

IETF Working groups

asap
asdf
calext
cbor
cellar
dispatch
dmarc
ecrit
emailcore
extra
jmap
mailmaint
mediaman
mimi
mlcodec
regext
satp
sipcore
sml
stir
vcon
wimse
gendispatch
6lo
6man
add
bpf
deleg
dhc
dmm
dnssd
dprive
drip
dtn
intarea
madinas
ntp
schc
snac
tictoc
anima
bmwg
dnsop
grow
iotops
ippm
ivy
mboned
mops
netconf
netmod
nmop
opsawg
sidrops
srv6ops
v6ops
babel
bess
bfd
bier
cats
ccamp
detnet

idr

lisp
lsr
lsvr
manet
mpls
nvo3
pals
pce
pim
rift
roll
rtgwg
savnet
spring
tvr
ace
acme
cose
dance
dult
emu
gnap
ipsecme
jose
keytrans
kitten
lake
lamps
mls
oauth
ohai
openpgp
ppm
pquip
privacypass
radext
rats
scim
scitt
secdispatch
spice
suit
teep
tls
uta
avtcore
ccwg
cdni
core
httpapi
httpbis
masque
moq
nfsv4
quic
taps
tcpm
tsvwg
webtrans
wish

IETF Working groups / Adoptions / etc

- The IETF is split into different WG (Working Groups) for various common subjects/areas of work
- You upload stuff to datatracker, talk about it on the WG mailing list
- Assuming you think people are not going to vote against it, you can ask the WG Chairs to Adopt the draft
- This means that the draft is now "part" of the working group, and generally means it's taken a lot more seriously, and tracked on chair's agenda
- It does not mean that you are going to "succeed" however in getting your draft published (aka, turned into a RFC)

IETF Timeline

- Apr 2021: "Hunting down the stuck BGP routes" and "draft-**spaghetti**-idr-bgp-sendholdtimer" is published
  - Spaghetti because you can put "anything" there before it is "adopted", so spaghetti it is.
- Apr 2021: *Lively* initial mailing list discussion
  - About 40 emails exchanged with various people
- Jul 2022: Request adoption to the working group
  - ***Lively*** email discussion, About 90+ emails exchanged with a smaller group of people
  - Eventually Adopted in May 2023
- Dec 2023: Working Group Last Call request

Working Group Last Call

- Basically the last point where the mailing lists views are collected, if the "vibe" is right, then they are accepted

- After the last call (and it gets accepted) then you are basically just dealing with the WG Chairs, IESG, IANA etc

IETF Timeline

- Apr 2021: "Hunting down the stuck BGP routes" and "draft-**spaghetti**-idr-bgp-sendholdtimer" is published
  - Spaghetti because you can put "anything" there before it is "adopted", so spaghetti it is.
- Apr 2021: *Lively* initial mailing list discussion
  - About 40 emails exchanged with various people
- Jul 2022: Request adoption to the working group
  - ***Lively*** email discussion, About 90+ emails exchanged with a smaller group of people
  - Eventually Adopted in May 2023
- Dec 2023: Working Group Last Call request
- May 2024: Submitted to IESG for publication
- July 2024: ISEG Last Call Requested

IESG

- Just before a working group chair write up happens that tries to summarize all of the discussion that has happened to your draft so far.
- Assuming all goes well, then it gets punted to the IESG, a different set of people who look over the document to see if it makes sense to them, it then goes for a "IESG Ballot"

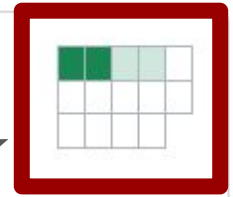
Informational	**DNSSEC Trust Anchor Publication for the Root Zone**
IETF stream	draft-ietf-dnsop-rfc7958bis
Token	Warren "Ace" Kumari (OPS)
IANA review	Version Changed - Review Needed
Consensus	Yes
Reviews	DNSDIR Telechat review by Petr Špaček **Ready w/nits** DNSDIR Telechat review (of -05) by Petr Špaček **Ready w/issues** DNSDIR Telechat review (of -04) by Petr Špaček **Almost ready** GENART Last Call review (of -03) by Dan Romascanu **Ready w/issues** ARTART Last Call review (of -03) by Scott Hollenbeck **Ready w/nits** SECDIR Last Call review (of -03) by Klaas Wierenga **Has nits** DNSDIR Last Call review (of -03) by Petr Špaček **On the right track** DNSDIR Early review (of -00) by Florian Obser **Ready w/nits** SECDIR Telechat Review due 2024-09-17 **Incomplete** OPSDIR Last Call Review due 2024-08-08 **Incomplete**

Datatracker

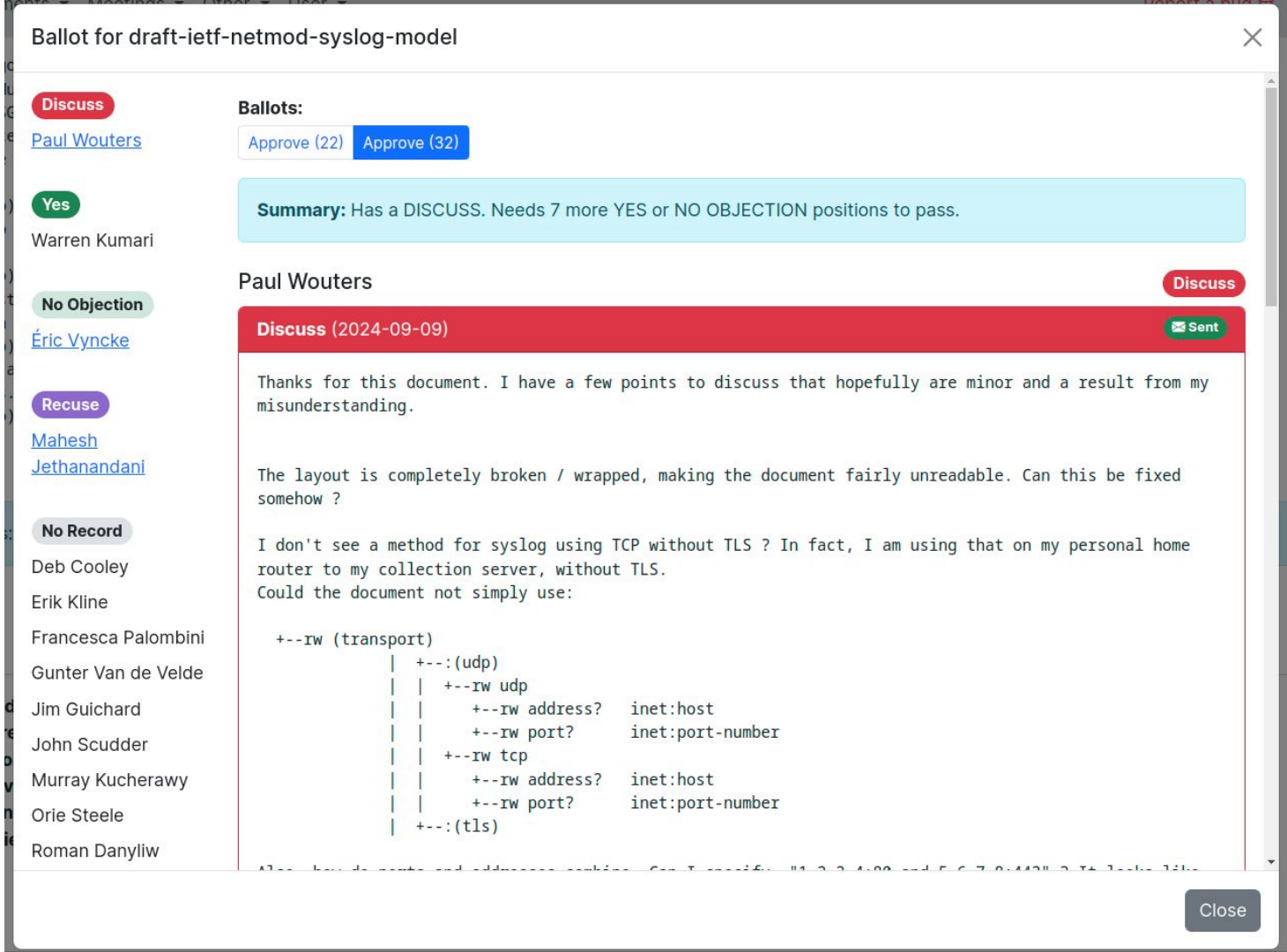
Sign in

Document search

Informational	**DNSSEC Trust Anchor Publication for the Root Zone**
IETF stream	draft-ietf-dnsop-rfc7958bis
Token	Warren "Ace" Kumari (OPS)
IANA review	Version Changed - Review Needed
Consensus	Yes
Reviews	DNSDIR Telechat review by Petr Špaček **Ready w/nits**
	DNSDIR Telechat review (of -05) by Petr Špaček **Ready w/issues**
	DNSDIR Telechat review (of -04) by Petr Špaček **Almost ready**
	GENART Last Call review (of -03) by Dan Romascanu **Ready w/issues**
	ARTART Last Call review (of -03) by Scott Hollenbeck **Ready w/nits**
	SECDIR Last Call review (of -03) by Klaas Wierenga **Has nits**
	DNSDIR Last Call review (of -03) by Petr Špaček **On the right track**
	DNSDIR Early review (of -00) by Florian Obser **Ready w/nits**
	SECDIR Telechat Review due 2024-09-17 **Incomplete**
	OPSDIR Last Call Review due 2024-08-08 **Incomplete**



- 4 possible states
- Discuss
  - As close to "no" as possible for the IETF
- Yes
- No Objection
  - For when people don't have a strong/educated view
- Recuse
  - Used when it is not appropriate for someone to vote (Conflict of Interest etc)

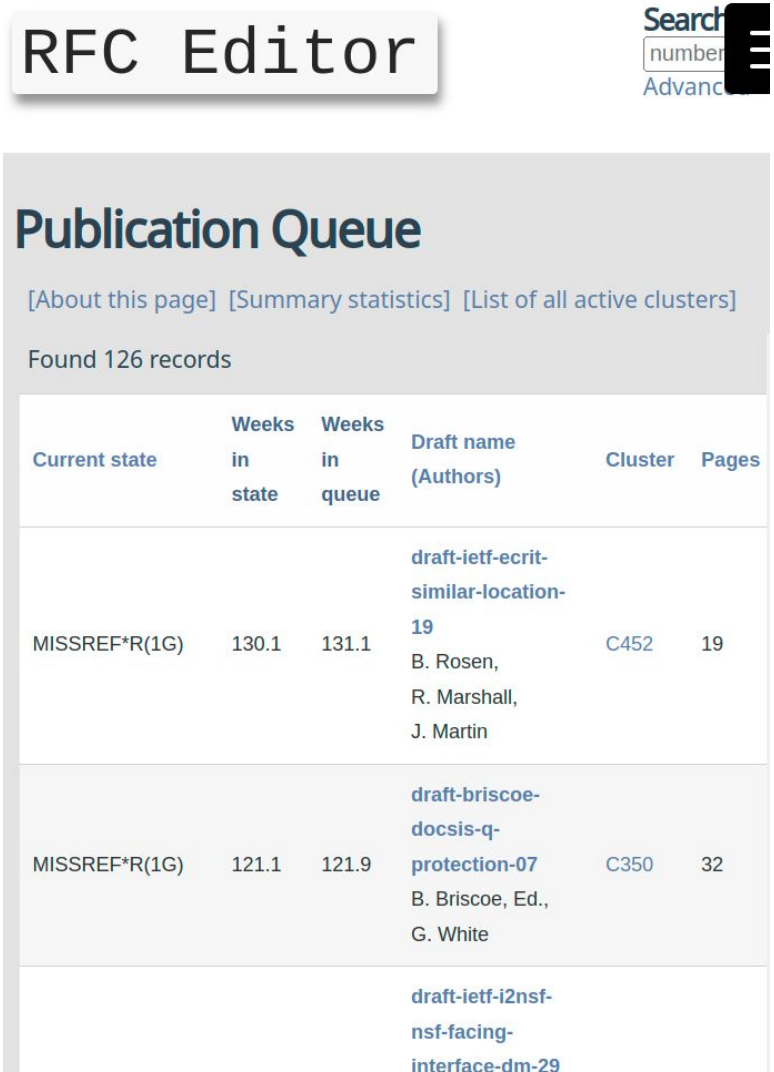


IETF Timeline

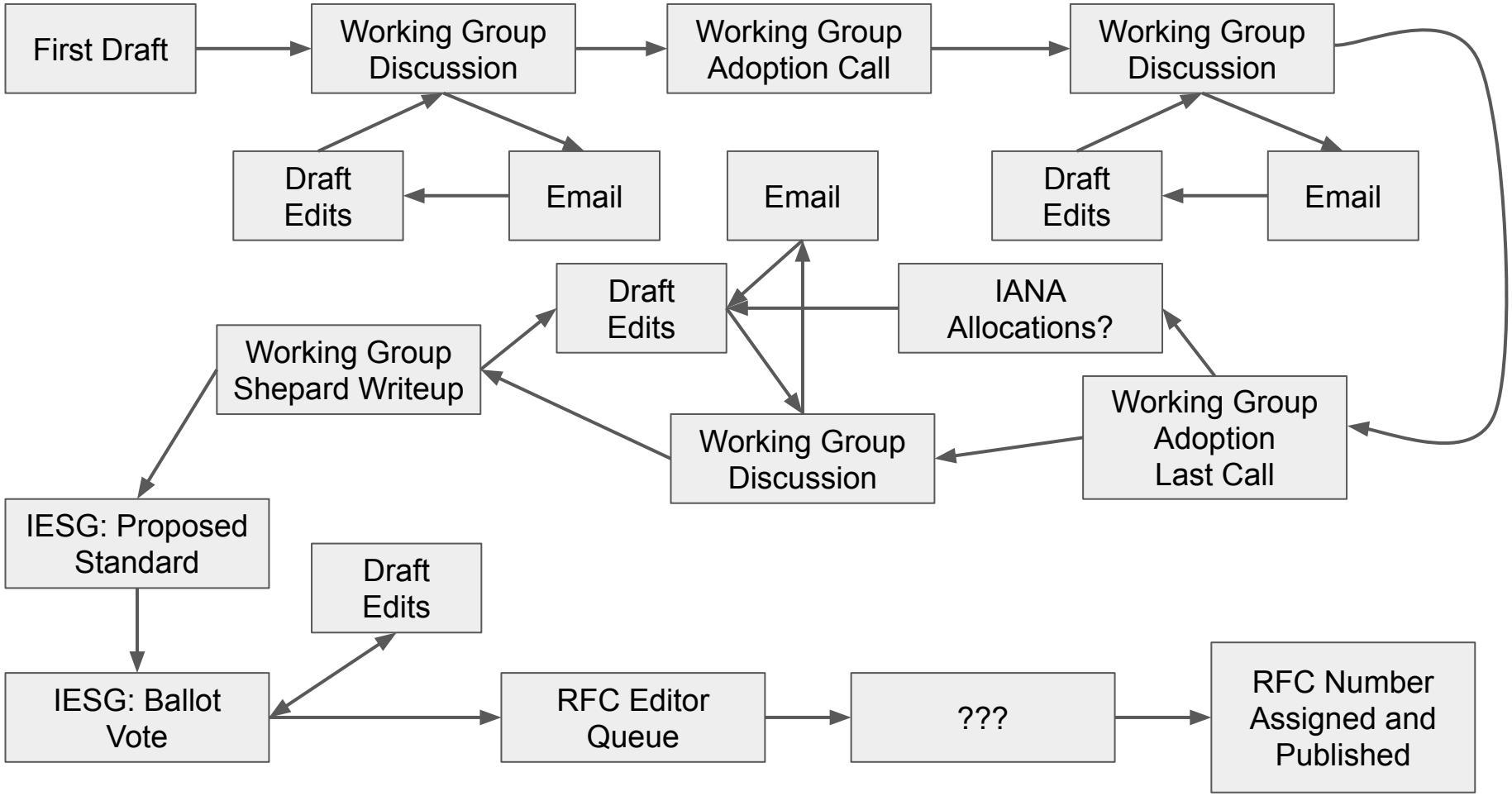
- Apr 2021: "Hunting down the stuck BGP routes" and "draft-**spaghetti**-idr-bgp-sendholdtimer" is published
  - Spaghetti because you can put "anything" there before it is "adopted", so spaghetti it is.
- Apr 2021: *Lively* initial mailing list discussion
  - About 40 emails exchanged with various people
- Jul 2022: Request adoption to the working group
  - ***Lively*** email discussion, About 90+ emails exchanged with a smaller group of people
  - Eventually Adopted in May 2023
- Dec 2023: Working Group Last Call request
- May 2024: Submitted to IESG for publication
- July 2024: ISEG Last Call Requested
- Aug 2024: RFC Editor Status

IETF Editor

- A group of people who are paid to be the final editor for drafts
- This is basically the last step before numbers are assigned and draft is published
- Stuff can spend a long time in here, if errors are found or if the document is really long!
- **This is where sendholdtimer currently is**
  - I ***think*** we are only weeks away from being published



So, yeah


First Draft
Working Group Discussion
Working Group Adoption Call
Working Group Discussion
Draft Edits
Email
Email
Draft Edits
Email
Draft Edits
IANA Allocations?
Working Group Shepard Writeup
Working Group Discussion
Working Group Adoption Last Call
IESG: Proposed Standard
Draft Edits
IESG: Ballot Vote
RFC Editor Queue
???
RFC Number Assigned and Published

IETF Timeline

- Apr 2021: "Hunting down the stuck BGP routes" and "draft-**spaghetti**-idr-bgp-sendholdtimer" is published
  - Spaghetti because you can put "anything" there before it is "adopted", so spaghetti it is.
- Apr 2021: *Lively* initial mailing list discussion
  - About 40 emails exchanged with various people
- Jul 2022: Request adoption to the working group
  - ***Lively*** email discussion, About 90+ emails exchanged with a smaller group of people
  - Eventually Adopted in May 2023
- Dec 2023: Working Group Last Call request
- May 2024: Submitted to IESG for publication
- July 2024: ISEG Last Call Requested
- Aug 2024: RFC Editor Status

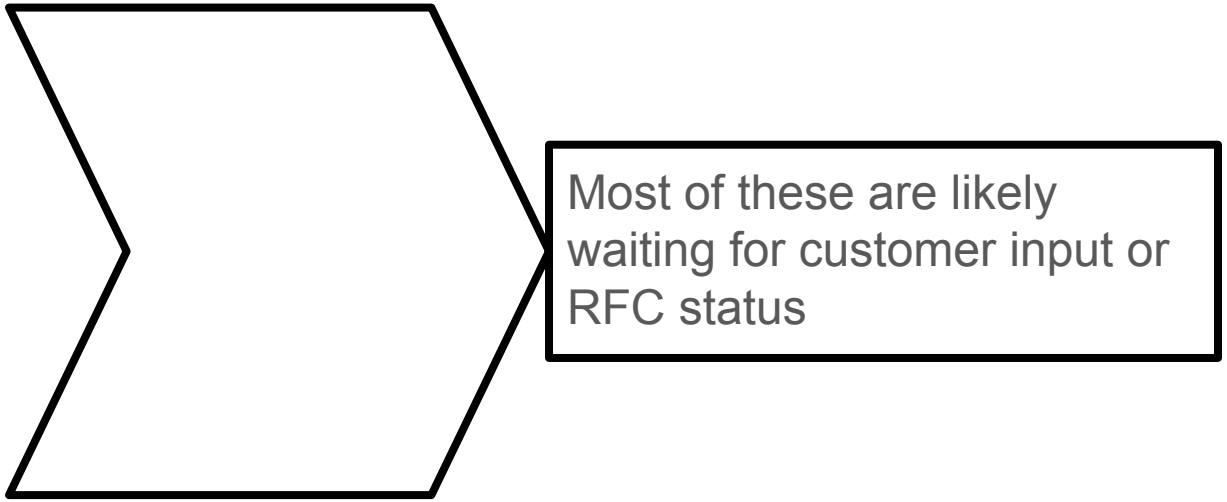
Hindsight

- I would likely not do this again, but I don't regret doing this
- You have to have someone experienced help you with this
  - Thank you to Job Snijders, who has been extremely knowledgeable
  - And also motivating me the many times when I wanted to give up on the IETF to keep going
- It's possible that this process was more painful due to COVID
  - But it's hard to prove
- Changing BGP behaviour is a extremely contentious thing to do
  - I was not expecting that much of a fight, I got quite a lot of one from a small-ish group of people

Anyway

Software Availability

- ✅ BIRD - 2.15+
- ✅ FRR - 8.3+
- ✅ OpenBGPd - 6.9+
- ❌ Cisco
  - ❌ NX-OS
  - ❌ iOS-XR
  - ❌ iOS-XE
  - 🪦 iOS-"Classic"
- ❌ MikroTik
- ❌ EXOS
- ❌ JunOS
- ❌ EOS
- ❌ SR-OS



Closing remarks

- If you are running software that could have this feature today, I would highly recommend upgrading to a version that has it
- Getting something to RFC is a whole skill level on it's own

- The fight is not over on stuck routes, I suspect these TCP "stick" failures are only ~30% of all stuck routes
  - The rest are unidentified(*) vendor bugs
- But we do already have a confirmed case of SendHoldTimer triggering as intended against a hardware vendor's implementation

Questions?

Shy? nonog-talk@benjojo.co.uk