The routing challenges in IPv6 DoS mitigation

David Freedman – Claranet – IPv6 Security Workshop – July 2017

claranet
hosting | applications | networks

This talk is about…

- **Using routing to steer or block traffic.**

This talk is not about…

- **What you do with it afterward.**

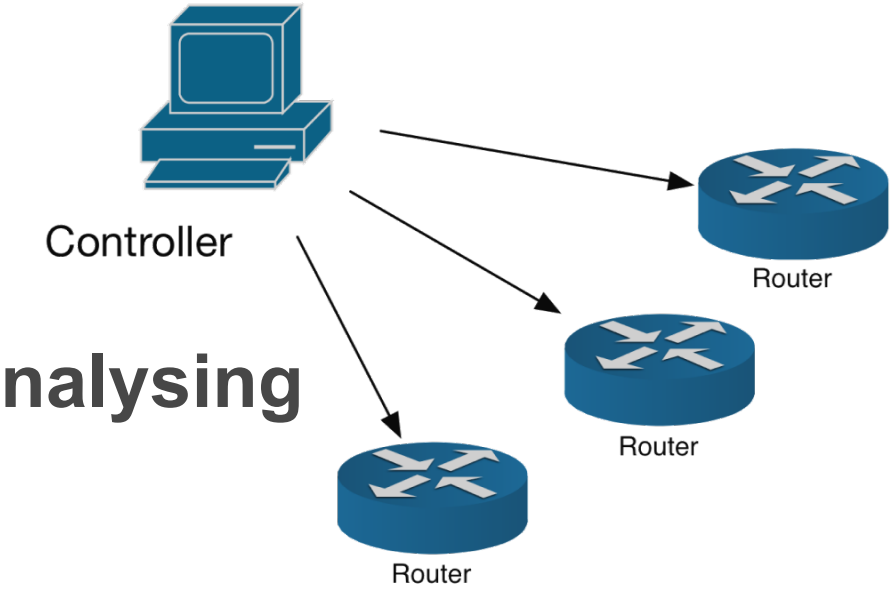
claranet
hosting | applications | networks

General principles – mitigation.

- Traffic being blocked announced as a host route.
- Applies to destination, but also source, using loose uRPF.
- You need a dummy next-hop for discard.
  - All routers should forward traffic to the next-hop to the bit-bucket.
  - Until 2012, our next-hop was `2001:a88::dead`
  - That was until RFC6666 came about, then we switched to `100::dead`
- Traffic selectively filtered of course has a real next-hop.
  - Real next-hop is the filtering platform.

In practice

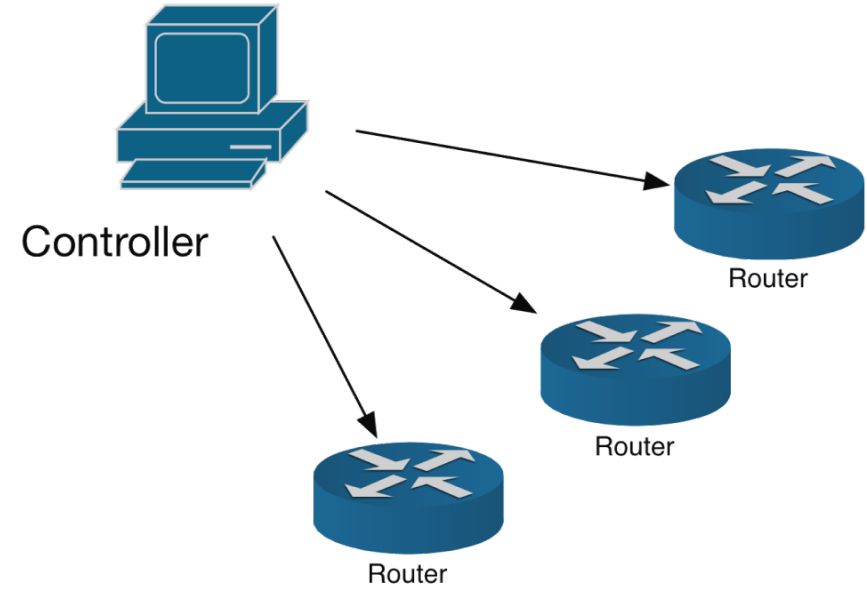
- A controller lives behind the scenes.
- Controller makes decisions based on analysing flow data.
  - It could decide to block.
  - It could decide to filter.
- Controller also has full BGP feeds.
- Controller runs iBGP.
  - Needs full visibility of all paths.
  - Needs to inject with access to all attributes.
  - Can't have next-hop overwritten (more on this later).



claranet
hosting | applications | networks

In practice

- Controller lives out-of-band.
  - Along with supporting infrastructure.
- OOB network directly attached.
  - Via dedicated ports/circuits.
  - To a number, but not all routers.
  - Exists to support our network globally.
- OOB network has a dedicated routing domain
  - Signalled eBGP on interconnects.
  - Overlaid iBGP to ISP routers for traffic steering.



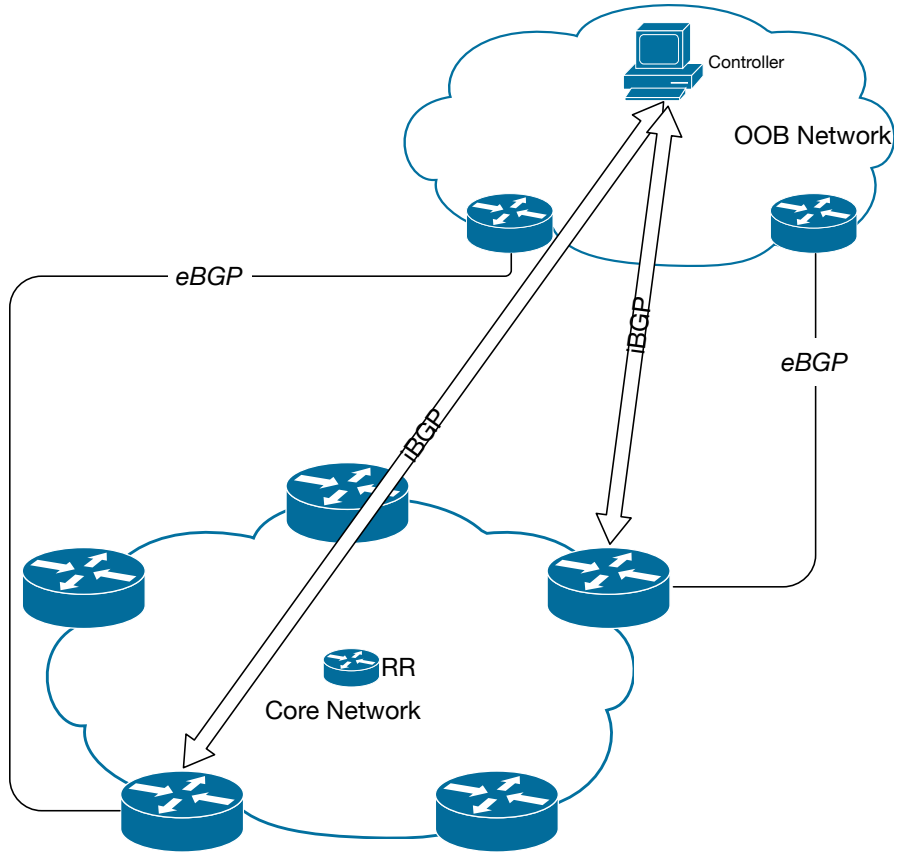
A detour

- Network had native IPv6 in 2001
- MPLS deployed in 2003
  - IPv4 Label switched (LDP signalled).
  - IPv6 routed.
  - Only solution for MPLS was 6PE.
  - I stubbornly held out for LDP6.
    - Which never came, as I lamented my decision in a 2009 interview (see picture).
- We eventually deployed 6PE in late 2013, 6VPE was an added bonus.
- DoS mitigation platform does not directly support 6PE
  - But attachment to core network doesn't require it.
  - Plan was to use iBGP 2/1 (unicast) and then iBGP 2/4 (label) internally.
  - This is where it all started going wrong…

claranet
hosting | applications | networks

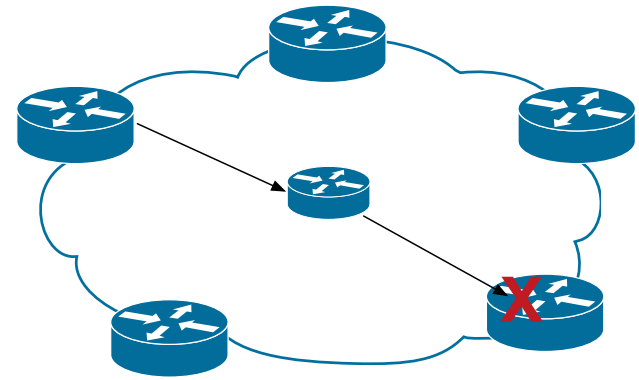
The scene is set.

- **OOB Network attached to core (IOS-XR for reference).**
- **OOB Network infrastructure signalled via eBGP 2/1 (ipv6 unicast).**
- **Controller has (congruent) iBGP 2/1 sessions to relevant core routers.**
- **Routers treat controller as RR-Client (important).**
- **Controller signals prefixes with relevant n-h (e.g. `100::dead`) via iBGP 2/1.**
- **Routers send prefixes back to RR as iBGP 2/4.**



claranet
hosting | applications | networks

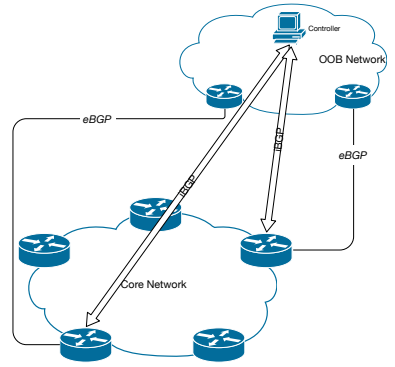
The original problem



- It didn't work ☹
- Prefix was accepted at core<->OOB edge, sent to RR and then onward to RR clients.
- However, the clients didn't accept it.
- next-hop (`100::DEAD`) a local discard route, in 2/1 LOC-RIB, but prefix accepted over 2/4 RR session.
- IOS-XR doesn't like this, and we couldn't make it like this, neither could Cisco make it like this.
- The controller can't speak eBGP or iBGP 2/4.
- Seemingly the only solution was to re-write the next-hop.

claranet
hosting | applications | networks

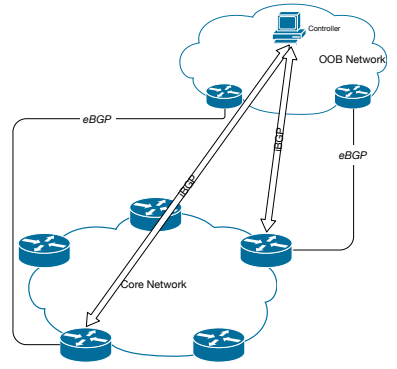
Rewriting the next-hop.



- No candidates in 2/4 to do this with.
- Nor can you specify a 2/4 next-hop manually.
- We were faced with next-hop-self.
- Traffic would have to be dragged across the core and terminated at the OOB interconnect point, then discarded.
- Messy solution, but it was our only hope of getting it working.
- Route-policy configured on OOB interconnect, set n-h-s.

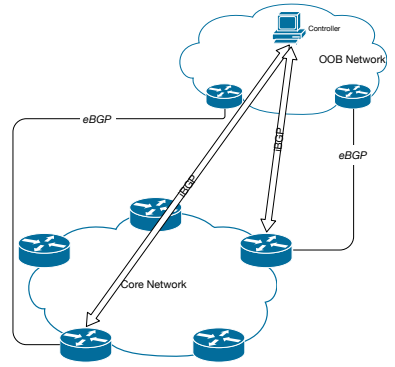
claranet
hosting | applications | networks

But it didn't work.



- Of course, you can't re-write next-hops in iBGP, unless you are the originating router (we are not).
- Special IOS-XR knob for doing this (CSCsh33618):
  - `ibgp policy out enforce-modifications`
- This is a global command to the router.
- Also, n-h-s needs to be configured toward the reflector.
- Adding it enabled next-hop-self to work.
- But it also re-wrote next-hop for other configured clients, unintentionally.

Selective re-write.

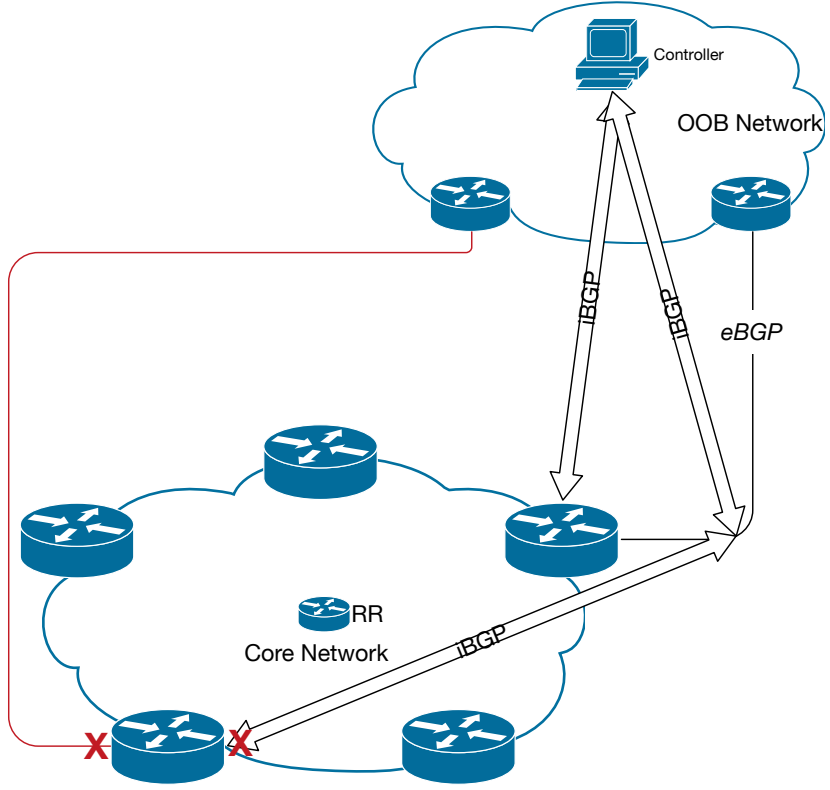


- At which point we are forced into a corner.
- Selective next-hop re-write was needed:
  - If next-hop was to be discard or filter, re-write as next-hop-self
- Policy again has to be toward reflector.
  - Ugly and getting uglier.
- Applied and it seemed to work.
- Then broke mysteriously one day when one of the infrastructure links failed.

claranet
hosting | applications | networks

Fault intolerance.

- One day, one of the infrastructure links failed between the OOB and Core.
- eBGP session was **down**.
- (now incongruent) iBGP session was **also down.**
  - Why? The iBGP used the infrastructure as a transport network.
  - iBGP TCP segments were being delivered via the OOB infrastructure to the alternate core router, destined to the original core router where the iBGP session endpoint lived.
- Core router was <u>discarding</u> the BGP TCP segments
  - Session torn down, couldn't be re-established.

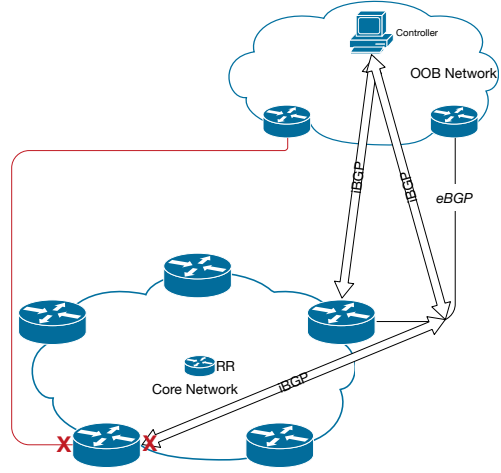


claranet
hosting | applications | networks

Yup, IOS-XR again…

- Not possible to peer over un(v6)numbered 6PE interface.
- Cisco eventually raised a DDTS (CSCuc56355).
- Even worse, the TCP SYNs cause buffer leaks ☹
- Eventually, SPP buffers exhaust, TCP stops working.
- Only option at that point is to restart NETIO.
- Workaround is supposedly numbering the core interface.
- But we can't use this as it has implications for our IGP.
  - We migrated from dual to single topology IS-IS during the 6PE implementation.
  - Adding v6 numbering to these interfaces would break our ST IS-IS ☹

What options did we have?

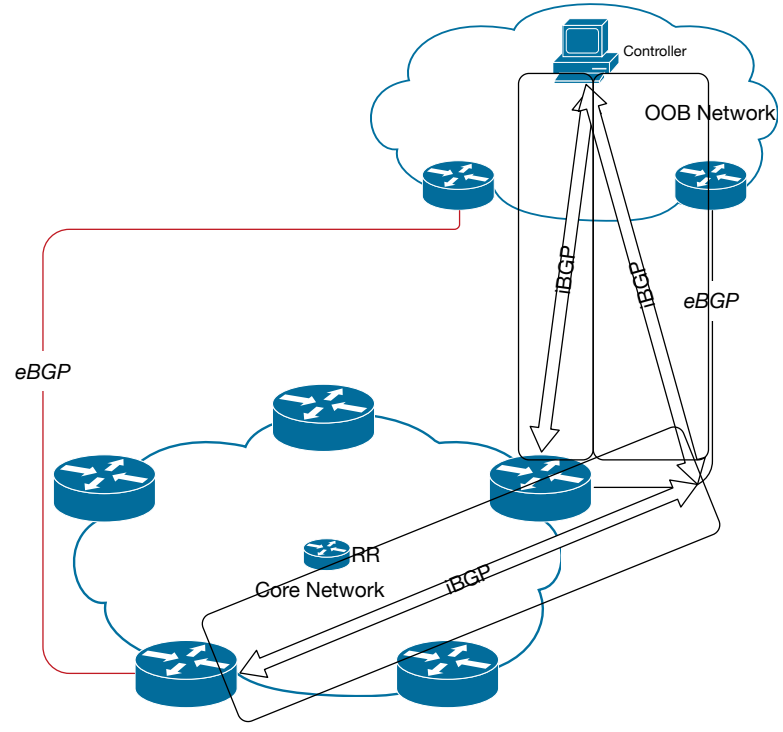


- We needed to deliver BGP TCP segments to the core routers.
- This delivery could not be labelled.
- Routing it would mean:
  - Re-adding v6 to the core p2p.
  - Adding v6 AFI back into the ST IS-IS or adding v6 AFI as a second topology (using MT-IS-IS).
  - Adding iBGP 2/1 sessions between core and RRs.
  - Adding BGP 2/1 to the RRs.
- At this point, it felt as if we were backing out of 6PE.

claranet
hosting | applications | networks

Solution : IP overlay

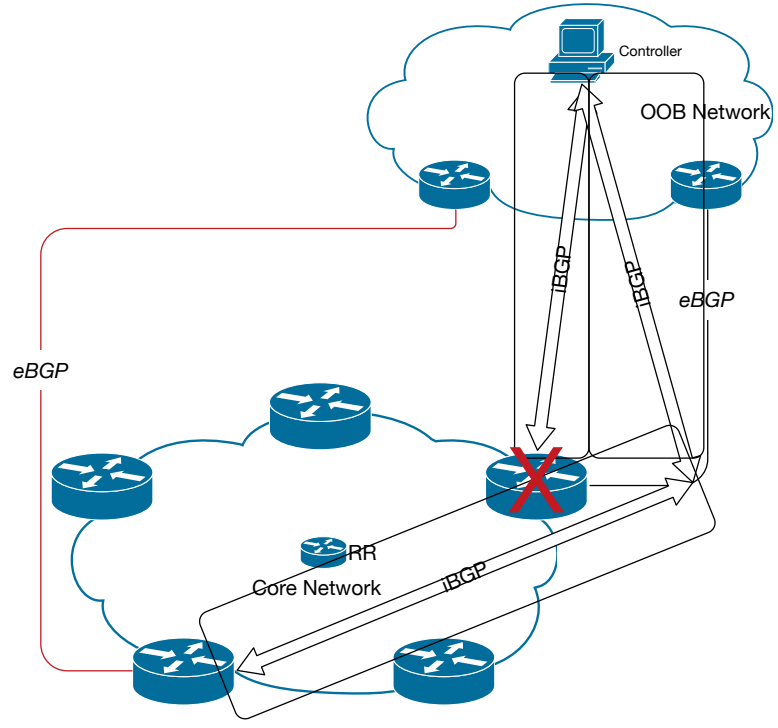
- At this point, we had to reach for the big sticking plaster, some form of overlay.
- No MPLS, as TCP couldn't be labelled.
- We deployed uni-directional GRE from the OOB infrastructure to the core handoff points.
- This meant the TCP arrived in IP, and made things work.



claranet
hosting | applications | networks

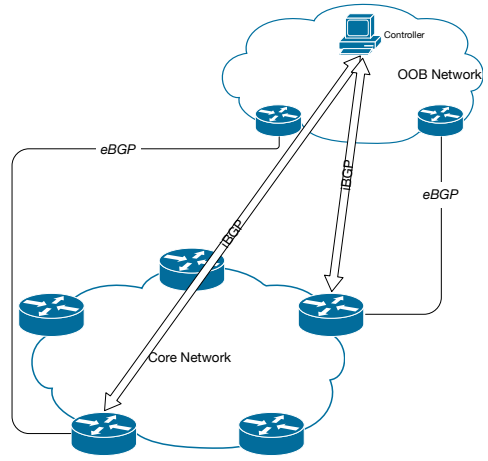
But then..

- We hobbled along with these tunnels for some time, just carrying part of the BGP signalling.
- One day we upgraded IOS-XR.
- The upgrade broke the GRE forwarding ☹
- It turns out that GRE over FRR is not officially supported.
- Actually, GRE over MPLS wasn't officially supported until the version we upgraded to.
- This means that adding official GRE over MPLS support broke our (working) GRE over MPLS over FRR support ☹



claranet
hosting | applications | networks

Just as we were about to give up..



- We noticed the BGP issue had been partially fixed.
- Though BGP active transport definitely didn't work, passive did.
- If the controller initiated the session, the core router could respond and establish it.
- We ripped out the GRE and moved back to native transport.

claranet
hosting | applications | networks

Was it all worth it?

- **Not really. 6PE + iBGP was a poor choice.**
  - **Wasted months of time.**
  - **Annoyed us all.**
- **In short term, we're looking to move to eBGP.**
- **Longer term, probably remove 6PE and move to 6SR.**
  - **6PE is still painful to troubleshoot.**
  - **Want to re-use label core.**

Any questions?

claranet
hosting | applications | networks