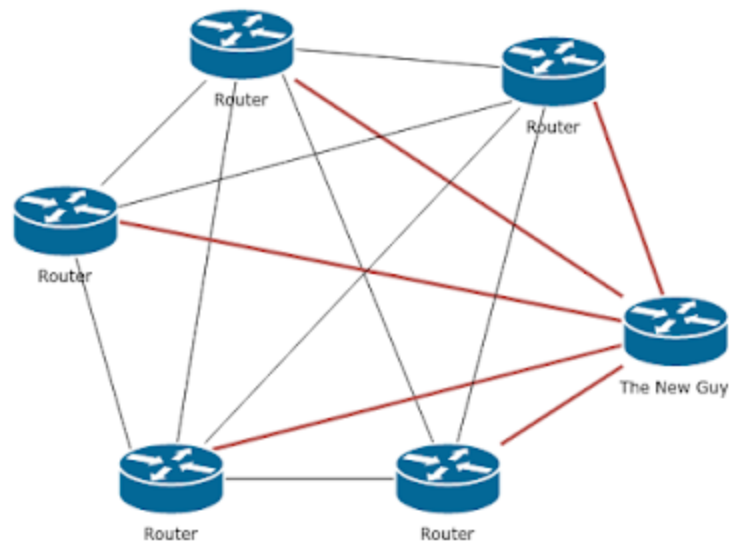April 15, 2018

## CREATING AN INTERNET EXCHANGE FOR EVEN MORE FUN AND LESS PROFIT
—

Last quarter, I was pulled into the slightly odd underground of people running their own autonomous systems, and since then, our circle of friends running autonomous systems at Hurricane Electric's FMT 2 has slowly been growing.

Which is great, except that we're all running autonomous systems, which means that we can set up peering links, and are you *really* friends with another network engineer if you're not running a cross connect between your two networks?

This wasn't too bad for the first few networks joining our little cabal of networks, but due to that pesky quadratic growth issue, the number of new cross connects needed when the fifth or sixth person joined started getting ridiculous. (It's like, four or five!)
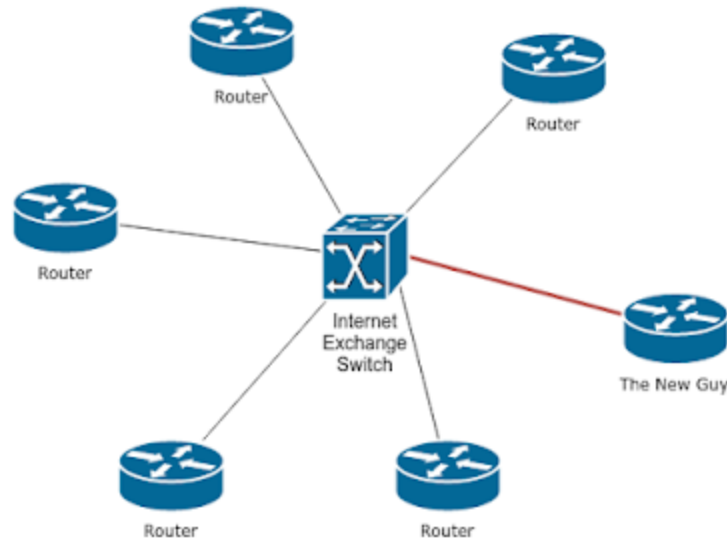
This is, of course, an issue that real networks have to deal with as well, so when we had an eighth friend sign a service agreement with Hurricane Electric this week, the idea was (half jokingly) floated that we should just start our own Internet Exchange Point to cut down on the number of cross connects we need for each new member.

An Internet Exchange is basically just a single L2 Ethernet switch which every network plugs into, such that every network can directly set up BGP peering / route packets to each other network on the fabric. Furthermore, to make it even easier to add new networks to an Internet Exchange, many IXs run "route servers," which are BGP peers which re-distribute all the connected routes. This is convenient because it means that only the IX operator and the new network need to adjust their BGP configuration when a network joins; everyone else is already peered with the route server and start getting the new routes (and which router on the switch to send that traffic to) as part of their already existing connection to the route server.

So we were all sitting there, contemplating the idea of ordering seven more cross connects and once again all logging into our routers to update our configs, and at that point, the idea of creating an Internet Exchange instead didn't seem too bad.

We could instead have all gotten cross connects into one of the existing Internet Exchanges in the HE FMT 2 building, such as SFMIX, but they charge $995/year for a port on their fabric, which is more money than it's worth for all of us to cross connect for amusement's sake (most of us are amateurs and not making money on our networks). So screw it, *hold my other beer*, and away we go!

And that's how the Fremont Cabal Internet Exchange was born.

We even made a website and everything.

We allocated a /64 IPv6 subnet from my /48 (which was originally allocated from another guy's /32), drummed up an IPv4

/24 that was currently between projects, and very carefully selected the private ASN 4244741280, and all that was left to get was a switch to all connect to.
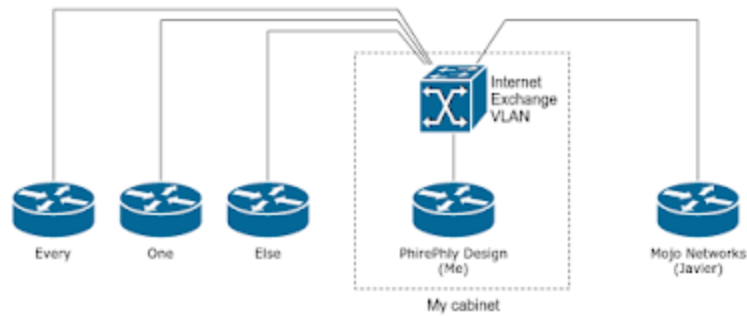

Thankfully, my entire network in my cabinet is built on a Cisco 6506, which is *technically* a switch, so we called that close enough, and instead of having to find another piece of hardware, just allocated a VLAN on my 6506 as the switch fabric, and we were all set. Besides, we were getting a little worried that there were getting to be too few Internet Exchanges running on Cisco 6500s these days.

Now whenever someone wants to connect to the FCIX (Fremont Cabal Internet Exchange) fabric, they just get a cross connect to my cabinet, I set another port to be an access port to the FCIX VLAN, and they're hooked up to everyone.

It's only 1Gbps to each network, but most of us are only originating a few prefixes for a few servers, so we aren't really pushing the limits of single 1G links per participant yet, but just like in any real IX, as soon as someone starts saturating their link to FCIX, they can start setting up direct peering links to other networks to start shedding that traffic off their exchange links. You know... when that happens...

Ideally we would have applied for a public ASN for the exchange, but that $550 + $100/yr for a registered ASN kind of went against the objective of saving money on cross connects, and I figured the chances of someone connecting to FCIX already using one random 4 byte private ASN inside their network was pretty low. Since the IX ASN is never appended to any routes going through the exchange, there's also the fact that no one outside the exchange will ever see this ASN, so it seems like a pretty acceptable trade-off for a group of amateurs for now. ~~(The biggest downside I can think of is that we might not be able to register this IX on~~ peeringDB ~~with a private ASN, to further prop up the facade that this is an Internet Exchange to be taken seriously)~~

Edit: OK, I stand corrected. peeringDB had no problem and we're now live on there as well. That was not expected.
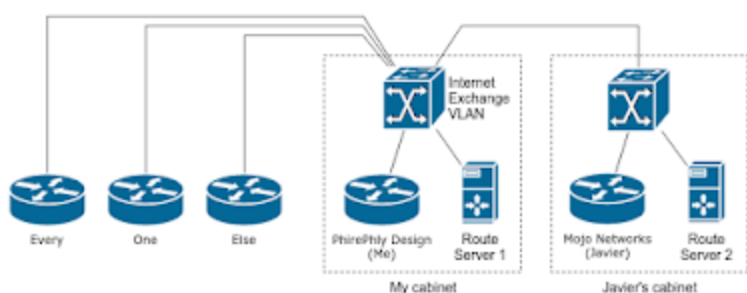
The last piece to really make adding new members to this peering fabric convenient is setting up two route servers, so that each new member doesn't trigger everyone needing to log into their routers to add a new BGP peer. Instead, everyone peers with the route servers and they handle the full N-to-N exchange of routes. When a new member joins, they set up their router on the fabric's /24+/64, and peer with the two route servers, and the only other involvement

needed is from one of the IX admins (which is really just me, currently) to add them to the route server. Every other member doesn't need to be involved and can just enjoy the new routes appearing on their router.
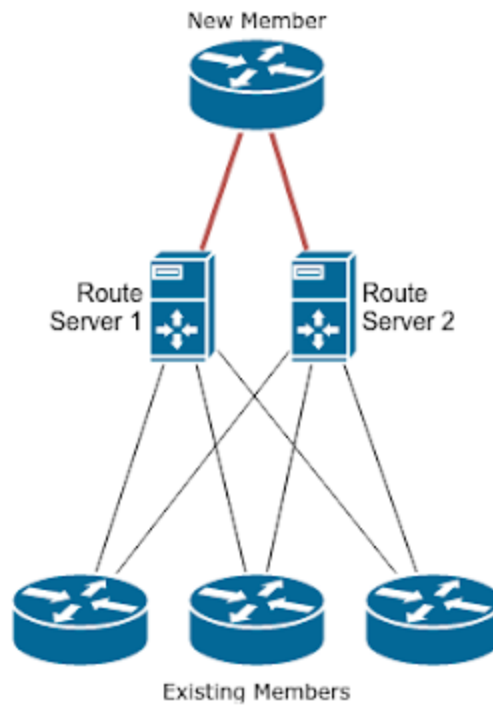
We have two BGP route servers so as I need to restart each one for maintenance reasons, everyone can still trade routes over the other one and I don't trigger a reconvergence every time I restart the daemon or VM. We even managed to get the second VM on a different hypervisor in Javier's cabinet instead of mine, for further fault tolerance.

We're still working to figure out exactly which route server software we want to use. I'm the most familiar with Quagga, but Quagga tries to emulate the Cisco model of all config changes are made on the fly through the console, where I don't want to be hand crafting config changes every time we add a member, so I'm currently taking a crash course in running BIRD as one of our route servers, and will likely be swapping various daemons in for each route server as we learn more.

Share

POPULAR POSTS

CREATING AN AUTONOMOUS SYSTEM FOR FUN AND PROFIT

May 05, 2021

UNLOCKING THIRD PARTY TRANSCEIVERS ON OLDER ARISTA SWITCHES

Share

Kenneth Finnegan

Archive