BGP: The Internet’s Fragile Beast

By Mike Dank for Radical Networks 2019
https://famicoman.com/bgp-radnets2019.odp
@famicoman   @famicoman@mastodon.sdf.org

What We’re Covering

- Who are you?
- What is BGP?
- Some history of the protocol
- How it works!
- What goes wrong?
- How can I play with it?
- Questions!

Who Am I?

- Not a network engineer!
- I do like mesh networks, though
  - https://phillymesh.net
- I also like knowing how the networks around us work
  - https://networksofphilly.org

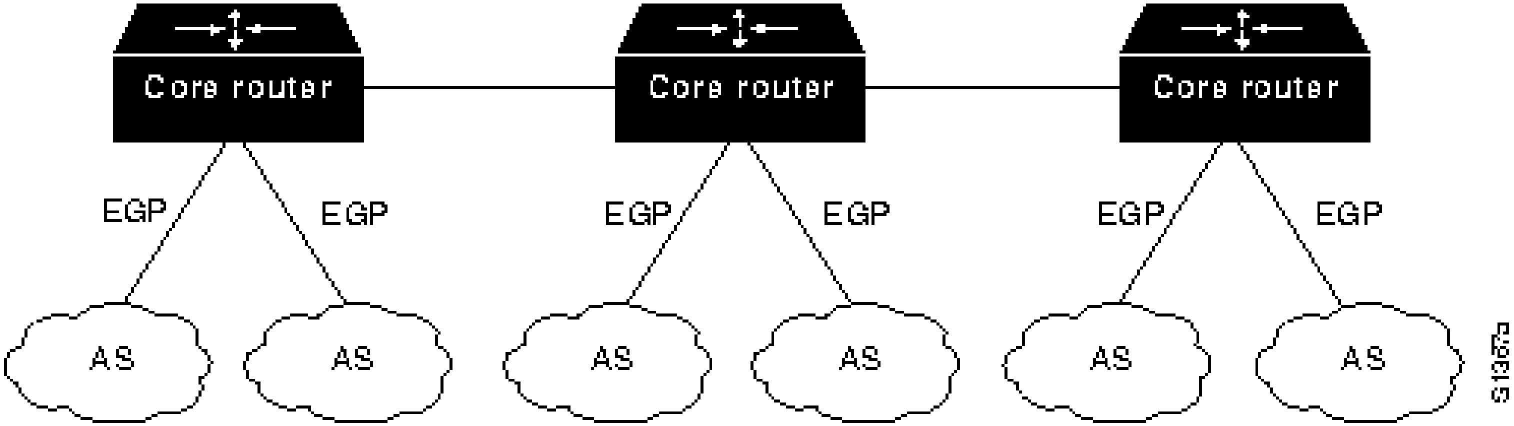
What is BGP?

- **BGP stands for Border Gateway Protocol**
  - It's the protocol that makes the Internet work!
    - It facilitates the routing of IP packets with routing tables!
  - Think about it like the postal system
    - You need to send a letter to a friend
    - You drop the letter in the mailbox
    - The postal service picks the best route for the letter
    - The postal service uses that route to deliver the letter quickly and efficiently.
  - This is a *best-effort* protocol

State of the Internet in 1989

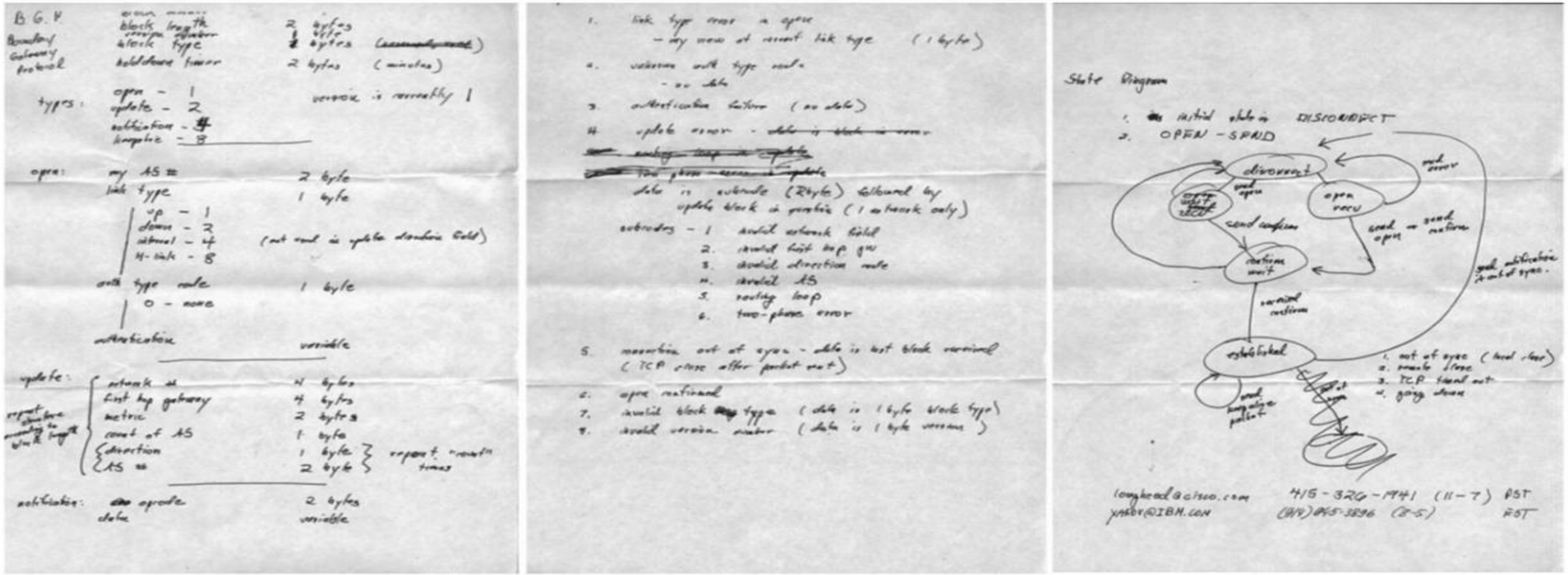
- NSFNET (National Science Foundation Network) is doing very well!
- The ARPANET is about to be shut down
- The existing routing protocol, Exterior Gateway Protocol (EGP), has problems[0]
  - The Internet is growing at an exponential rate
  - Centralized topology
  - Routing table updates are too large for maximum transport size

EGP Topology[2]



BGP - A Two-Napkin Protocol

- Kirk Lougheed of Cisco and Yakov Rekhter of IBM[1]



B.G.P.
Boundary Gateway Protocol

~~[illegible]~~
block length 2 bytes
~~version number~~ 1 byte
block type ~~2~~ 1 bytes (~~[illegible]~~)
holddown timer 2 bytes (minutes)

types: open - 1 version is currently 1
update - 2
notification - ~~3~~ 4
keepalive - 8

open: my AS # 2 byte
link type 1 byte
up - 1
down - 2
internal - 4 (not used in update direction field)
H-link - 8
auth type code 1 byte
0 - none
authentication variable

update: network # 4 bytes
first hop gateway 4 bytes
metric 2 bytes
count of AS 1 byte
{direction 1 byte}
{AS # 2 byte} repeat "count" times
repeat structure according to block length

notification: ~~[illegible]~~ opcode 2 bytes
data variable

1. link type error in open
   - my view of recent link type (1 byte)
2. unknown auth type code
   - no data
3. authentication failure (no data)
4. update error - ~~data is block in error~~
~~— routing loop in update~~
~~— two phase error in update~~
data is subcode (2byte) followed by update block in question (1 network only)
subcodes - 1 invalid network field
2. invalid first hop gw
3. invalid direction code
4. invalid AS
5. routing loop
6. two-phase error
5. connection out of sync - data is last block received (TCP close after packet sent)
6. open refused
7. invalid block ~~[illegible]~~ type (data is 1 byte block type)
8. invalid version number (data is 1 byte version)

State Diagram

1. ~~[illegible]~~ initial state is DISCONNECT
2. OPEN - SEND

1. out of sync (local close)
2. remote close
3. TCP timed out
4. going down

lougheed@cisco.com 415-326-1941 (11-7) PST
yakov@IBM.COM (914)945-3896 (8-5) EST

BGP is Born

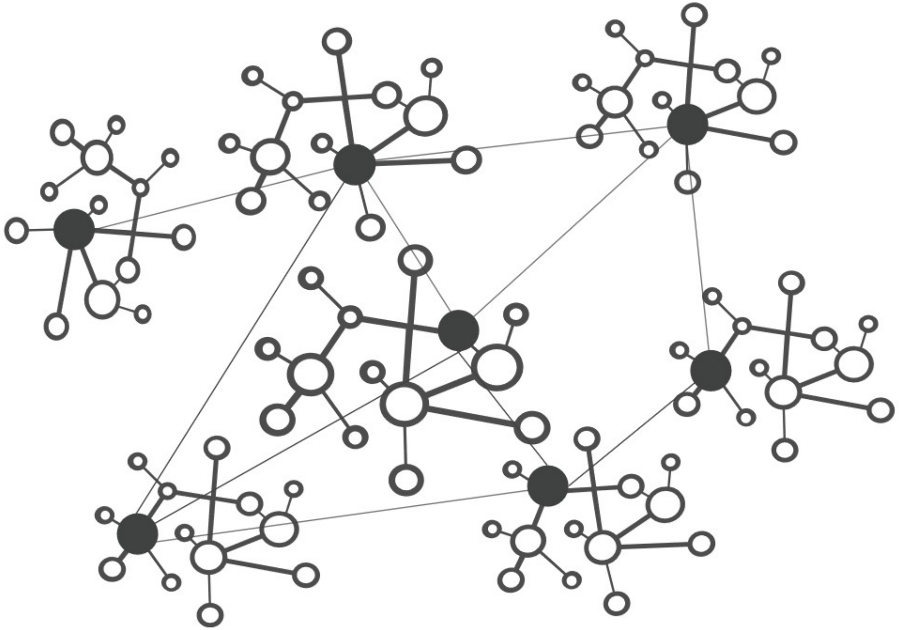
- RFC 1105 introduced in 1989[11]
  - At this time, protocol changes were done voluntarily. Working software prevailed!
- BGP works on top of TCP
  - Sessions created on TCP port 179
- We currently use BGP-4 (2006)

Advantages of BGP

- Mesh topology, connect many Autonomous Systems (independent networks)
- “Best path” algorithm (path vector routing)
    - Routers advertise their network routes
    - Routers can choose to not route through different networks
- Scalable and flexible
- Handles route “flapping” (unstable links that go down) via dampening

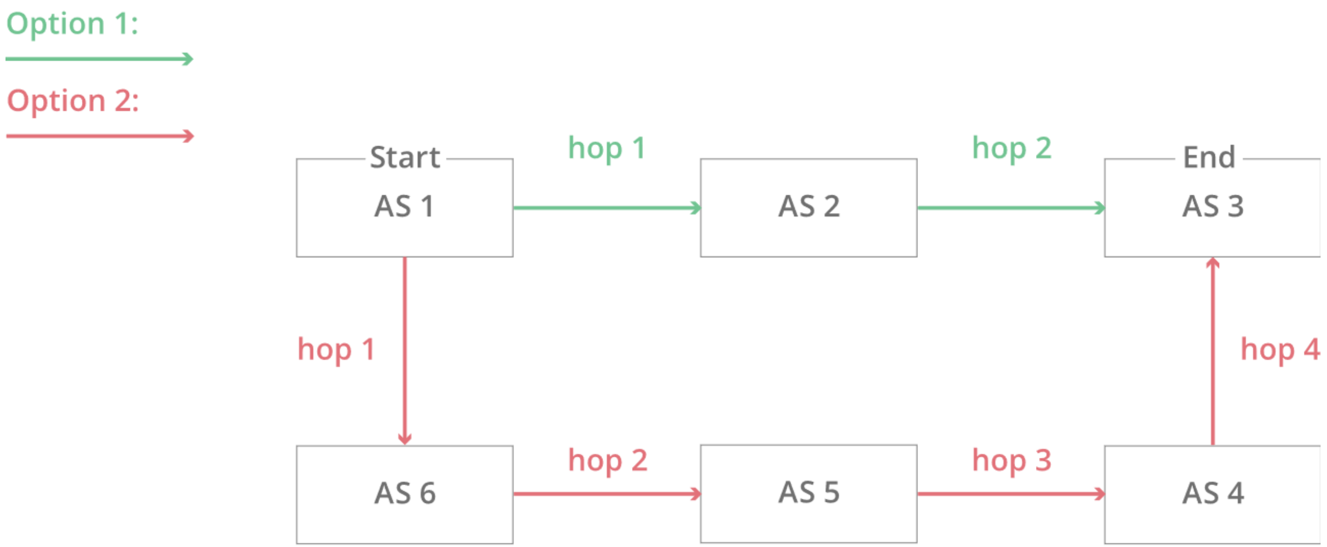
BGP Topology

- A network of networks[7]

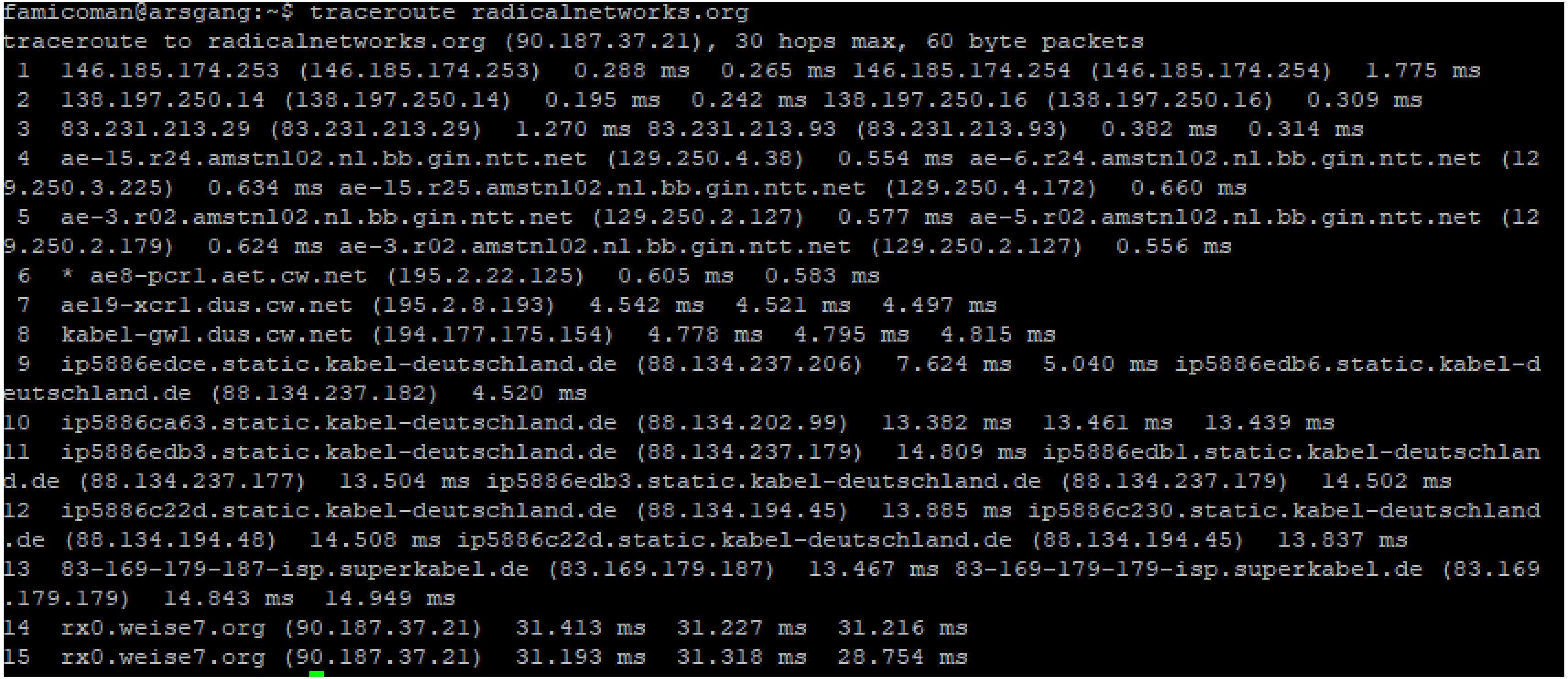


How Data Flows Through Networks

- Let’s go from AS 1 to AS 3[7]



You Can See Where Your Traffic Goes!



```
famicoman@arsgang:~$ traceroute radicalnetworks.org
traceroute to radicalnetworks.org (90.187.37.21), 30 hops max, 60 byte packets
 1  146.185.174.253 (146.185.174.253)  0.288 ms  0.265 ms 146.185.174.254 (146.185.174.254)  1.775 ms
 2  138.197.250.14 (138.197.250.14)  0.195 ms  0.242 ms 138.197.250.16 (138.197.250.16)  0.309 ms
 3  83.231.213.29 (83.231.213.29)  1.270 ms 83.231.213.93 (83.231.213.93)  0.382 ms  0.314 ms
 4  ae-15.r24.amstnl02.nl.bb.gin.ntt.net (129.250.4.38)  0.554 ms ae-6.r24.amstnl02.nl.bb.gin.ntt.net (12
9.250.3.225)  0.634 ms ae-15.r25.amstnl02.nl.bb.gin.ntt.net (129.250.4.172)  0.660 ms
 5  ae-3.r02.amstnl02.nl.bb.gin.ntt.net (129.250.2.127)  0.577 ms ae-5.r02.amstnl02.nl.bb.gin.ntt.net (12
9.250.2.179)  0.624 ms ae-3.r02.amstnl02.nl.bb.gin.ntt.net (129.250.2.127)  0.556 ms
 6  * ae8-pcr1.aet.cw.net (195.2.22.125)  0.605 ms  0.583 ms
 7  ae19-xcr1.dus.cw.net (195.2.8.193)  4.542 ms  4.521 ms  4.497 ms
 8  kabel-gw1.dus.cw.net (194.177.175.154)  4.778 ms  4.795 ms  4.815 ms
 9  ip5886edce.static.kabel-deutschland.de (88.134.237.206)  7.624 ms  5.040 ms ip5886edb6.static.kabel-d
eutschland.de (88.134.237.182)  4.520 ms
10  ip5886ca63.static.kabel-deutschland.de (88.134.202.99)  13.382 ms  13.461 ms  13.439 ms
11  ip5886edb3.static.kabel-deutschland.de (88.134.237.179)  14.809 ms ip5886edb1.static.kabel-deutschlan
d.de (88.134.237.177)  13.504 ms ip5886edb3.static.kabel-deutschland.de (88.134.237.179)  14.502 ms
12  ip5886c22d.static.kabel-deutschland.de (88.134.194.45)  13.885 ms ip5886c230.static.kabel-deutschland
.de (88.134.194.48)  14.508 ms ip5886c22d.static.kabel-deutschland.de (88.134.194.45)  13.837 ms
13  83-169-179-187-isp.superkabel.de (83.169.179.187)  13.467 ms 83-169-179-179-isp.superkabel.de (83.169
.179.179)  14.843 ms  14.949 ms
14  rx0.weise7.org (90.187.37.21)  31.413 ms  31.227 ms  31.216 ms
15  rx0.weise7.org (90.187.37.21)  31.193 ms  31.318 ms  28.754 ms
```

What Do I Need to Get on the Internet?

- **Find your IANA Regional Internet Registry: AFRINIC, ARIN, APNIC, LACNIC or RIPE NCC**
- **IP Addresses!**
  - IPv4 – A */24 (256 Addresses, xxx.xxx.xxx.0 – xxx.xxx.xxx-255)*
    - *$25/*address, $6,425 Total Upfront[4]
  - IPv6- A /48 (1,208,925,819,614,629,174,706,176 Addresses, xxxx:xxxx:xxxx:0000:0000:0000:0000:0000 - xxxx:xxxx:xxxx:ffff:ffff:ffff:ffff:ffff)
    - $250 TOTAL Upfront[5]
- **Autonomous System Number (ASN) (with info for two other networks agreeing to peer with you)**
  - Looks like AS####
    - $550 TOTAL Upfront[5]
- **Total Upfront Costs = $7,225, Total Annual Recurring Costs = $350[5]**

Find a Physical Location for the Internet

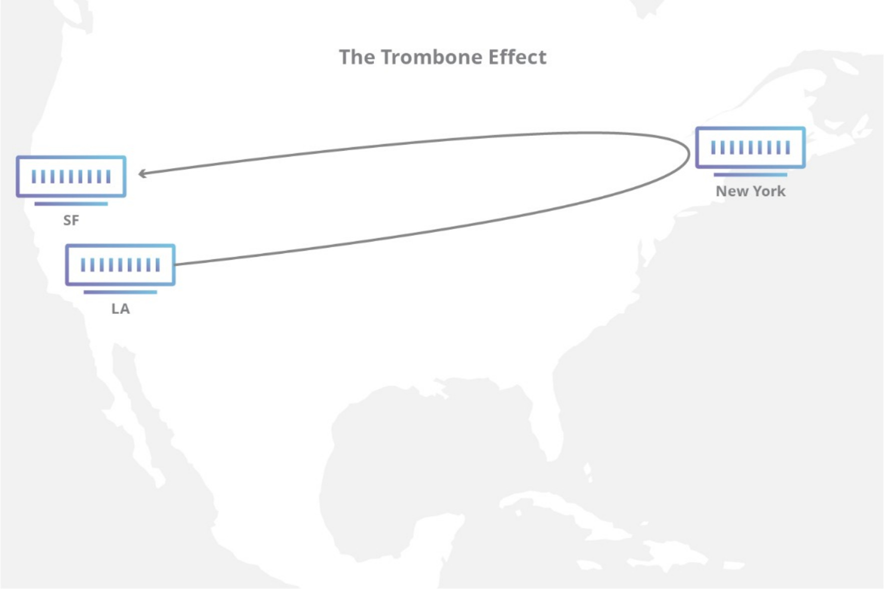
- IXPs (Internet eXchange Points) and Carrier Hotels
  - Building where many networks have a physical “edge”
    - PoPs (Point of Presence)
  - Facilitate links between networks to let data flow on the Internet
  - Robust buildings, built to last, often fireproof
  - Critical to keeping the Internet operating
  - Example: 60 Hudson, NYC[6]

The Internet is HALF A BLOCK AWAY FROM YOU

- 811 10th Avenue, NYC
- AT&T backbone network site
  - *Networks connect here!*
- Named in The Intercept's 2018 article on NSA spy hubs[17]
- AT&T transferred colocation assets and operations to Evoque in January 2019[18]

Why are IXPs Important?

- Backbone ISPs can sometimes route traffic through distant locations[8]



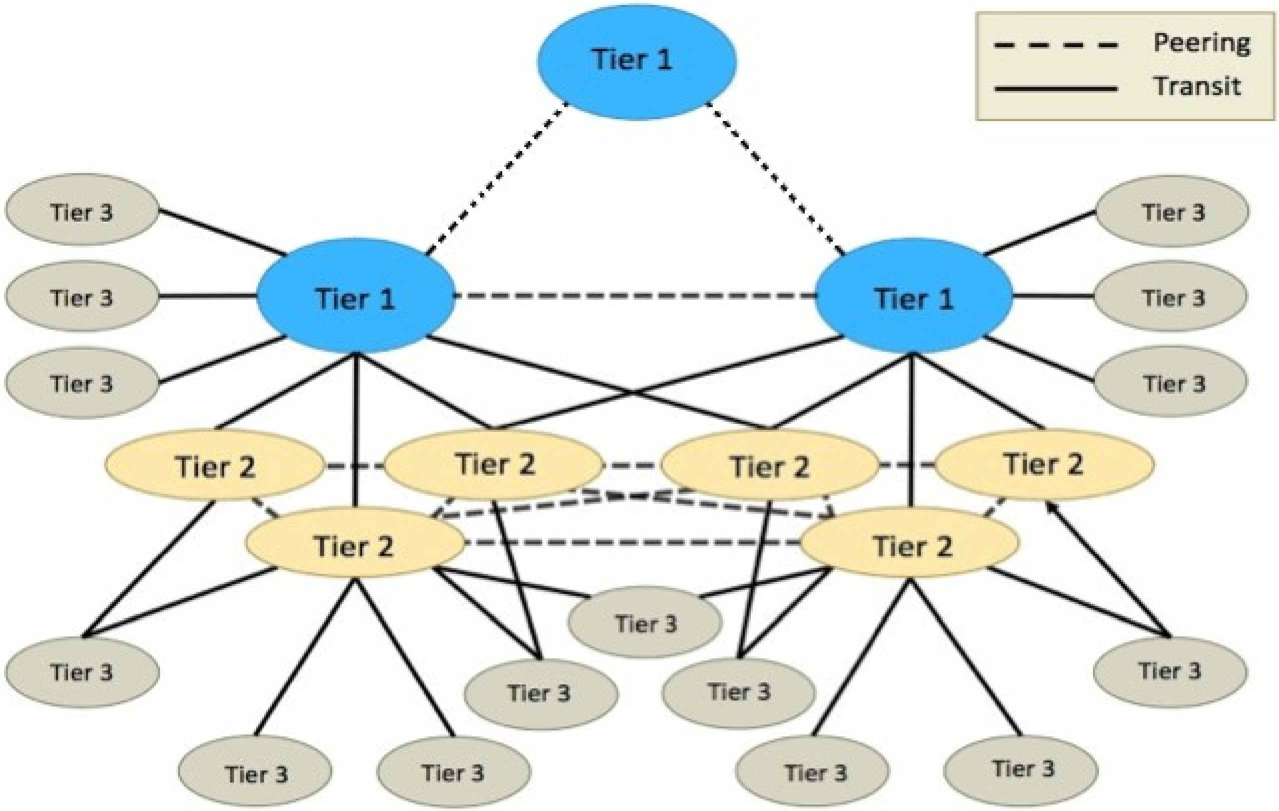
How Networks Connect

- Peering vs Upstream Transit
- Networks in data centers can connect with a layer 2 network, much like your home network (but with much faster speeds and bigger pipes)
- AS routers run BGP, and are generally Linux/BSD boxes or dedicated network gear (Cisco, etc.)
- Networks negotiate a connection deal. Free peering links are common, and mutually beneficial, but *upstream will almost always cost something*
- Networks announce routes to one another. You announce your IP range(s) to a peer, while they announce range(s) back.

The Tiered Internet

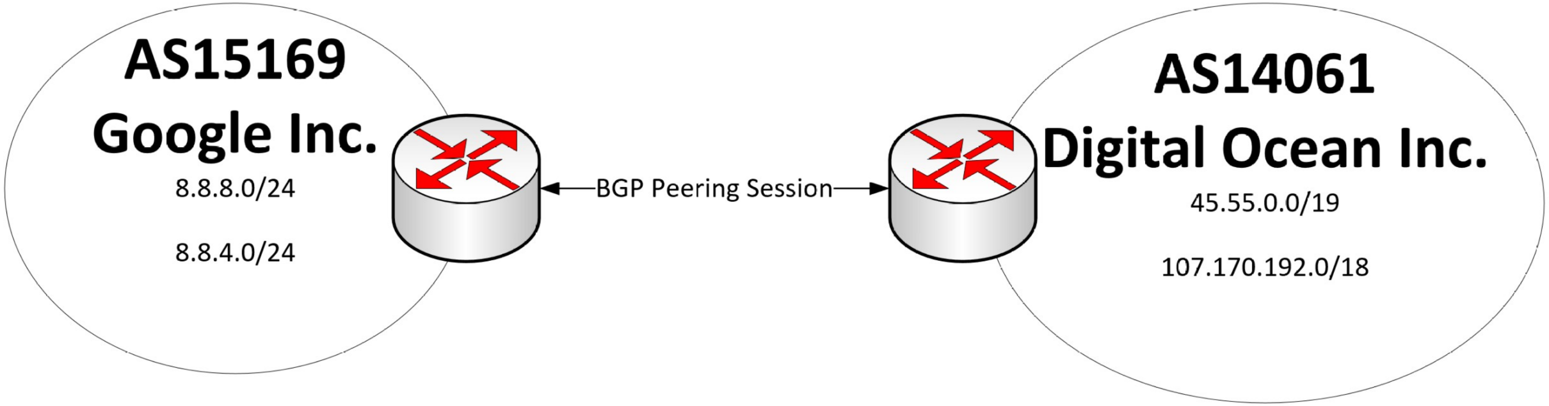
- **Tier 1 networks make the backbone of the internet**
  - Examples: AT&T, Sprint, Verizon, Century Link (Level 3), etc.
- **Tier 2 networks are large ISPs, usually purchase transit**
  - Examples: Cogent, Comcast, Hurricane Electric
- **Tier 3 networks are last mile ISPs, solely purchase transit**
  - Examples: Small ISPs, businesses, schools

Connecting the Tiers[16]



What Does Peering Look Like?

- Basic peering between two AS[9]



BGP Operation

- **Path Attributes**
  - Shortest AS path “wins”
  - Filtering to prefer certain neighbors, use different routes for different sources (internal traffic vs external), routes based on aggregating traffic together, etc.

BGP Security

- **BGP has few security precautions**
  - Most operators don’t configure anything for security!
- **What could go wrong?**
  - Route leak
    - Content of the BGP table is maliciously/accidentally altered, traffic can’t reach its destination
  - Route hijacking
    - Bad actor announces a victim’s prefix, rerouting target traffic to itself
  - Denial-of-service (DoS)
    - Bad actor sends undesirable BGP traffic to a victim, exhausting resources

“[Security] wasn’t even on the table.”[3] - Yakov Rekhter, Inventor of BGP

> **“There was no concept that people would use this to do malicious things. . . . Security was not a big issue.” - Kirk Lougheed, Inventor of BGP**

Some BGP Incidents

- April 1997 - AS 7007 incident, ISP in Virgina leaks routing table, blackholes the Internet
- May 1998 - L0pht testify before Congress, can “bring down the whole Internet in 30 minutes”
- February 2008 – Pakistan attempts to block YouTube
- April 2010 – Chinese ISP Hijacks Internet
- February 2014 – Canadian ISP Hijacked to steal bitcoin
- April 2017 – Russian Rostelecom originates 37 prefixes for Visa, Mastercard, etc.
- July 2018 - Iran Telecommunication Company originated prefixes of Telegram Messenger
- November 2018 - China Telecom site originated Google addresses
- June 2019 - Large European mobile traffic was rerouted through China Telecom
- June 2019 – Verizon advertises misconfigured routes from Allegheny Technologies

Pakistan Attempts to Block Youtube

- **February 24, 2008, Pakistan’s state-owned telecom attempted to block YouTube**
- **Accidentally announced 256 addresses in YouTube's 208.65.153.0 network space (hole-punching)**[21]
  - Hong Kong-based PCCW (Pakistan’s uplink) did not stop broadcasting the range
  - In 15 seconds, large Pacific-rim providers directed YouTube.com traffic to Pakistan ISP, in 45 seconds routers in the rest of the Internet to follow suit[21]
  - Availability for YouTube dropped to 0 in an hour, took two hours to correct[21]
  - YouTube countered in minutes, advertising 64-address ranges[21]

Canadian ISP Hijacked to Steal Bitcoin

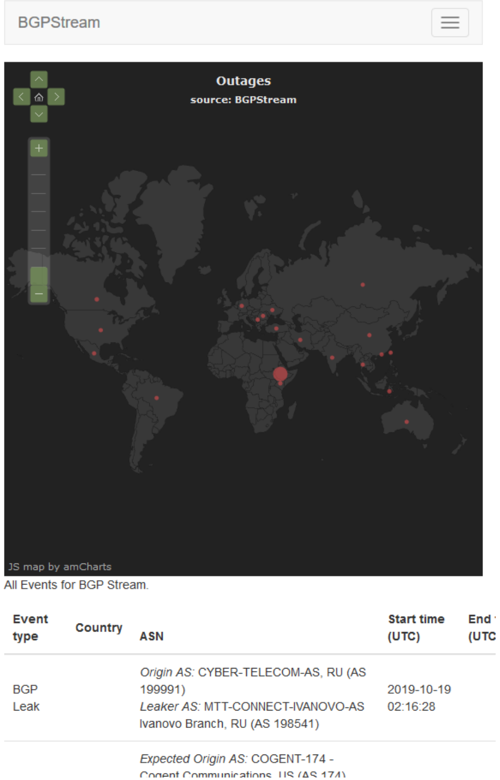
- **Between February and May 2014, a hacker used a Canadian ISP to announce addresses for a known Bitcoin mining pool**
- **Hacker changed config on ISPs router for 30 seconds at a time, 22 times total within the 3 month period[23]**
  - At least 51 different networks were compromised including Amazon,DigitalOcean, OVH, and 19 ISPs[22][23]
  - Address of Bitcoin pool server was redirected to a machine under the hacker’s control (running its own pool software)
  - Hacker was able to hijack mining pool to cash out $83,000[23]

European Mobile Traffic Routed Through China

- On June 6, 2019 Swiss data center colocation company Safe Host, accidentally leaked over 70,000 routes from internal routing tables to China Telecom[24]
- China Telecom re-announced Safe Host's routes, interposing itself as one of the shortest ways to reach Safe Host's network and other nearby European telcos and ISPs[24]
  - Mobile data from France, Holland, Switzerland was routed through China
  - Slow connection speeds for users
  - Route leak continued for 2 hours before being corrected
  - It is speculated that the Chinese government used this event for information gathering
    - Users don’t even know their data went through a different network!

BGP Incidents Happen Everyday!

- Cisco's BGPStream
  - Real-time monitoring for BGP changes
  - https://bgpstream.com/
- On 10/15 (last Tuesday) there were…
  - 15 outages
  - 3 possible hijacks
  - 2 route leaks



All Events for BGP Stream.

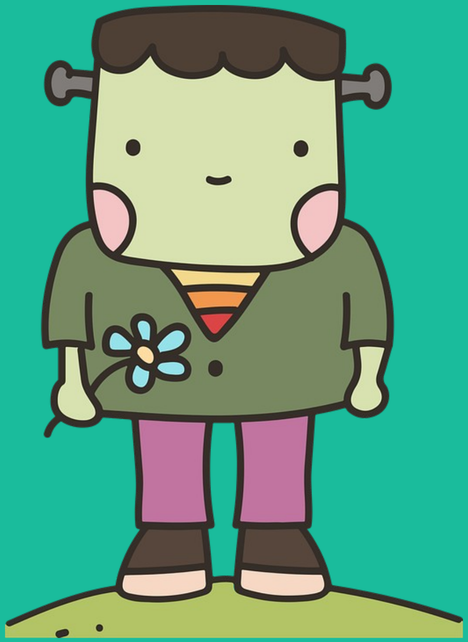
Event type	Country	ASN	Start time (UTC)	End (UTC
BGP Leak		*Origin AS:* CYBER-TELECOM-AS, RU (AS 199991) *Leaker AS:* MTT-CONNECT-IVANOVO-AS Ivanovo Branch, RU (AS 198541)	2019-10-19 02:16:28	
		Expected Origin AS: COGENT-174 - Cogent Communications, US (AS 174)		

How Can BGP Be Secured?

- **NIST's "proof-of-concept demonstration"**
  - Route Origin Validation (ROV) using Public Key Infrastructure verify routes are announced by proper AS. BGPSec has routers signing routes, creating a trusted chain[12]
    - RFC 6810 in 2013[13]
    - RFC 8210 in 2017[14]
    - RFC 8206 in 2017[15]
  - As of August 2019, there are 92,000 unique ASNs, **currently 84** use Route Origin Validation[19]
- **BGP Operations and Security, RFC 7454 (2015)**[20]
  - Like the missing BGP security manual, how to appropriately filter, TCP authentication settings, and more.

How You Can Play with BGP

- **AMPRNet aka 44Net - https://www.ampr.org**
  - Experimental network for Ham radio operators, free to use!
  - Can get a /24 (256 addresses)
- **DN42 - https://dn42.eu**
  - BGP test network, uses private ranges
  - Many amateur sysops
- **router.city - https://router.city**
  - BGP test network I helped build
  - Framework for others to easily setup their own BGP testnet



Questions?

Mike Dank

https://famicoman.com/bgp-radnets2019.odp

@famicoman @famicoman@mastodon.sdf.org

Thank you!

Sources

- Title Slide - https://pixabay.com/vectors/monster-hairy-halloween-creature-3764868/
- [0] - https://computerhistory.org/blog/the-two-napkin-protocol/?key=the-two-napkin-protocol
- [1] - https://www.stuff.co.nz/technology/digital-living/69048160/
- [2] - http://wwwlehre.dhbw-stuttgart.de/~schulte/htme/55146.htm
- [3] - https://www.washingtonpost.com/sf/business/2015/05/31/net-of-insecurity-part-2/?noredirect=on
- [4] - https://www.ipv4connect.com/
- [5] - https://www.arin.net/resources/fees/fee_schedule/#registration-services-plan
- [6] - https://en.wikipedia.org/wiki/60_Hudson_Street#/media/File:Western_Union_building,_Manhattan_jeh_crop.jpg
- [7] - https://www.cloudflare.com/learning/security/glossary/what-is-bgp/
- [8] - https://www.cloudflare.com/learning/cdn/glossary/internet-exchange-point-ixp/
- [9] - https://blog.cdemi.io/beginners-guide-to-understanding-bgp/
- [10] - https://datapacket.com/blog/bgp-network-how-does-it-work/
- [11] - https://tools.ietf.org/html/rfc1105
- [12] - https://duo.com/decipher/nist-outlines-how-to-secure-bgp

Sources (Continued)

- [13] - https://tools.ietf.org/html/rfc6810
- [14] - https://tools.ietf.org/html/rfc8210
- [15] - https://tools.ietf.org/html/rfc8206
- [16] - https://orhanergun.net/2017/01/tier-1-tier-2-tier-3-service-providers/
- [17] - https://theintercept.com/2018/06/25/att-internet-nsa-spy-hubs/
- [18] - https://about.att.com/story/2018/att_data_center_colocation_operations_assets.html
- [19] - https://rov.rpki.net/
- [20] - https://tools.ietf.org/html/rfc7454
- [21] - https://www.cnet.com/news/how-pakistan-knocked-youtube-offline-and-how-to-make-sure-it-never-happens-again/
- [22] - https://www.zdnet.com/article/hacker-hijacks-isps-steals-83000-from-bitcoin-mining-pools/
- [23] - https://www.wired.com/2014/08/isp-bitcoin-theft/
- [24] - https://www.zdnet.com/article/for-two-hours-a-large-chunk-of-european-mobile-traffic-was-rerouted-through-china/

CC0 – No Rights Reserved!

https://creativecommons.org/share-your-work/public-domain/cc0/