Welcome to the Internet

BGP, ASNs and decentralisation

https://www.youtube.com/watch?v=k1BneeJTDcU
Welcome to the Internet - Bo Burnham (from “Inside")

Plan B Academy - Feb. 26th 2026 - Michel ‘ketominer’ L.

What I do

nodl

VPS - Domain Names

```
 __               __   _____              __
|  |--.-----.-----|  |_|  |  |.----.-----.|__|.-----.-----.
|     |  _  |__ --|   _|__    |  __|  _  ||  ||     |__ --|
|__|__|_____|_____|____|  |__||____|_____||__||__|__|_____|
```

NoKYC - BTC Only

PRECISION
LABS

Infrastructure • Security • Space Systems

Agenda

- The Origins
- Networking Fundamentals
- Deep Dive
- Trust & Security
- Implementation

Why should you care?

Bitcoin runs on top of the Internet

- A system built on top of other systems requires understanding the underlying systems
- Network level attacks and errors are real
  - MyEtherWallet DNS hijack via AWS BGP hijack (2018)
  - Telekom Malaysia misconfiguration propagated by Level3
  - Pakistan Telecom blocks Youtube, takes down parts of Internet instead
- There is a lot of misunderstanding about the decentralization of the Internet

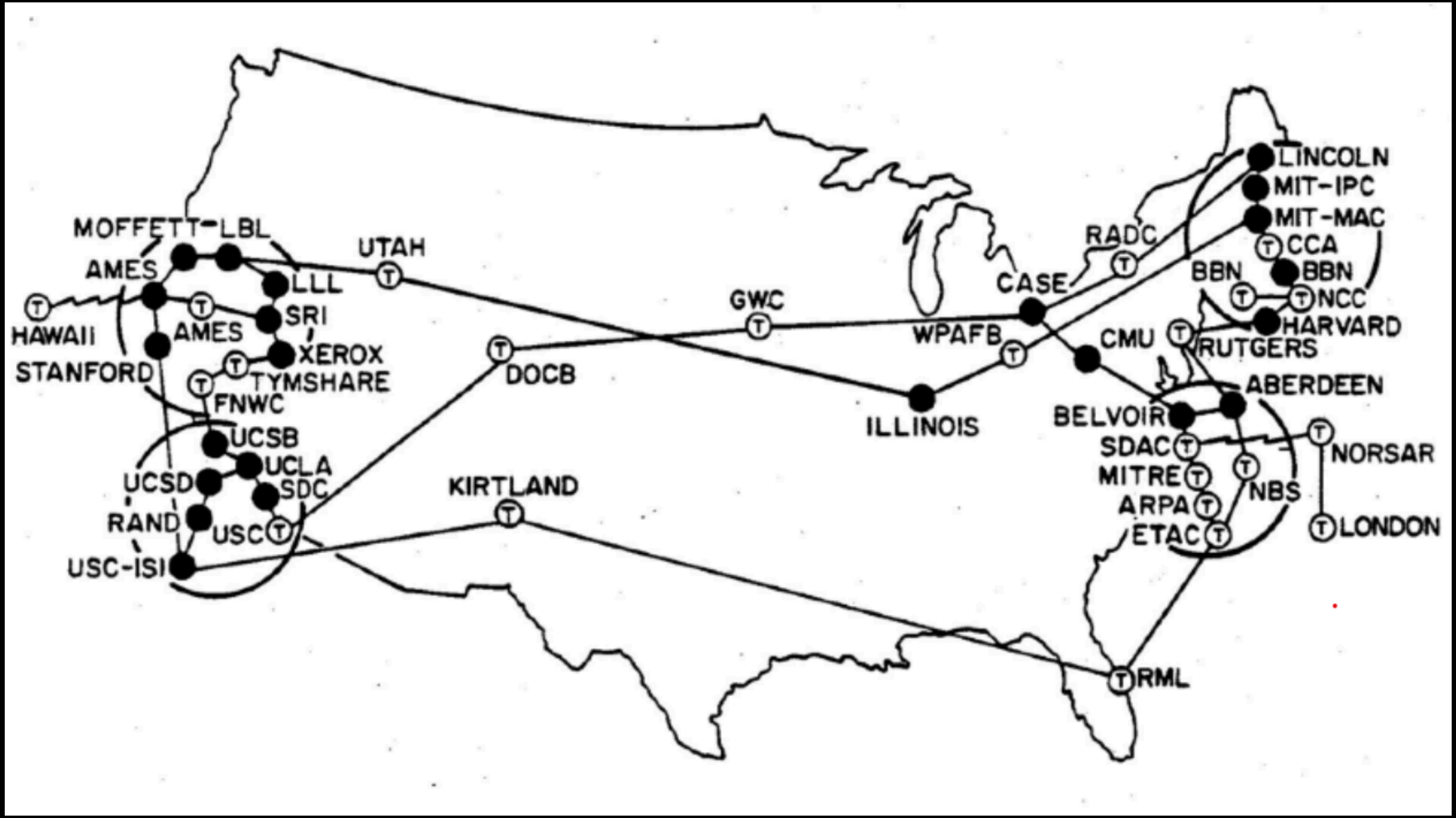
So is this thing decentralized?

- The “web services” are concentrated (and getting more so)
  - “big cloud” (AWS, GCP, Azure, …)
  - SaaS (shopify, …)
- But don’t have to
  - many other “cloud providers” / VPS / dedicated servers - linode/hetzner/infomaniak/ovh/scaleway/…
  - you can still self host many things at home without Web3 - static IP helps but there are other ways
- The network is totally decentralized
  - 100.000+ interconnected networks of all shapes and sizes

=> The infrastructure is totally decentralized. Your use of it, much less.

Chapter One: The Origins

How it started



Key nodes:
UCLA - Stanford - MIT - Harvard
Military: ARPA, RAND, RCS

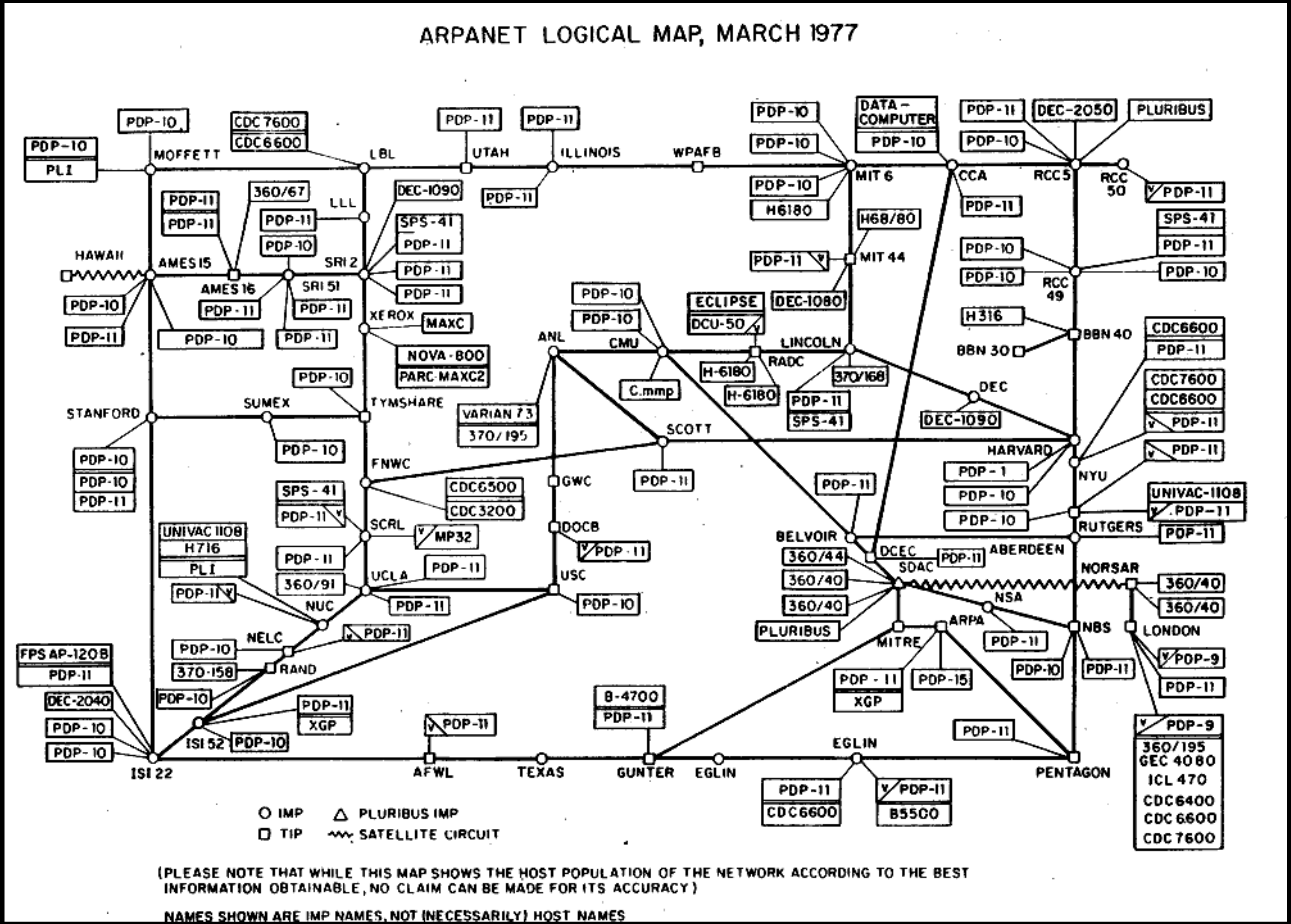
First international link UK and Norway

MILNET quickly split up (1984)

56k was a “high speed” link

Physical map of ARPANET - 1970s

How it started



How it started - what made it possible

- January 1st 1983: whole network switches to TCP/IP “flag day” (transition plan https://www.rfc-editor.org/rfc/rfc801.txt)
- 1986: NSFNET connects major universities with a 56kbps backbone
- 1988: finished upgrade to T-1 (1.5 Mbps)
- 1991-1995: progressively open to commercial traffic and phase-out of the old NSFNET backbone
- 1995: BGP4 / RFC1771
- Now: 100 000+ independent networks interconnected

Main concept

Circuits vs Packets vs Nuclear war

Circuit switched network

- One communication = one circuit (path)
- Resources reserved end to end - even if no traffic!
- One link breaks = connection lost

Packet switched network

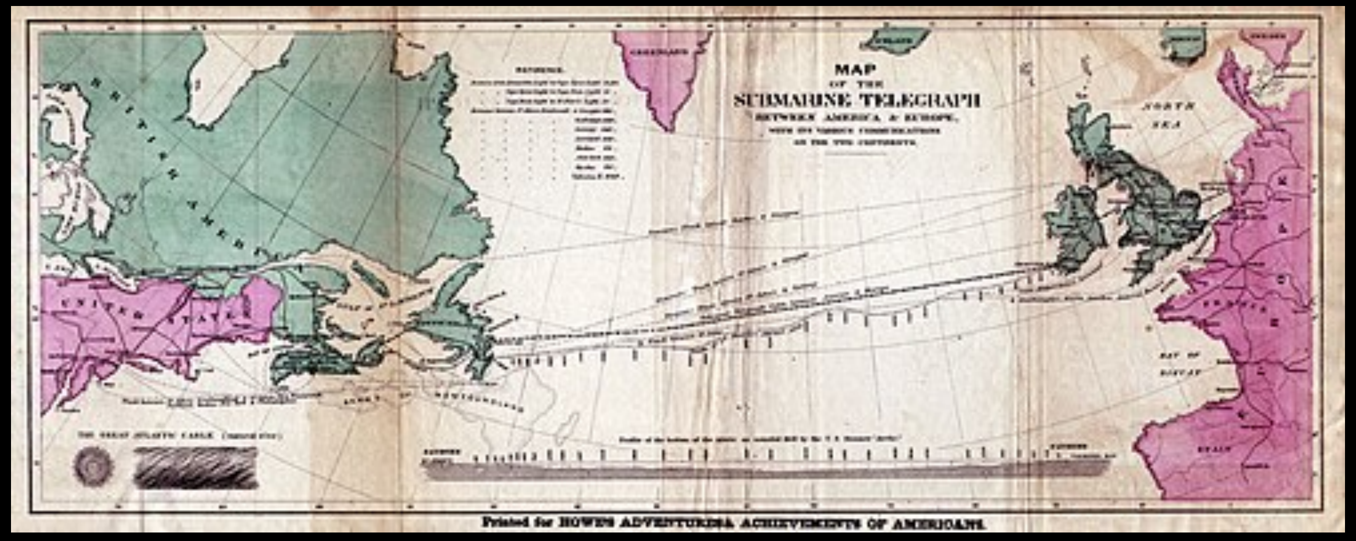
- Each packet (can) follows a different path - packets can arrive out of order
- Resources reserved hop by hop - can lead to congestions (no way to know capacity of next hop)
- Packets reroute around failures - can take time to “converge”

The Internet is a packet switched network (TCP/IP)

It was built to survive a partial destruction (nuclear war)

Higher level protocols exist to provide circuit switched network reliability on top of the packet switched Internet (SR, MPLS, …)

Satellites!



Many people believe that overseas communications happen over satellites.

If we were in 1971-73 this would mostly be true (for the Internet).

Telegraph and then phone communications are undersea cable based since 1858

2024: ~550 submarine optical cables - 1.4+ million km

Capacity: 100-1000 Tbps per cable (satellites = Gbps scale)

Latency: ~5ms per 1000km - Transatlantic = 5500-7500km = 55-75ms RTT (GEO satellites = 600-800ms RTT)

Cost per bandwidth (and latency) makes satellites impractical for long distance communications

Owned by consortiums, telcos, a very few by GAFAMs

Explore Telegeography's https://submarinecablemap.com to see every cable, every landing point, and owner

But Starlink?

Starlink is not “part of the Internet” (of its backbone). It’s a leaf (actually several).

- Buys “transit” from Tier-1/2 providers - in many locations they’re reselling connectivity from 3rd party ISPs
- Ground stations connect to terrestrial Internet - for most users, located in the country they subscribed
- Satellites = wireless last mile (like cell towers)
- NOT a backbone - traffic still goes through cables

- Satellite to satellite comms doesn’t make it part of the backbone, internal use only and still has to go out through “local” spaceport

ISP - Transit - Peering - Tiers

English “ISP” is too generic. What most people call ISP is what French call FAI (Fournisseur d’Accès à Internet) = Internet **Access** Provider = what you have at home

Internet **Service** provider regroups anything running a network part of the Internet (hosting, transit, tier-1, tier-2, etc.)

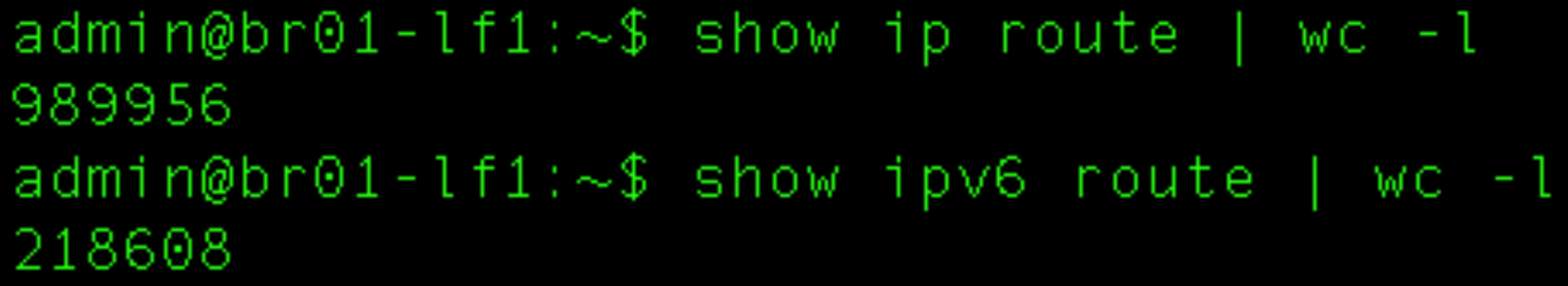
More specific vocabulary:

- Peering: connection between two ISPs (in this context, two ASNs)
- Transit: specific type of peering providing a “full view” of the Internet (more on this later)
- Tier-1: network that sees the whole Internet through peerings - typically involves a network covering the whole planet (there is a handful of those)
- Tier-2: network that sees close to the whole Internet through peerings but has to buy some connectivity to complete the view
- All others: networks that see the whole Internet by buying one or several Transits and has zero to several Peerings
- IXP (Internet eXchange Point): huge switch (or network of switches) interconnecting several networks

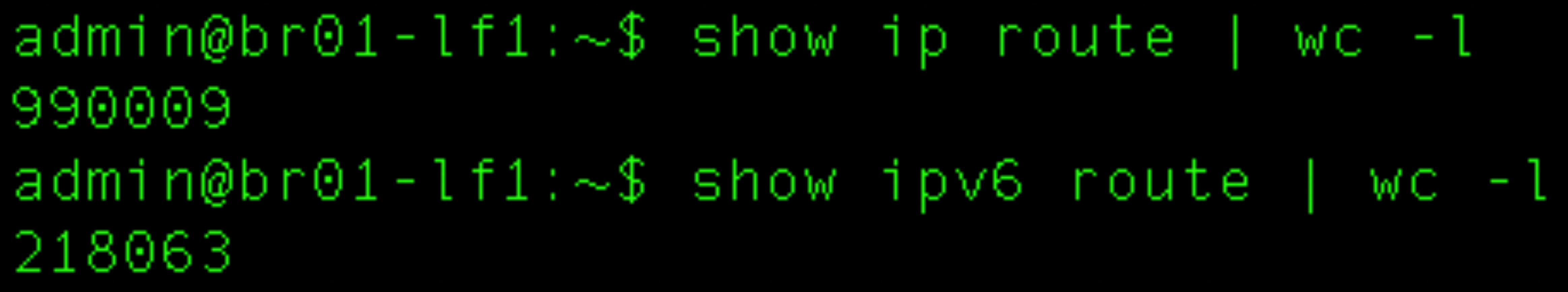
How it’s going

There is no single view

- https://as2914.net/ (as seen by NTT)
- https://www.pacnog.org/pacnog17/presentations/MappingTheInternet.pdf (as mapped by the NSRC and the University of Oregon)
- As seen by one of my routers (2 days apart):

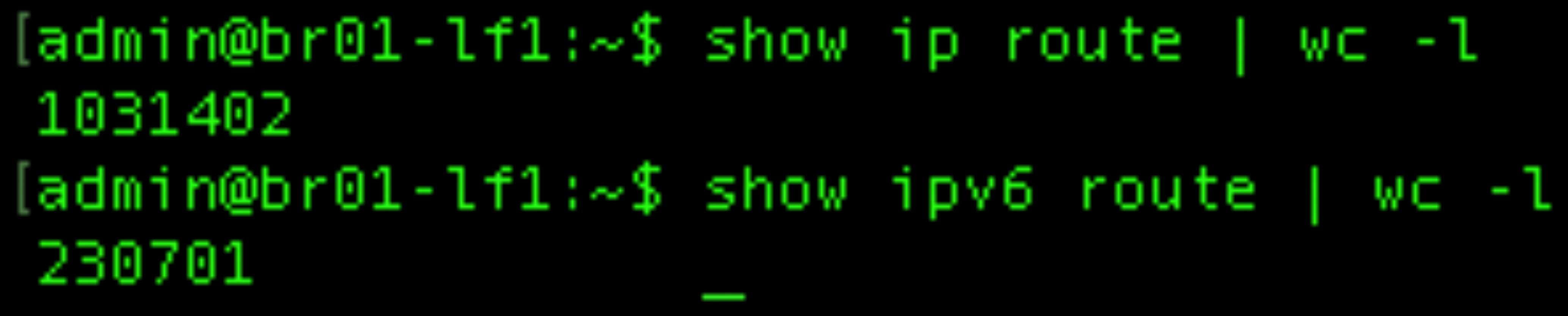


```
admin@br01-lf1:~$ show ip route | wc -l
989956
admin@br01-lf1:~$ show ipv6 route | wc -l
218608
```



```
admin@br01-lf1:~$ show ip route | wc -l
990009
admin@br01-lf1:~$ show ipv6 route | wc -l
218063
```

18 months later:



```
[admin@br01-lf1:~$ show ip route | wc -l
1031402
[admin@br01-lf1:~$ show ipv6 route | wc -l
230701
```

Chapter Two: Networking Fundamentals

Hub vs Switch vs Router

- Hub (mostly disappeared): receives packet on one port, sends out on all ports - collisions!

- Switch: receives packet on one port, looks up MAC address, sends out on port connected to MAC address (or none) - Layer 2 “routing”

- Router: receives packet on one port, looks up destination subnet, sends out to next router (“next hop”) - Layer 3 routing - what you see in a traceroute

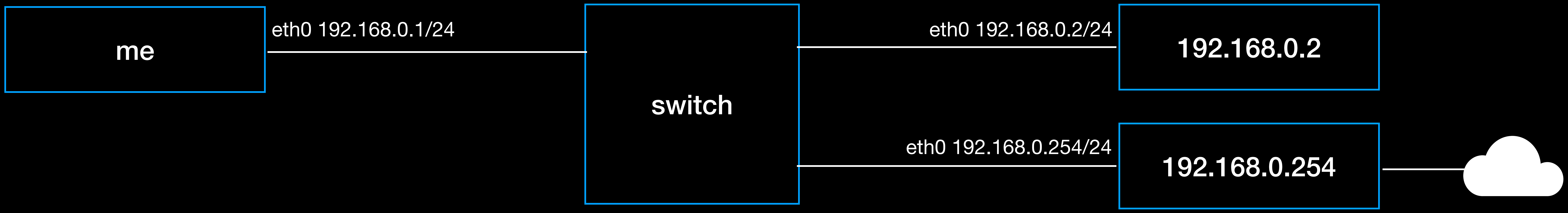
- Most modern high end switches can also route

- Most routers can be made to switch

- We (network administrators) try to avoid mixing L2 and L3 on the same equipment

Routing table

- Routers maintain (several instances of) routing tables
- “directly connected”: my eth0 is 192.168.0.1/24 so I can reach all of 192.168.0.0/24 there
- “static route”: to reach 192.168.10.0/24 I can use gateway 192.168.0.2 - to reach 0.0.0.0/0 (everything else) I can use gateway 192.168.0.254
- Imagine maintaining this for 1,000,000+ routes



Enter: dynamic routing protocols

- “I know 192.168.20.0/24” (and a million other things)
- Internally we use IGPs (ex. RIP, IS-IS, OSPF, …)
- Externally we use EGPs (mainly - exclusively BGP)

Internal/Interior Gateway Protocol

External/Exterior Gateway Protocol

BGP and ASNs

- RFC1771 - could as well be called “the Internet RFC” - https://datatracker.ietf.org/doc/html/rfc1771
- March 1995 - Defines AS, BGP4, …
- Pretty much nothing changed
- AS: “The classic definition of an Autonomous System is a set of routers under a single technical administration, using an interior gateway protocol and common metrics to route packets within the AS, and using an exterior gateway protocol to route packets to other ASs.”
- 1 ASN = 1 “ISP” = 1 participant in the global Internet

ASNs

- 16 bit: 1-65535 (not many available)
- 32 bit: current “standard” but not widely adopted yet - unsupported on older equipment
- Private ASNs: 64512-65534 (like 192.168.x.y but for ASNs)
- Regional Internet Registries (mostly non profit)
    - RIPE (Réseaux IP Européens) - covers Europe + parts of ME
    - ARIN (North America)
    - APNIC (Asia Pacific)
    - LACNIC (Latin America)
    - AfriNIC (Africa)

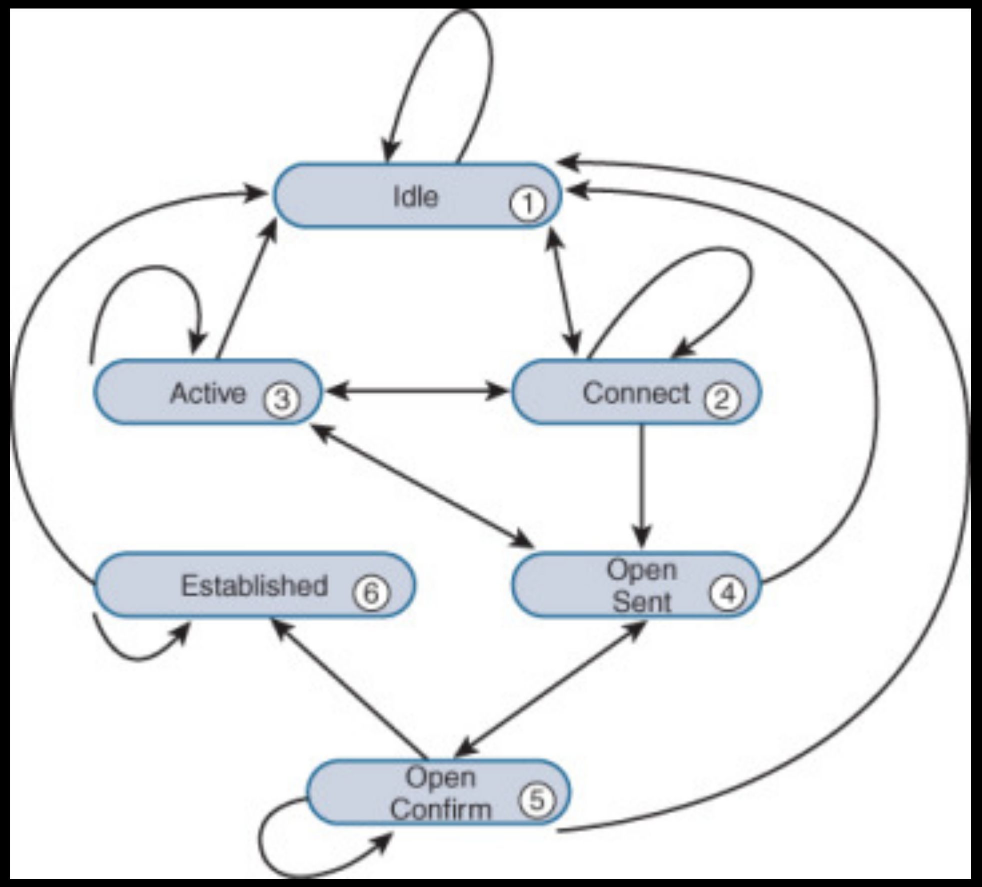
IPs

- Minimum IPv4 announce is /24
- Minimum IPv6 announce is /48
- IPv4 can only be obtained via waitlist or secondary market ($10k-$15k for /24)
- IPv6 mostly free but membership to local RIR required to get ASN
- PA - Provider Aggregatable - tied to an ISP, can't be taken elsewhere - only way, become your own ISP / LIR (Local Internet Registry)
- PI - Provider Independent - mostly gone - allowed to have multiple upstreams (be multi-homed) without being a LIR

Chapter Three: Deep Dive

BGP - RIB - FIB

WTFBBQ



- RFC4271 superseeds RFC1771 (but it’s more of the same)
- BGP exchanges routes: “I know how to reach x.y.z.0/24”
- TCP port 179 - state machine - notice there is no “end”
- iBGP vs eBGP - Internal (same AS - must be continuous!) vs External (different AS / ISP)
- RIB: Routing Information Base (multiple options for one destination)
- FIB: Forwarding Information Base (selected route, inserted in kernel or hardware)

AS-PATH

- Route: 200.132.0.0/16
- AS-PATH: 2914 1916 2716
- “To reach 200.132.0.0/16 through me, you’ll continue through AS 1916 then AS 2716”
- Loop prevention: shouldn’t see own AS in path
- Path selection: shorter = better (see how it doesn’t include capacity?)
- Debugging: example when AWS became unreachable through 5511 but still announced

Route selection

1. Highest weight (cisco specific, local)
2. Highest local preference (within AS)
3. Locally Originated (prefer own routes)
4. Shortest AS-PATH (fewer hops, in theory)
5. Lowest origin type (IGP < EGP < Incomplete)
6. Lowest MED
7. eBGP over iBGP
8. Lowest IGP metric to next-hop

Also, more specific route **always** wins.

/24 beats /22

/24 beats /16

etc.

This allows some level of control on ingoing traffic… but also hijacks!

Traffic engineering

A (short) story about control

- Very easy to influence outgoing traffic - on top of BGP “natural” decisions, you can do whatever you want and “force” traffic to a particular destination through a particular upstream

- Well, you can force the first hop

- Used to control price and capacity planning

- Very hard to impossible to control incoming traffic - precisely because it’s easy to control outgoing - hacks: announcing more specific (hard to counter), prepending AS-PATH (easy to counter)

- number of hops >= AS-PATH length

- actual number of hops >= number of hops visible in traceroute

More about Transit vs Peering

- Transit
  - Paid
  - Provider gives the “full view” (1M+ routes = 0.0.0.0/0)
  - You pay for bandwidth, not traffic (paying for traffic is a SCAM)
  - Commit vs Burst and 95th percentile
- Peering
  - Usually free (you pay the port if IXP)
  - Each side announces only own network (+ optionnaly customers/downstreams)
  - Private (PNI) or via IXP
  - Based on mutual benefit
- Tier-1 networks
  - Can reach 100% of the Internet via Peerings only
  - NTT, Lumen/Level3 (Colt), GTT, Telia, Orange International, …

IXPs

- Networks meet to exchange traffic
- Large switch fabric in carrier neutral datacenters
- Monthly port fee (10/100/400 Gbps)
- Route servers: one BGP to access many “open peering” networks
- Typical traffic: 10+ Tbps on large ones
- DE-CIX, AMS-IX, LINX, France-IX, …

Building blocks of the internet
In the 1980s engineers and scientists devised ways to link sprawling computer networks into a single unified ‘inter-network’ or internet. The internet grew into a global system of interconnected computers, requiring an enormous physical infrastructure. The internet initially relied on existing telephone networks, but increasingly data was shared through high-speed fibreoptic cables.

1 **Fibreoptic cable, used on the Hitchin–Stevenage telephone link, 1977** Following trials in the 1970s, fibreoptic cables were used to create the backbone of the internet.

Donated by: STC plc
Object No: 2014-57

2 **Cisco Catalyst 1200 switch, 1994**
This switch was used by the London Internet Exchange (LINX), the first exchange to route internet traffic for internet service providers (ISPs) across UK networks.

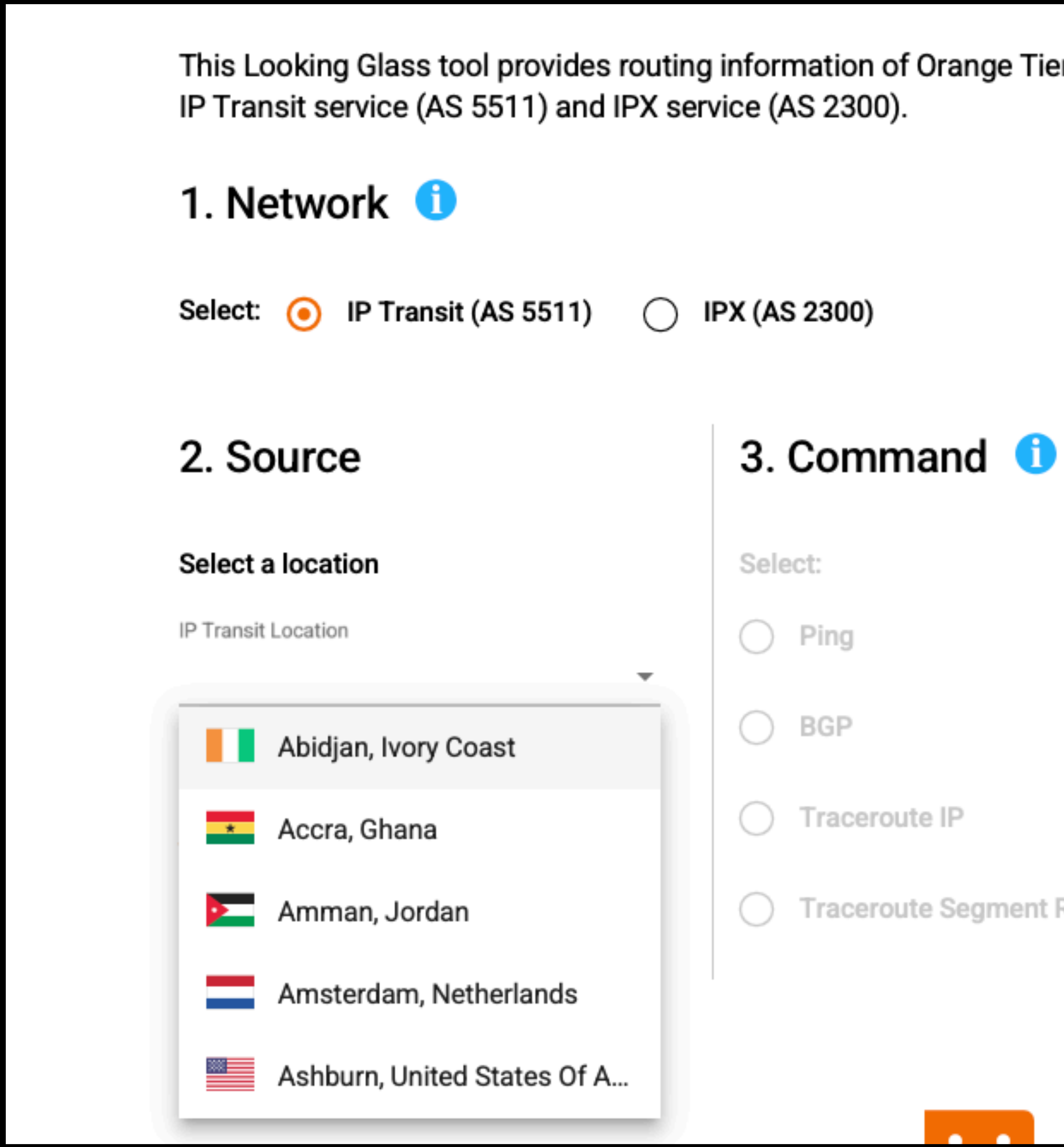
Donated by: London Internet Exchange
Object No: 2004-268

Charles Kao (born 1933) pioneered the use of fibreoptic cables for telecommunications. As well as early theoretical work on the properties of light, Kao played a key role in the practical realisation of the technology. In the 1980s and 1990s the burgeoning internet made use of fibreoptic cables for sending information around the globe.

Image: Science Museum/SSPL

Go explore

- Access directly routers (through a website a.k.a. “looking glass”) or via telnet: http://traceroute.org/ - used daily to troubleshoot issues, check reachability, …
- Web view of the Internet as seen from Hurricane Electric (wanabee Tier-1): https://bgp.he.net/
- Type ASN + Looking glass in a search engine - ex: AS5511 Looking Glass: https://looking-glass.opentransit.net/
- Many big players make looking glass public because it’s a proof of their interconnections and helps troubleshooting
- Learn on a small scale simulated Internet: https://dn42.eu/ (it has all the features of the real one!)
- bgpq3/bgpq4 (https://github.com/bgp/bgpq4)



This Looking Glass tool provides routing information of Orange Tie
IP Transit service (AS 5511) and IPX service (AS 2300).

1. Network

Select: ● IP Transit (AS 5511) ○ IPX (AS 2300)

2. Source

Select a location

IP Transit Location

- Abidjan, Ivory Coast
- Accra, Ghana
- Amman, Jordan
- Amsterdam, Netherlands
- Ashburn, United States Of A...

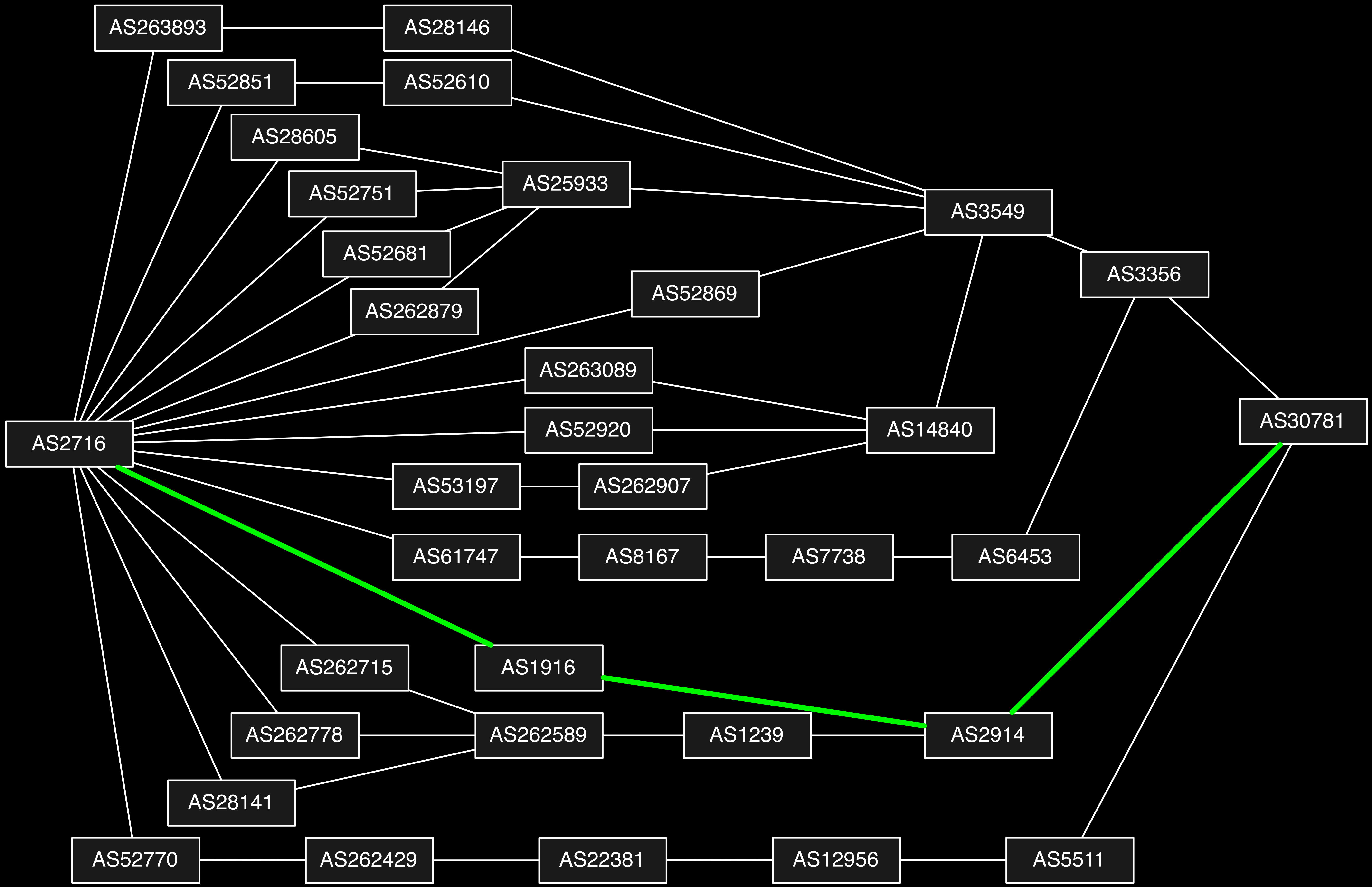
3. Command

Select:

○ Ping
○ BGP
○ Traceroute IP
○ Traceroute Segment

Asymetric routing

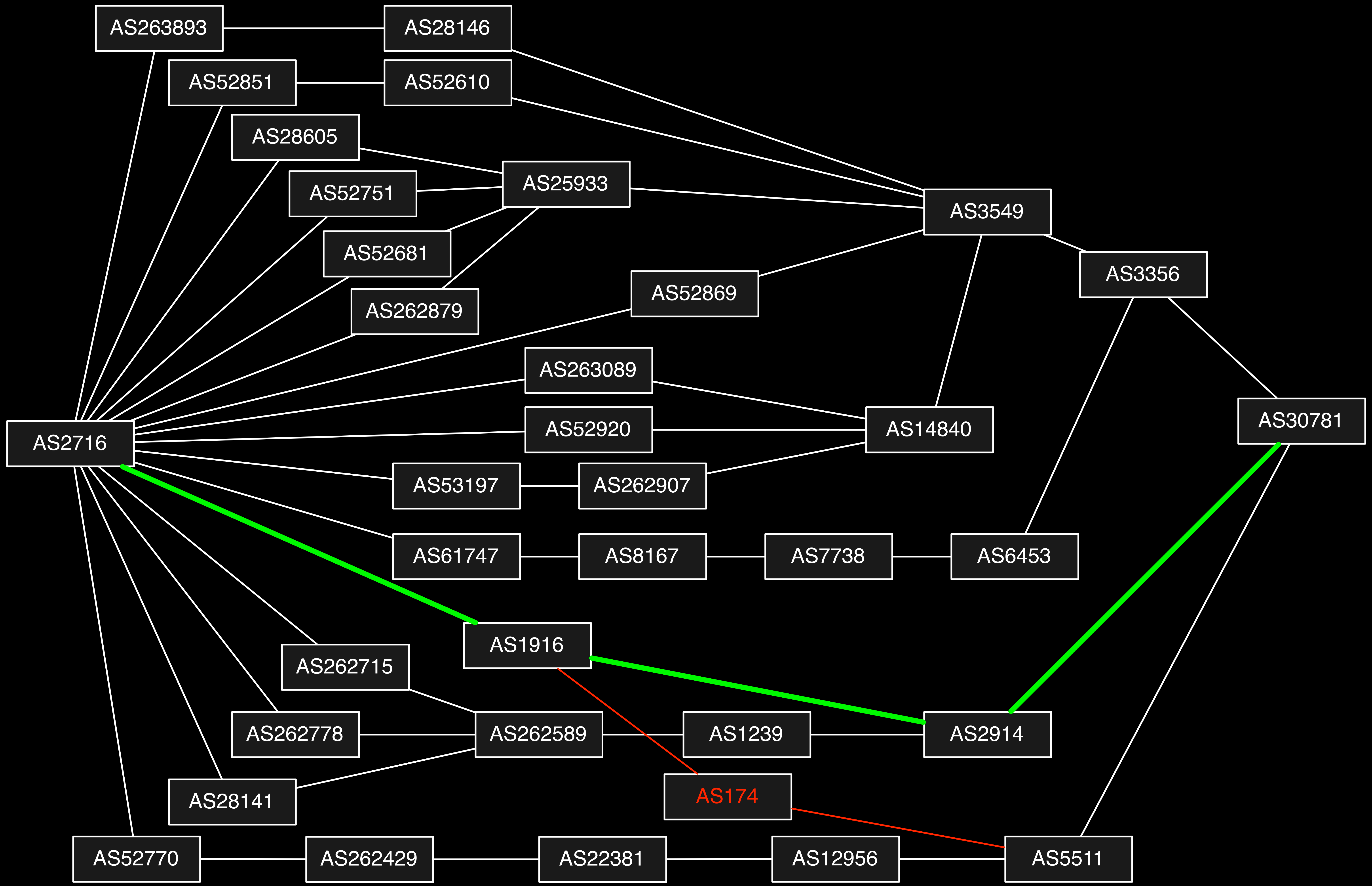
From Rio to Paris



```
ASHandle:       AS3549 (ex-GBLX)
OrgName:        Level 3 Communications, Inc.

aut-num:        AS3356
as-name:        Level3
descr:          Level 3 Communications

aut-num:     AS25933
owner:       Sul Americana Tecnologia e Inform?tica Ltda.

aut-num:        AS5511
as-name:        Opentransit
descr:          Orange S.A.
remarks:        Orange - Worldwide IP Backbone

aut-num:        AS3215
as-name:        AS3215
descr:          Orange S.A.
```

(A)symetric routing

From Paris to Rio



Chapter Four: Trust & Security

BGP’s trust model

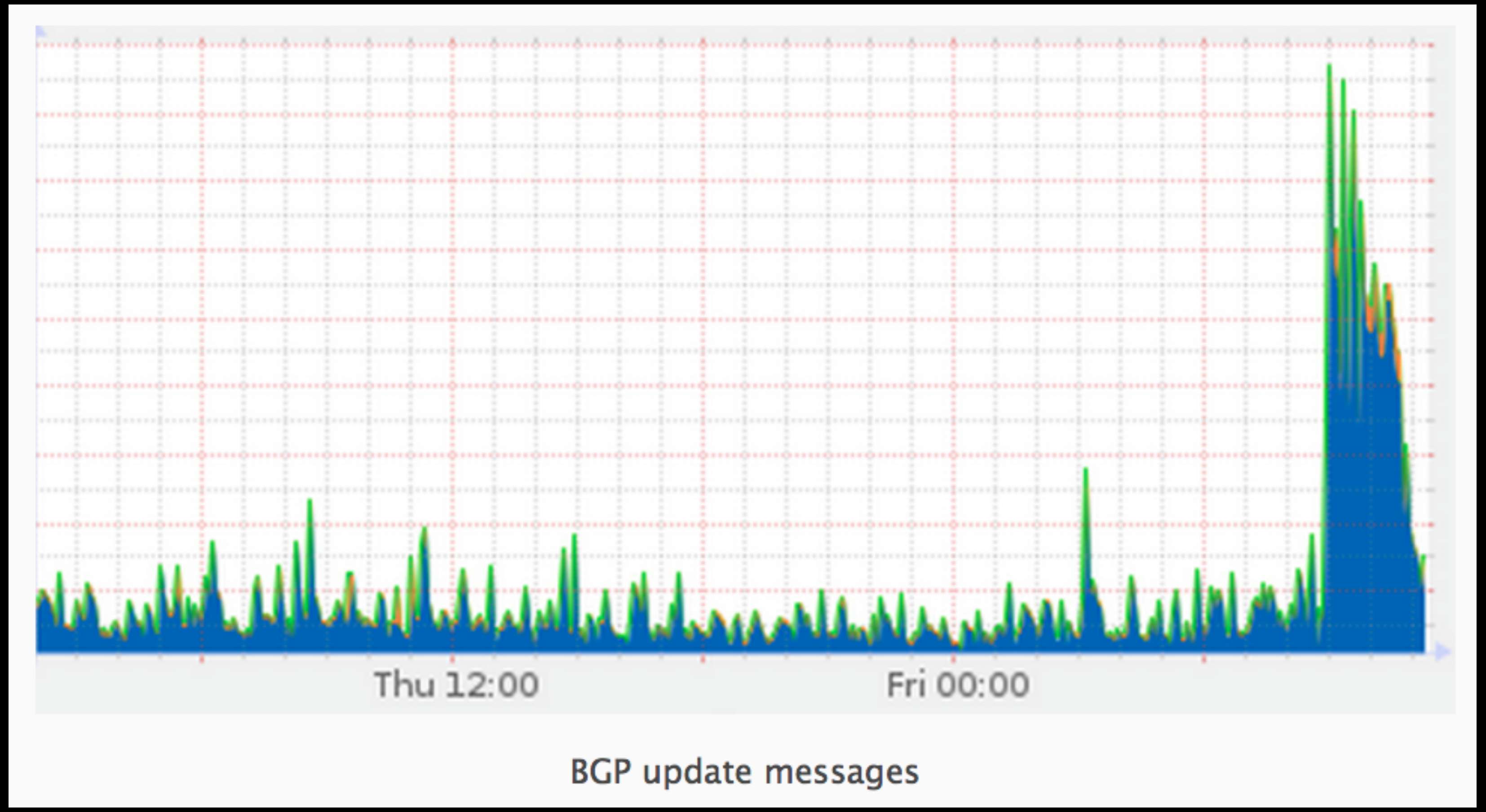
Trust, don’t verify

- By default, we accept everything
  - peer says they can reach 8.8.8.0/24 -> we send them traffic for 8.8.8.0/24
  - peer says they have a quick path to AWS -> we send them traffic for AWS
- Why this works
  - small community
  - reputation matters
  - mostly mistakes, few malicious

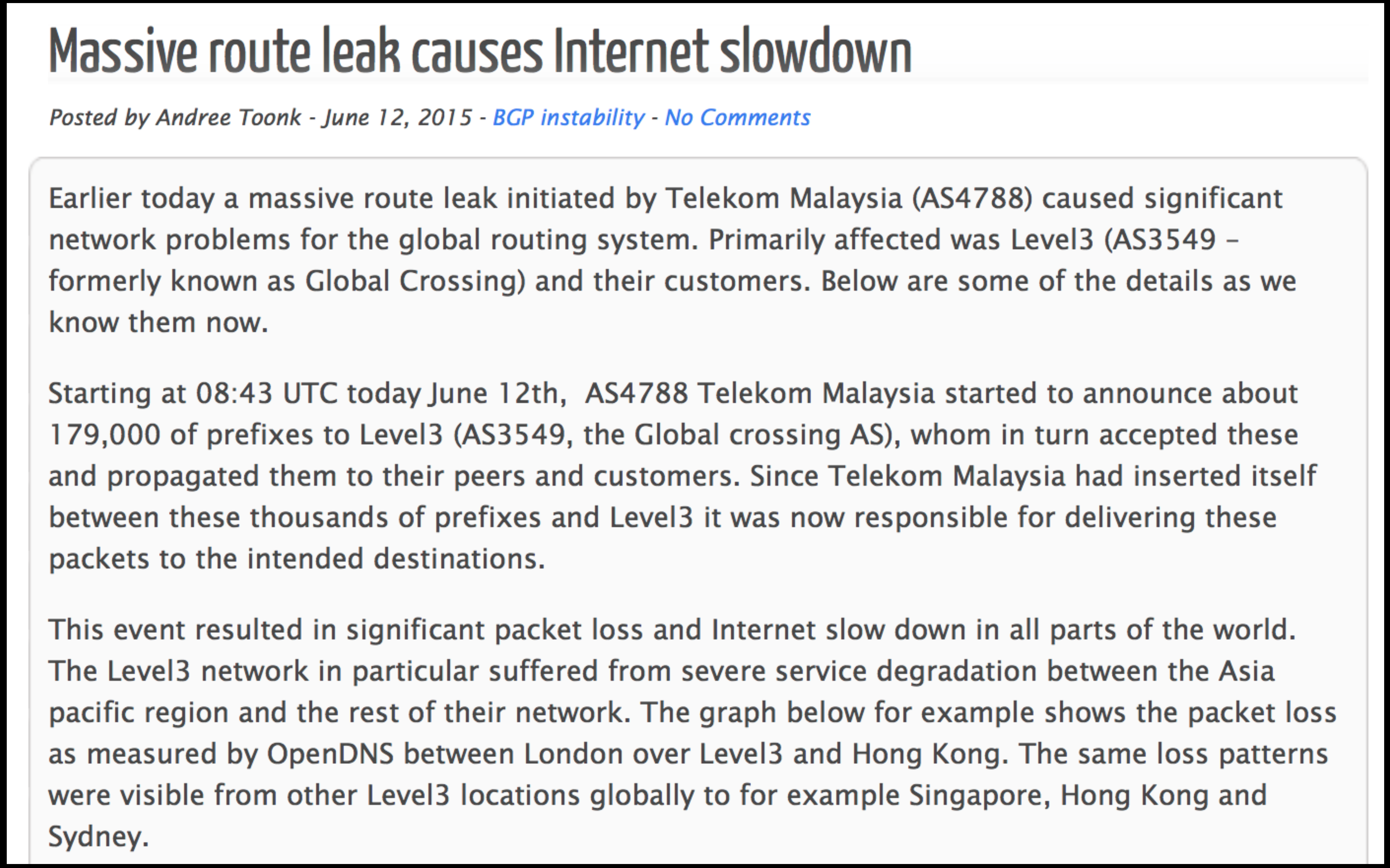
The Telekom Malaysia incident

AS4788 announced 179,000 prefixes to Level 3 (Tier-1)

- Level 3 accepted and propagated to their other customers
- Global internet slowdown
- Mostly affected APAC
- Level 3 network seriously impacted



BGP update messages



Massive route leak causes Internet slowdown

Posted by Andree Toonk - June 12, 2015 - BGP instability - No Comments

Earlier today a massive route leak initiated by Telekom Malaysia (AS4788) caused significant network problems for the global routing system. Primarily affected was Level3 (AS3549 – formerly known as Global Crossing) and their customers. Below are some of the details as we know them now.

Starting at 08:43 UTC today June 12th, AS4788 Telekom Malaysia started to announce about 179,000 of prefixes to Level3 (AS3549, the Global crossing AS), whom in turn accepted these and propagated them to their peers and customers. Since Telekom Malaysia had inserted itself between these thousands of prefixes and Level3 it was now responsible for delivering these packets to the intended destinations.

This event resulted in significant packet loss and Internet slow down in all parts of the world. The Level3 network in particular suffered from severe service degradation between the Asia pacific region and the rest of their network. The graph below for example shows the packet loss as measured by OpenDNS between London over Level3 and Hong Kong. The same loss patterns were visible from other Level3 locations globally to for example Singapore, Hong Kong and Sydney.

The Youtube incident

Countrywide censorship goes wrong

- Pakistan Telekom tried to block Youtube domestically
- Routed 208.65.153.0/24 internally to block traffic
- They announced the route to their upstream (PCCW)
- PCCW propagated the announce globally
- Youtube down worldwide for ~2 hours, all traffic sent to Pakistan

The Facebook incident

Angle grinders to the rescue

- 2021: BGP configuration mistake in Facebook network
- 6 hours downtime
- Offices access control on same network, no access to office to fix issue
- History says angle grinders were used to regain access
- $6 billion estimated loss

The MyEtherWallet incident

Only example here with malicious intent

- Malicious actor at small ISP (or client of) announced Route53 DNS service IPs
- Upstream propagated to whole Internet
- Redirected users to malicious version of MyEtherWallet
- ~$150k stolen

Defense: filtering

- Extensive filtering must exist
- For ex. BCP-38 (Best Common Practice) - for Ingress - applies to both Telekom Malaysia and Pakistan Telekom examples
- Basic filters such as max-prefix (an ISP with 100 routes shouldn’t be sending 1000 or 100000)
- Bogons (routes that shouldn’t exist)
- AS-PATH filters (length and content)
- IRR-based filters (use RIR databases to build filters - declarative only)

This doesn’t sound very advanced (because it’s not)

RPKI: Route Origin Validation

An attempt to make things better

- Cryptographic validation of routes
- ROA “Route Origin Authorization” - signed statement “AS12345 may announce x.y.z.t/24”
- But… validation states: Valid (exists and matches) - Invalid (exists and doesn’t match, route should be refused) - Unknown (no ROA)
- Deployment
  - ~50% of routes have ROAs
  - ~30% of networks check them
  - ~15% reject bad
- Honorable mention for BGPSEC (exists, not really used)

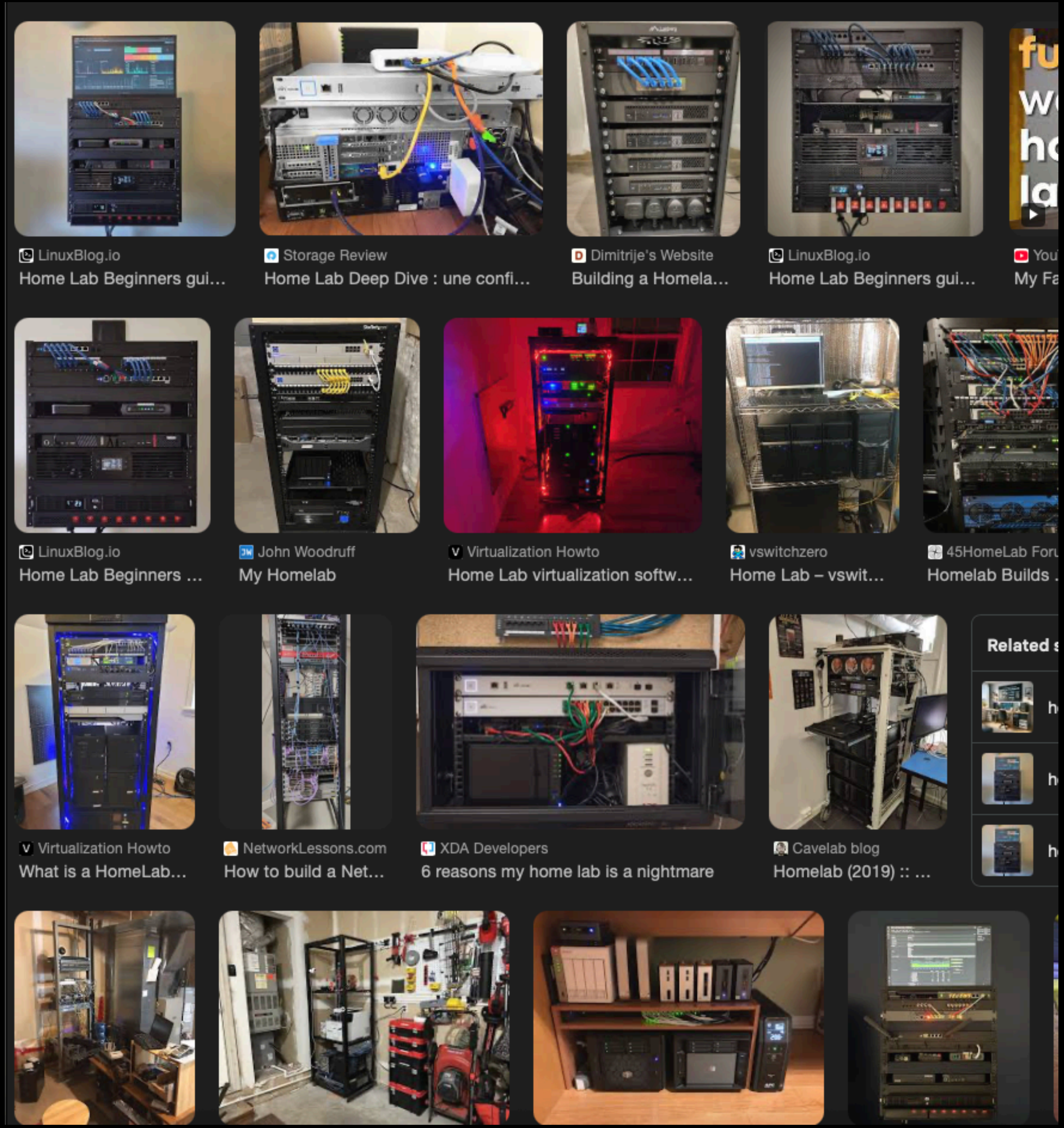
BCP, MANRS, …

We are your friends

- https://manrs.org
- 1000+ participants, no enforcement, naming and shaming
- Four actions:
  - Filtering: prevent propagation of incorrect routing info
  - Anti-spoofing: validate source addresses (BCP-38)
  - Coordination: maintain contact infos, joins NOGs, …
  - Validation: publish routes in IRR/RPKI…

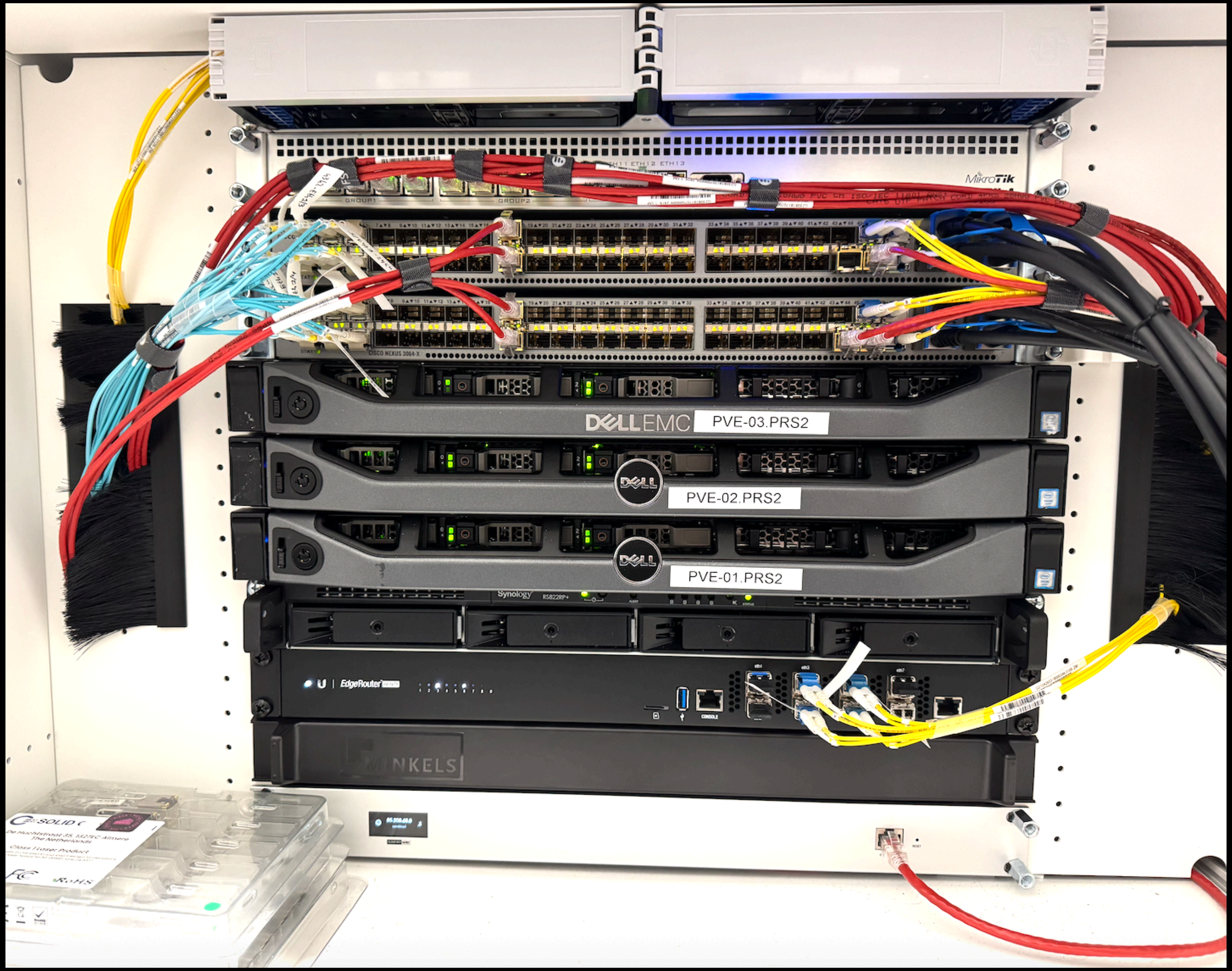
Chapter Five: Implementation

“Self-hosting”



Single power source, usually single attachement to “the Internet”, etc.
- connected TO the Internet -

or



Redundant power with N+N UPS, ASN, multiple IP subnets and upstreams, connections to IXPs, etc
- part OF the Internet -

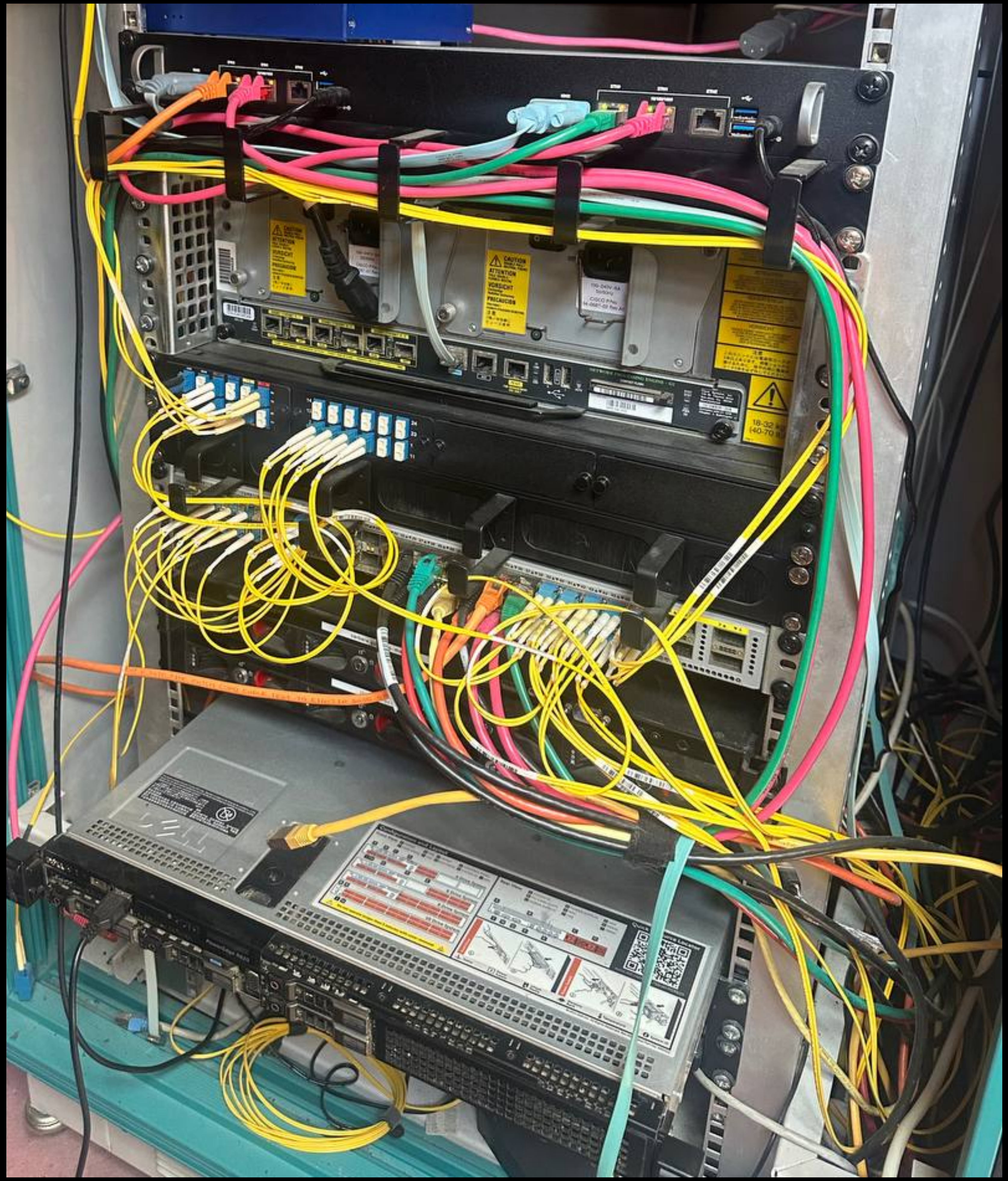
Start small and iterate

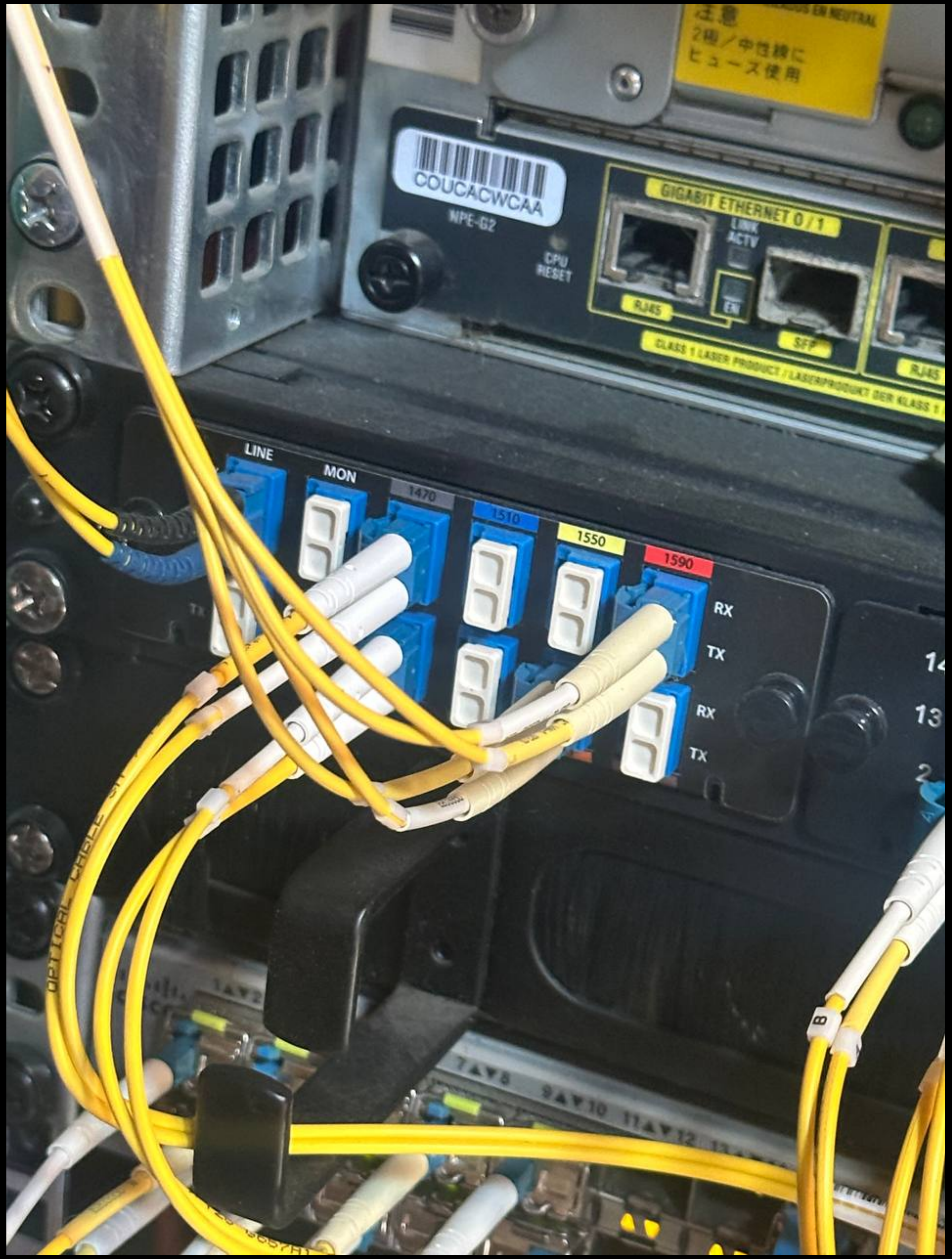
- Optimisations for the “home-lab” hosting:
  - get an ISP proposing fixed IPv4 (hint: starlink (business plans) is one!)
  - get UPS, generator, etc.
- If you live in a (european) city, you can go further
  - get dedicated (or else) fiber(s) to one or several datacenters
  - get an ASN + IPv6 block (comes with RIPE membership - 1000 Eur sign-up + 1800 Eur yearly + 50 Eur yearly for ASN) - pricing model evolves yearly
  - get an IPv4 block (/24 minimum for BGP) - 8-12kEur right now / or wait list (3 years?)
  - use existing DSL/FTTH for OOBM (very important!)
- Or you can simulate this with VPNs into a couple of friendly ISPs that will provide BGP to you

Personal ISP/ASN

- You don’t have to be a legal entity
- Many of RIPE NCC members are private individuals
- ex. https://bgp.he.net/AS35360#_asinfo - directly connected to 1000+ other ASNs… from home
- ex. https://bgp.he.net/AS44097 - less peers but has 100 Gbps fiber at home :)
- ~3000 new ASNs each year!
- Keep in mind: it still mainly happens in IRL events, mailing lists, handshake agreements

Personal ISP/ASN (AS35360)





A special experiment

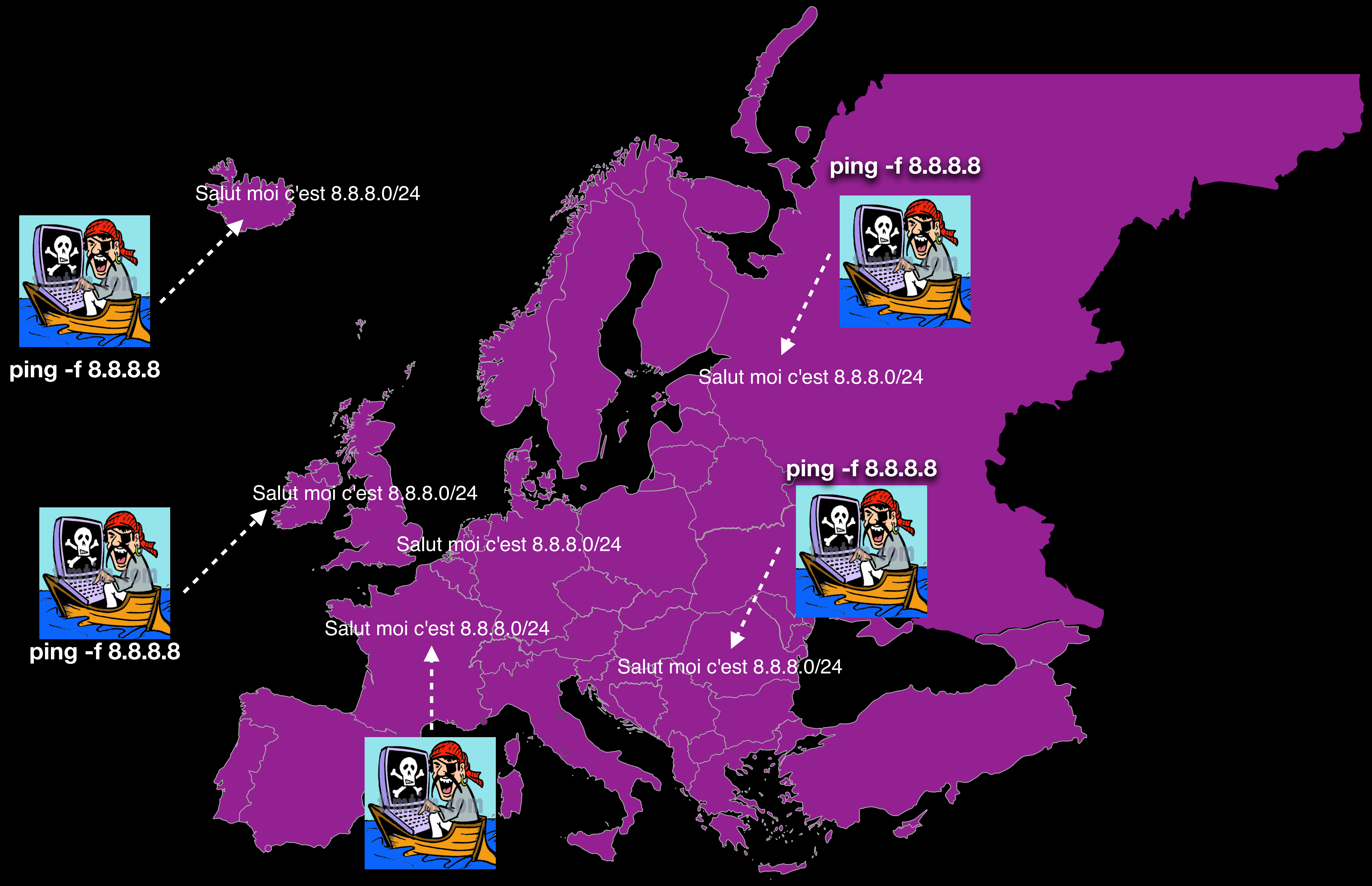
Anycast: the art of being everywhere

- One announced block has no reason to be present in only one location

- Works well for single question/response UDP*, less for TCP

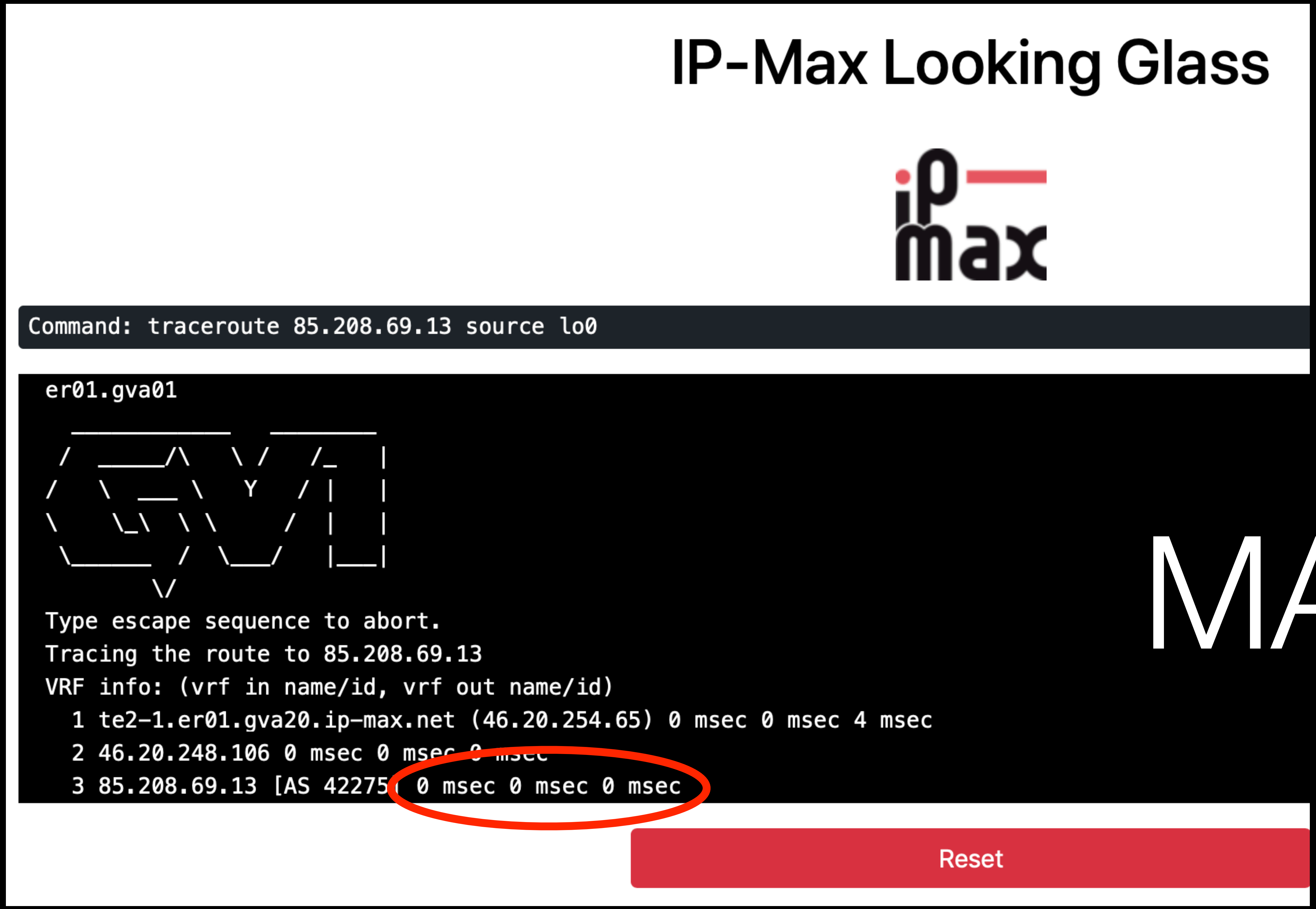
- Typically used for DNS then directing users to the closest service server farm

*this does not include HTTP over UDP, which is TCP-like


Salut moi c'est 8.8.8.0/24
ping -f 8.8.8.8
ping -f 8.8.8.8
Salut moi c'est 8.8.8.0/24
Salut moi c'est 8.8.8.0/24
Salut moi c'est 8.8.8.0/24
ping -f 8.8.8.8
Salut moi c'est 8.8.8.0/24
ping -f 8.8.8.8
Salut moi c'est 8.8.8.0/24

A special experiment

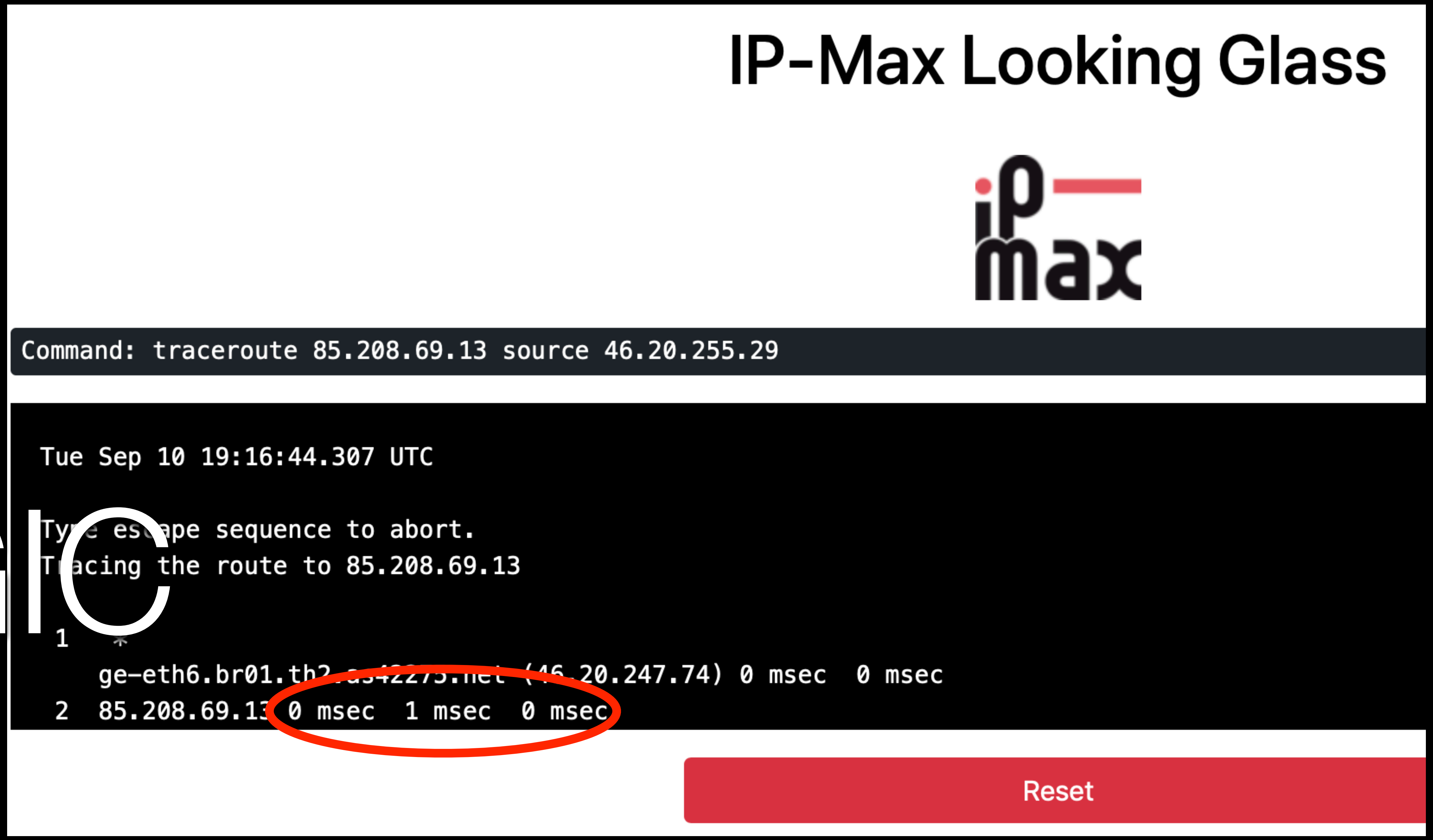
Let’s have some fun (2019)



```
Command: traceroute 85.208.69.13 source lo0
```

```
er01.gva01
   ____________   ________
 /   _____/\   \ /   /_   |
/   \  ___\   Y   / |   |
\    \_\  \ \     /  |   |
 \______  /  \___/   |___|
        \/
Type escape sequence to abort.
Tracing the route to 85.208.69.13
VRF info: (vrf in name/id, vrf out name/id)
  1 te2-1.er01.gva20.ip-max.net (46.20.254.65) 0 msec 0 msec 4 msec
  2 46.20.248.106 0 msec 0 msec 0 msec
  3 85.208.69.13 [AS 42275] 0 msec 0 msec 0 msec
```

MAGIC



```
Command: traceroute 85.208.69.13 source 46.20.255.29
```

```
Tue Sep 10 19:16:44.307 UTC

Type escape sequence to abort.
Tracing the route to 85.208.69.13

 1  *
    ge-eth6.br01.th2.as42275.net (46.20.247.74) 0 msec  0 msec
 2  85.208.69.13 0 msec  1 msec  0 msec
```

traceroute from Geneva

traceroute from Paris

Average latency from Geneva to Paris = ~9ms

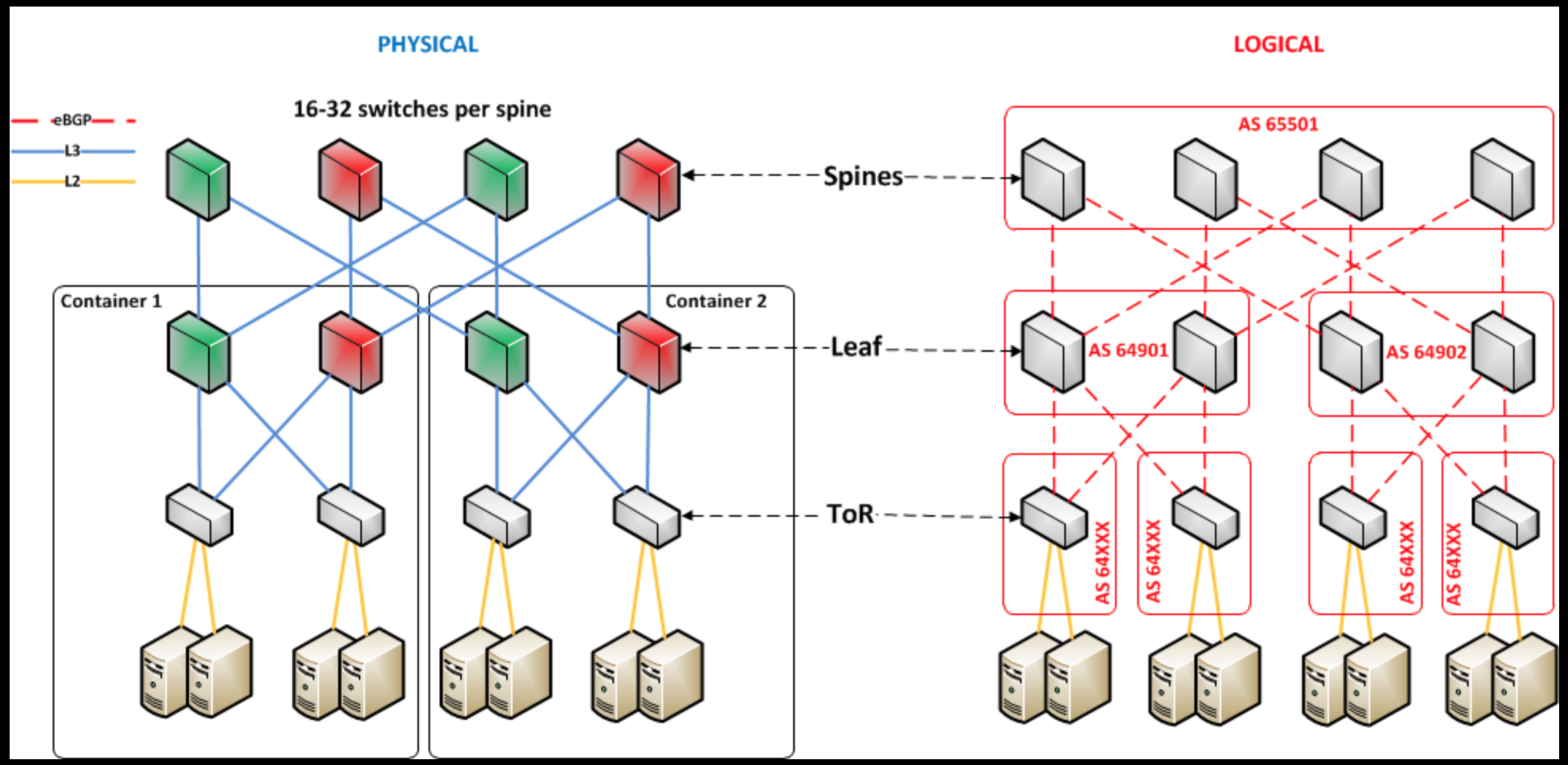
A special experiment

Why?

- Geographic load distribution
- Resilience (one location fails, next closest one takes over)
- DDoS absobtion (targets gets distributed too despite being a “single” IP)
- Replaces DNS seed for Bitcoin?

Going deeper

into the network



- Tools like ExaBGP allow announcing services, not networks
- CLOS topology
- Used by “big cloud” everywhere

https://en.wikipedia.org/wiki/Clos_network

http://www.networkworld.com/article/2226122/cisco-subnet/clos-networks--what-s-old-is-new-again.html

http://conferences.sigcomm.org/sigcomm/2015/pdf/papers/p123.pdf

https://code.facebook.com/posts/360346274145943/introducing-data-center-fabric-the-next-generation-facebook-data-center-network/

https://tools.ietf.org/html/draft-lapukhov-bgp-routing-large-dc-02

https://www.nanog.org/meetings/nanog55/presentations/Monday/Lapukhov.pdf

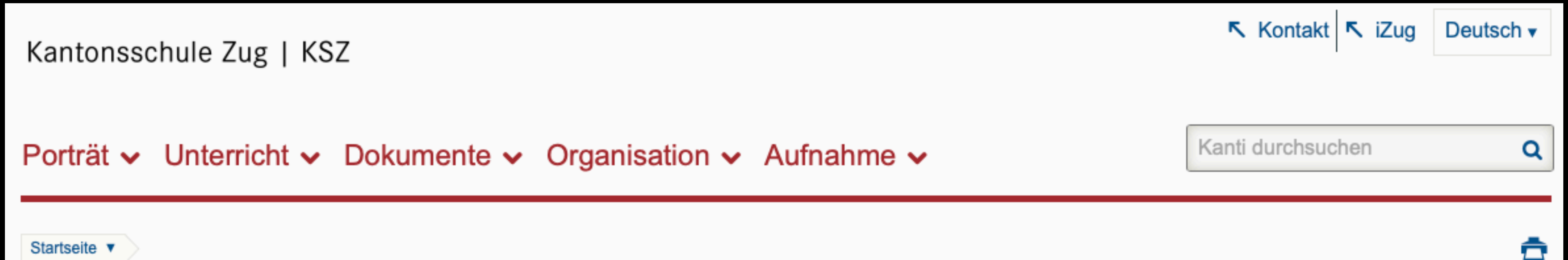
```
#!/bin/sh
while true; do
  if check-if-ok; then
    echo "announce"
  else
    echo "withdraw"
  fi
  sleep 5
done
```

Take aways

- No single point of control
- No accountability but Trust through Reputation - also Naming and Shaming!
- Engineer based community (and a small one!)
- NANOG, FRnOG, etc…
- Mostly acting in the best interest of their users
- work in progress: https://manrs.org/ (Mutually Agreed Norms for Routing Security)
- Bogons…
- Don’t let all this scare you!

Opinions (from someone running multiple ASNs for 20+ years)

- Bitcoin runs on top of the Internet (at least for now)
- A system built on top of another system requires understanding of the underlying system
- Some initiatives (asmap) go in the right direction
- Participate in the decentralisation by running your own AS ($$)
  - there are actors in the community who already do!
- We haven’t touched the fun parts yet: traffic engineering, etc.
- If they can do it, you can! - AS34288
- I can help!



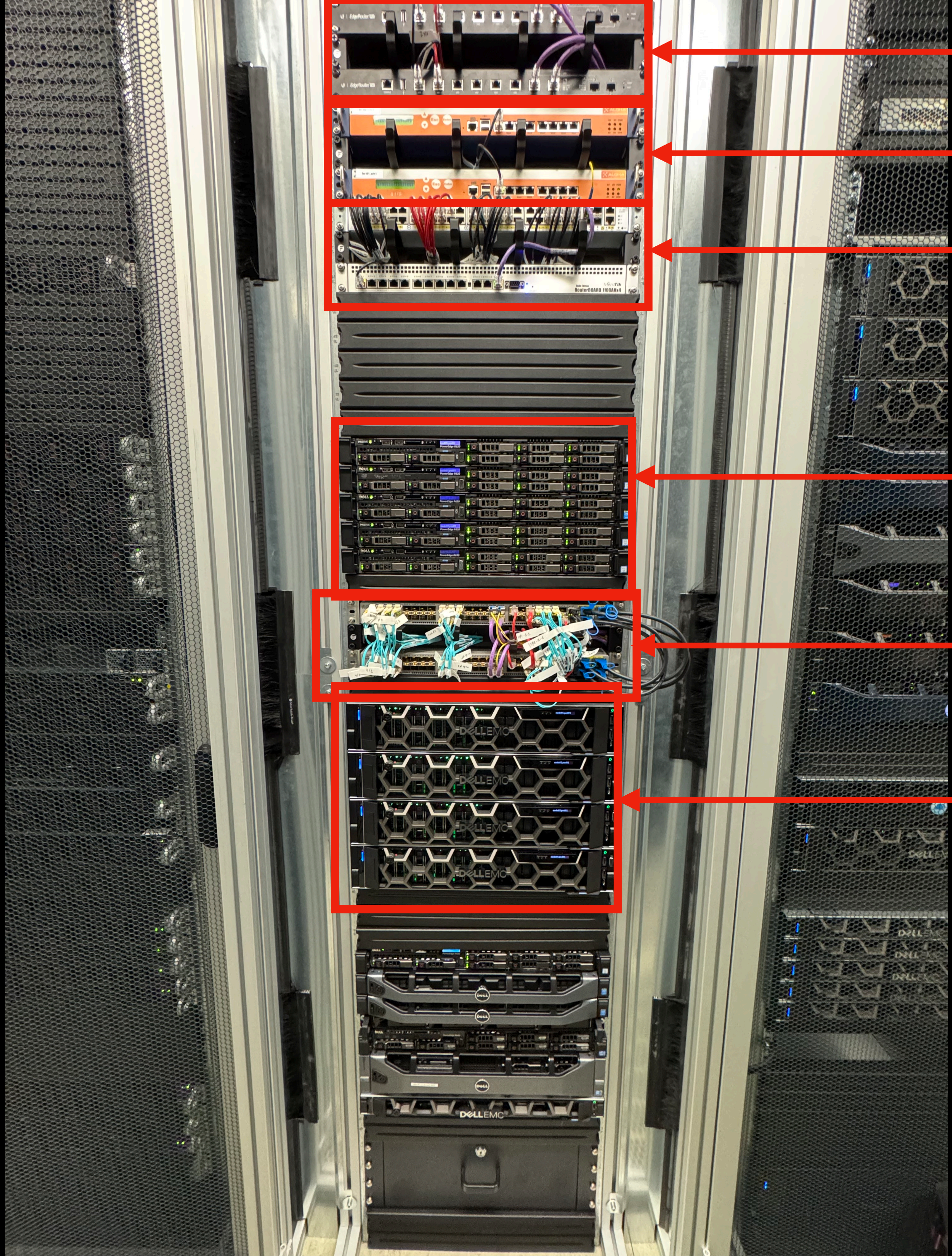
Q&A and Contact info

contact@ketominer.pw (personal)

ketominer@nodl.eu (2 ASNs - FR / CH)

michel@three-fourteen.eu (2 ASNs - FR / FR-CH)

440C 1576 9D19 E690 8CC1 DDB2 3070 DE97 72DB 8A48



Upstream ISP routers

nodl.cloud OpenBSD routers (recycled haproxy appliances)

Out of band management router and switch

“pod02” (general purpose compute servers)

L2 redundant switches

“pod01” (bitcoin nodes and high end general purpose compute - you are here!)

nodl cloud Zurich ZRH1 infrastructure