

Fundamentals of Data Analytics

SEN163A

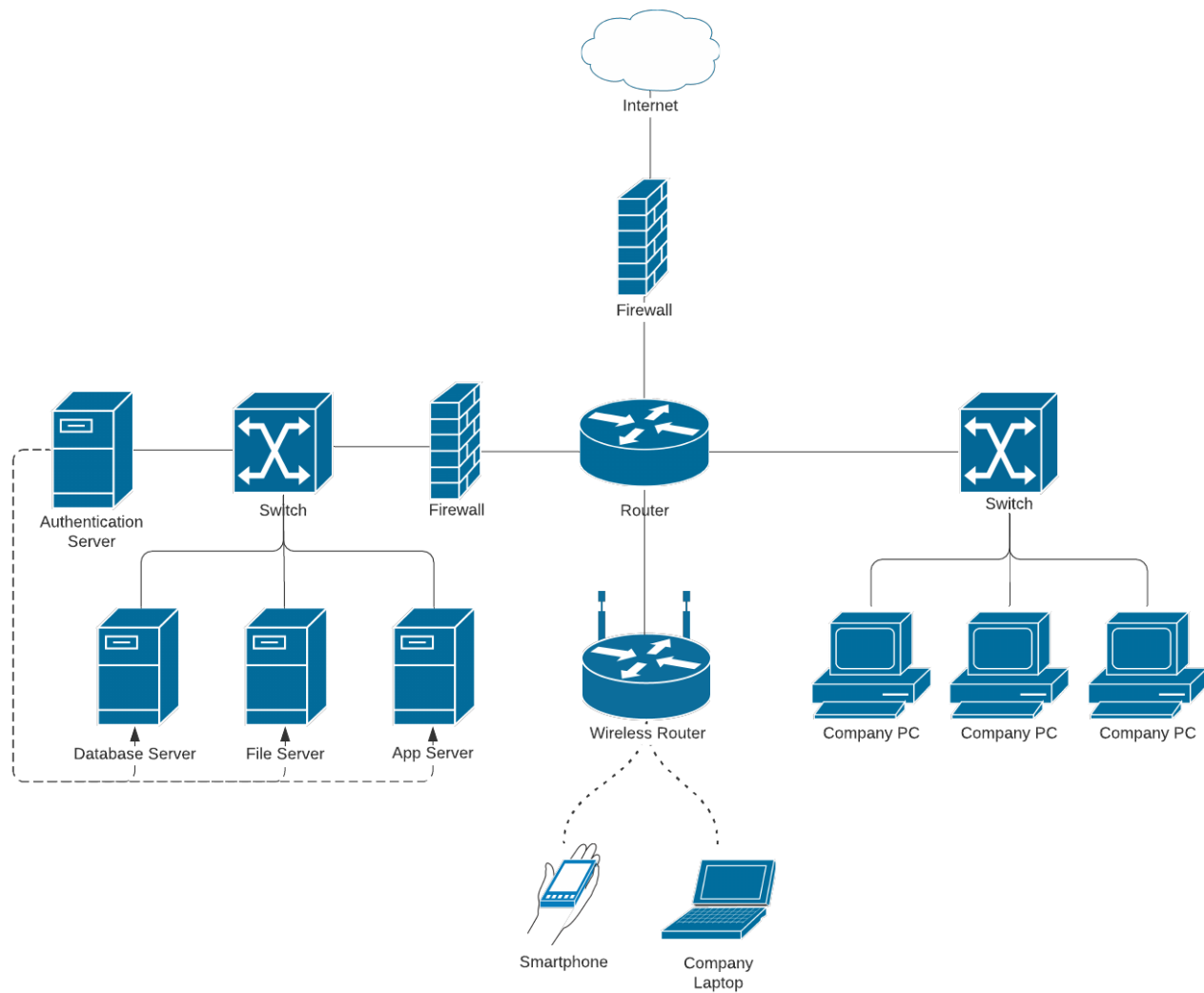
Internet measurements

Jacopo De Stefani

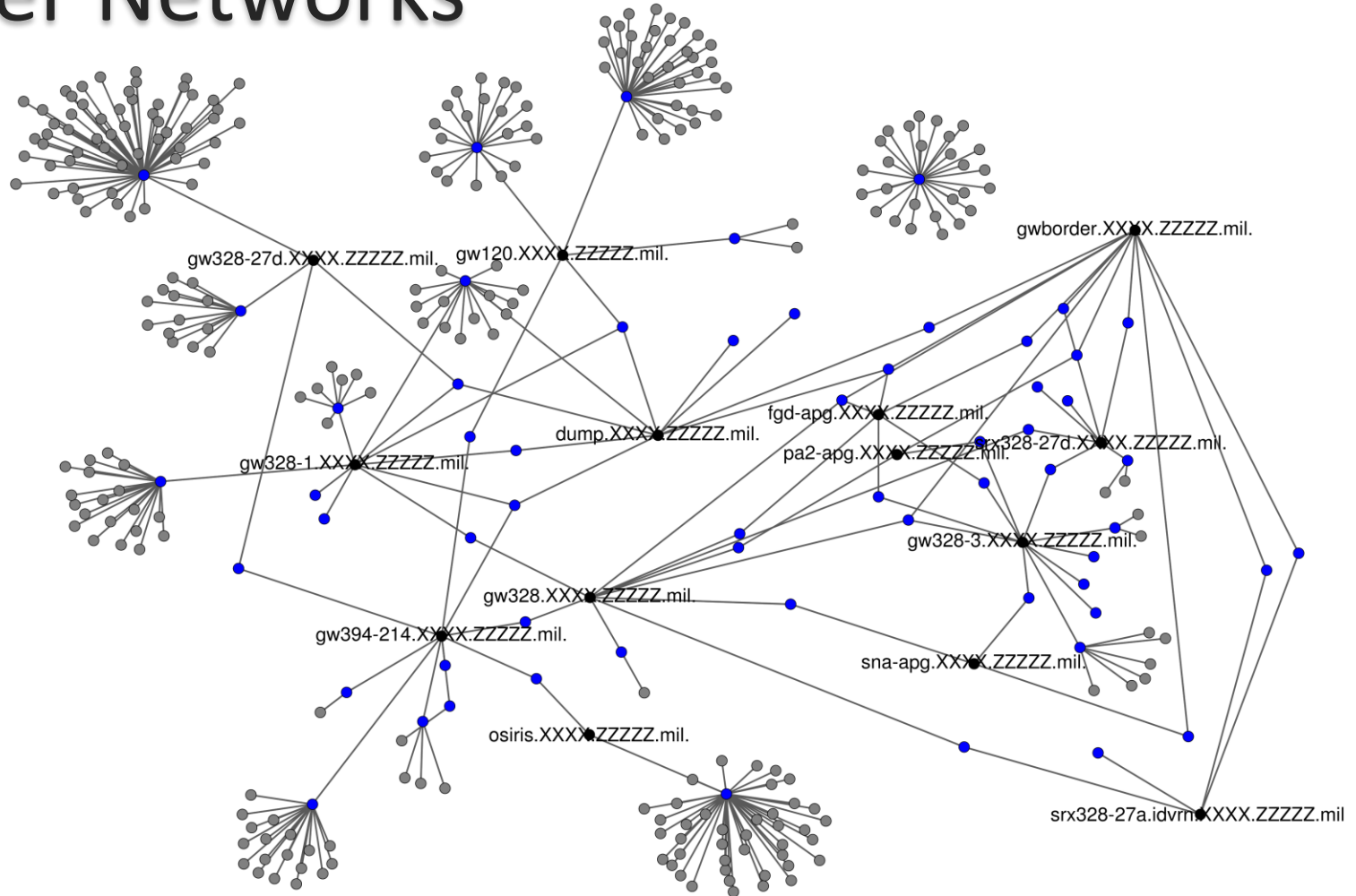
Based on the material by Tobias Fiebig

Outline of the lecture

- Review of the basic networking concepts
- Types of internet measurements: passive vs active measurements with examples
- Internet measurement in practice



Computer Networks



Key Concepts

- Host
- IP
- Port
- DNS : Host \leftrightarrow IP

Key Terms

- Autonomous Systems (AS): The individual networks which make up the Internet. Also has a more technical dimension
- Network: A set (or block) of IP addresses. ASes use (advertise) one or multiple networks.
- Prefix size: Determines the number of IP addresses in a Network (usually in CIDR notation)

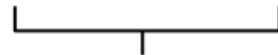
IPv4 addresses

IPv4 address in dotted-decimal notation

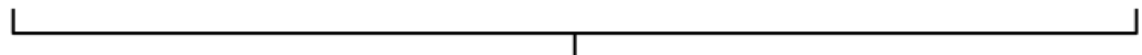
172 . 16 . 254 . 1



10101100 . 00010000 . 11111110 . 00000001



8 bits



32 bits (4 bytes)

IPv6 addresses

An IPv6 address (in hexadecimal)

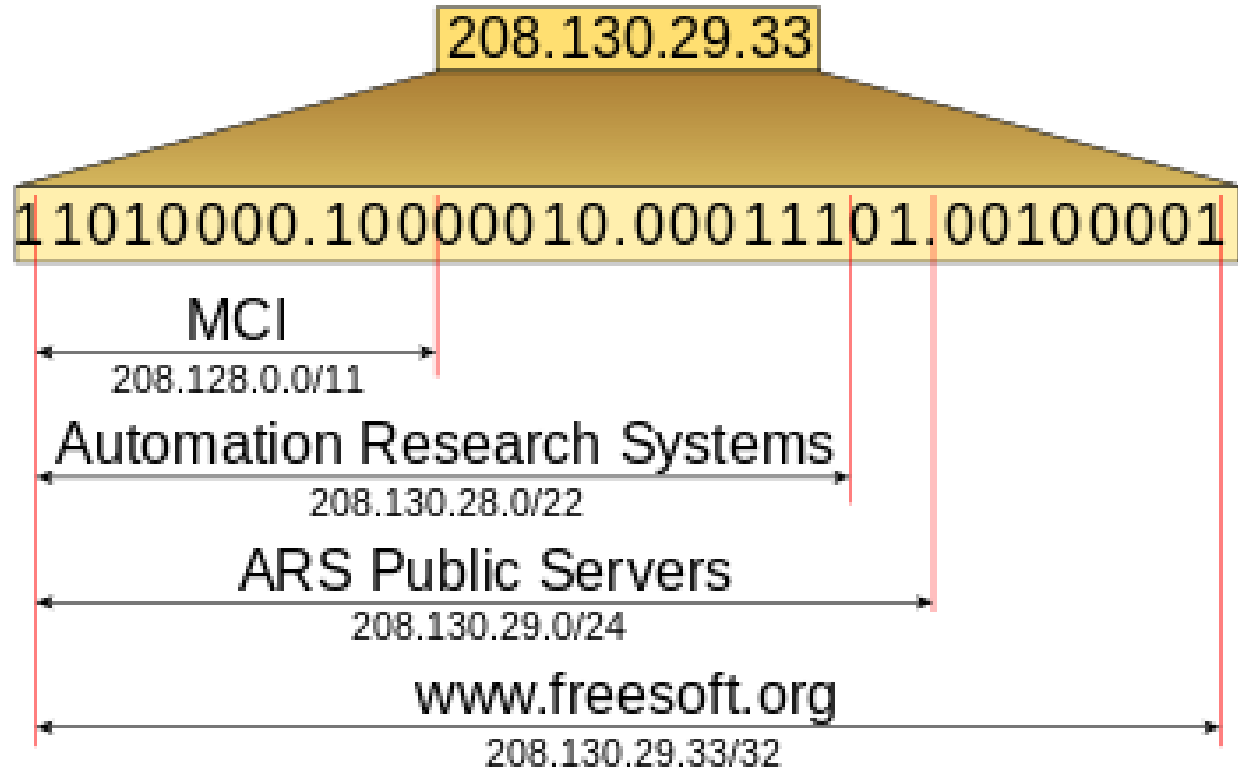
2001:0DB8:AC10:FE01:0000:0000:0000:0000



2001:0DB8:AC10:FE01:: Zeroes can be omitted

0010000000000001:0000110110111000:1010110000010000:1111111000000001:
0000000000000000:0000000000000000:0000000000000000:0000000000000000

CIDR Notation



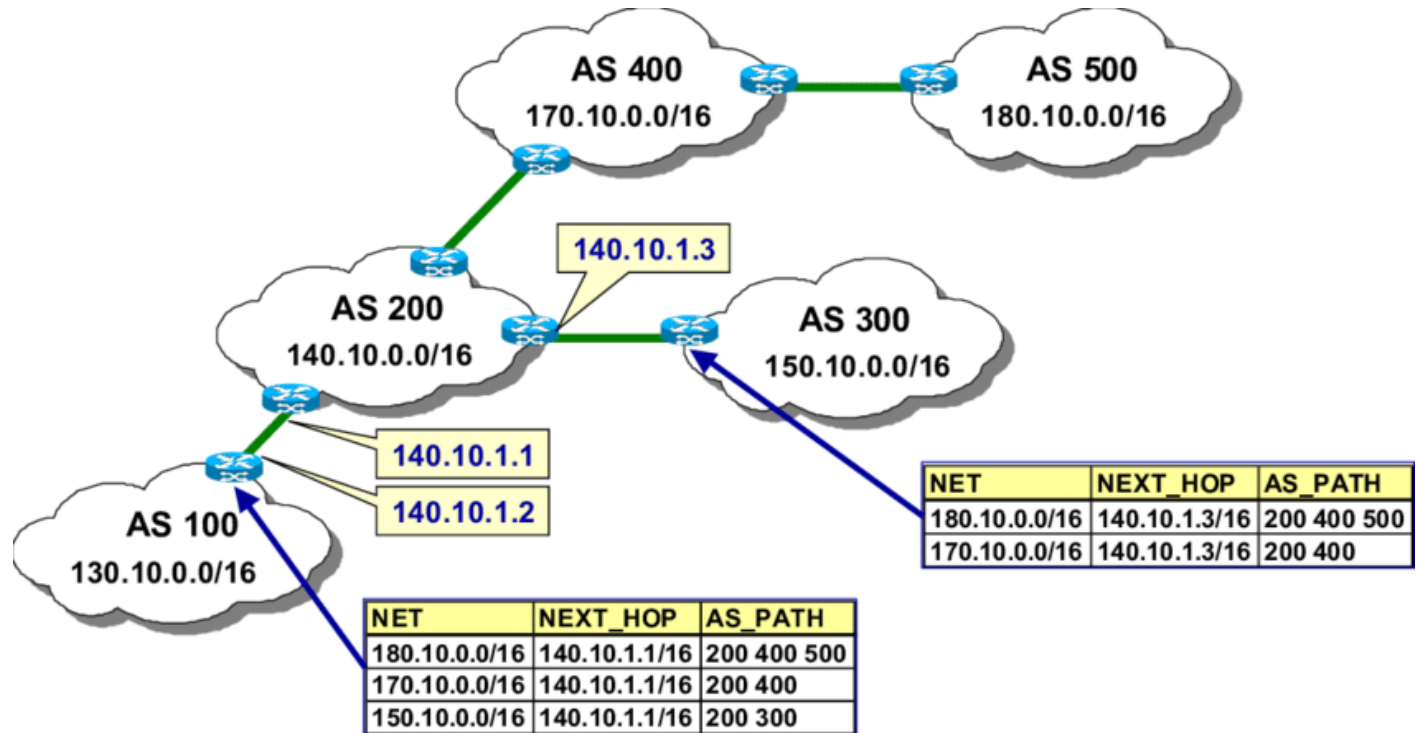
Hosts and names

- Hosts should have names
- We usually look at the forward direction (name to IP)
- There are also DNS measures to resolve the name for an IP (if configured by the operators)
 - 2a01:4f8:10b:37ef::186:
6.8.1.0.0.0.0.0.0.0.0.0.0.0.0.0.f.e.7.3.b.0.1.0.8.f.4.0.1.0.a.2.ip6.arpa
domain name pointer mail.aperture-labs.org.
 - 94.130.126.186
186.126.130.94.in-addr.arpa
domain name pointer mail.aperture-labs.org.



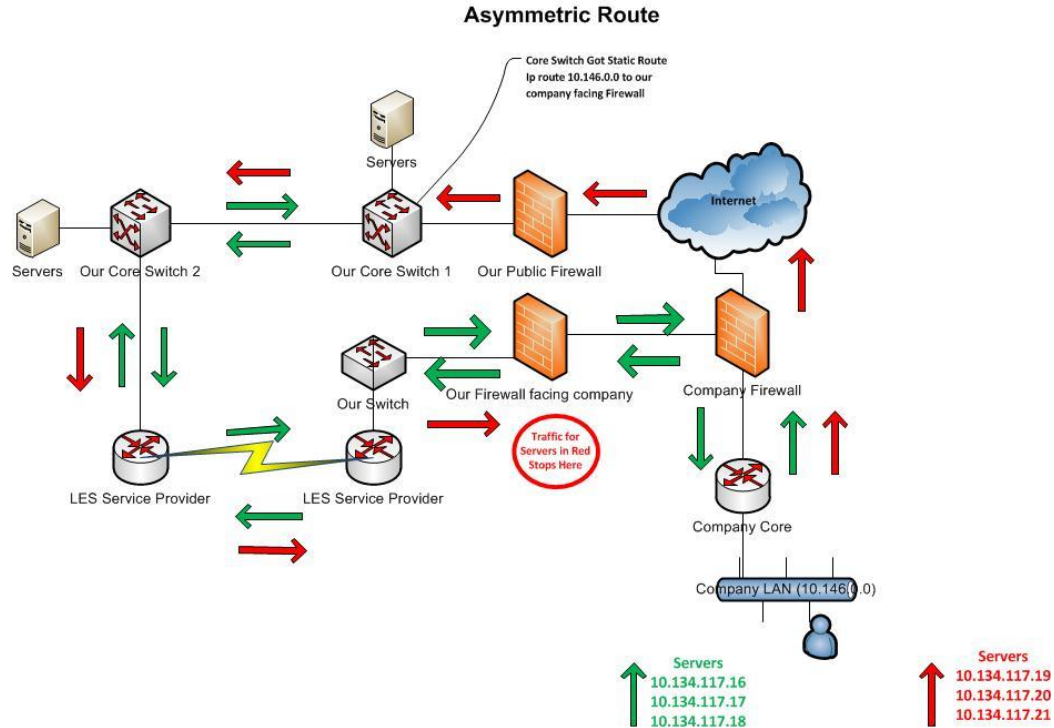
The Internet is a network of networks...

BGP: Finding a path for your packets



Source: https://www.researchgate.net/figure/BGP-routing-table-example_fig7_34930048

Asymmetric Routing





Break

Types of internet measurements



Passive measurement...

Passive measurement

- We just listen in somewhere
 - At your ISP
 - Your university
 - ...

Demo

~~Local WiFi traffic~~
Local Network Traffic

A cartoon character with a yellow face, brown hair, and a mustache is lying on a green surface, possibly grass. He has a wide-eyed, panicked expression with one eye visible and the other obscured by a blue, cracked lens. His mouth is open in a scream, showing a red tongue. He is wearing green pants and yellow shoes. The background is dark blue with some purple spots.

Active measurement...

Active measurement...

- We do something to solicit a (measurable) response
- Example: Traceroutes and RTT measurements, Portscans

Demo

Portscan and RTT measurements

The background of the slide features a series of concentric circles or ripples emanating from a central point. The circles are composed of many thin, overlapping lines, creating a textured, almost hypnotic effect. The colors are primarily dark blues and greys, with a small, bright blue dot at the very center.

RTT (Round Trip Time)

RTT Measurements

- Send a packet and note down when you sent it
 - Ask the recipient to send a reply
 - Check time when the reply comes in
- > You know how long a round trip takes.

Implications for active (path) measurements

- You never know the return path of your packets
- Hence: We can only do an RTT
- We can guestimate 1-directional TT using a traceroute (by looking at the path RTTs)

TTL

- IP Packets come with a TTL field
- Practical default value for the Internet: 64
- Idea:
 - Each host decrements the TTL by 1
 - If the TTL reaches 0, the packet is discarded and the sender notified of the discard
- We can measure the path we take to a destination with it. How?

Traceroutes

```
Ubuntu 18.04 LTS
My traceroute [v0.92]
tud278692.tudelft.nl (131.180.98.151) 2020-02-19T11:23:30+0100
Keys: Help  Display mode  Restart statistics  Order of fields  quit
```

Host	Packets		Pings				
	Loss%	Snt	Last	Avg	Best	Wrst	StDev
1. 131.180.98.1	0.0%	58	0.3	2.3	0.3	43.2	7.0
2. 10.200.246.52	0.0%	58	0.7	3.0	0.4	45.1	8.7
3. 10.200.246.49	0.0%	58	0.4	0.4	0.3	0.5	0.0
4. 10.200.246.4	0.0%	58	28.7	1.9	0.4	28.7	5.0
5. 10.200.24.1	0.0%	58	0.8	1.3	0.6	28.7	3.7
6. 145.145.26.97	0.0%	57	23.9	1.7	0.6	23.9	3.8
7. 80.249.208.50	0.0%	57	1.7	3.3	1.7	27.4	4.6
8. 80.249.209.128	0.0%	57	2.5	2.7	2.3	5.4	0.7
9. 80.255.14.6	0.0%	57	8.1	7.9	7.7	8.3	0.1
10. 80.255.15.122	0.0%	57	7.7	15.0	7.6	80.5	16.7
11. 213.239.224.245	0.0%	57	12.3	13.8	12.2	22.3	2.5
12. 213.239.245.98	0.0%	57	12.0	12.1	12.0	13.8	0.3
13. 94.130.90.73	0.0%	57	12.0	12.1	12.0	12.3	0.1
14. 94.130.126.186	0.0%	57	12.2	12.6	12.1	13.1	0.3

Routers are not pingers

- Remote sites might rate-limit ICMP requests
- If you see packets dropping at 8.8.8.8 or at an intermediate router, it might just be that these hosts are currently seeing too many packets
- The actual services (google DNS, packet forwarding) take precedence over ICMP

Scanning

- ICMP (Like RTT checks for system liveness)
- Portscans
 - See if a port is reachable
- Banner Scans
 - See what is running on a port

Scanning

- Usually a multi-step process:
 - Check liveness (ICMP)
 - Check if ports are open
 - Do banner scans for open ports

Scanning

```
Ubuntu 18.04 LTS
tfiebig@shells ~ % nmap -T insane mail.aperture-labs.org
Starting Nmap 7.80 ( https://nmap.org ) at 2020-02-19 11:51 CET
Nmap scan report for mail.aperture-labs.org (94.130.126.186)
Host is up (0.00051s latency).
Other addresses for mail.aperture-labs.org (not scanned): 2a01:4f8:10b:37ef::186
Not shown: 992 closed ports
PORT      STATE SERVICE
22/tcp    open  ssh
25/tcp    open  smtp
80/tcp    open  http
443/tcp   open  https
465/tcp   open  smtps
587/tcp   open  submission
993/tcp   open  imaps
2000/tcp  open  cisco-sccp

Nmap done: 1 IP address (1 host up) scanned in 34.50 seconds
tfiebig@shells ~ %
```

Scanning (with banners etc.)

```
Ubuntu 18.04 LTS
Nmap done: 1 IP address (1 host up) scanned in 34.50 seconds
friebig@shells: ~ % reset
friebig@shells: ~ % nmap -A -T insane mail.aperture-labs.org
Starting Nmap 7.80 ( https://nmap.org ) at 2020-02-19 12:22 CET
Nmap scan report for mail.aperture-labs.org (94.130.126.186)
Host is up (0.00049s latency).
Other addresses for mail.aperture-labs.org (not scanned): 2a01:4f8:10b:37ef::186
Not shown: 992 closed ports
PORT      STATE SERVICE
22/tcp    open  ssh      OpenSSH 8.1 (protocol 2.0)
| ssh-hostkey:
|   2048 2e:2b:9a:8c:a8:4f:8e:ad:dc:ba:b5:cf:7b:ba:37:bd (RSA)
|   256 fb:8a:5f:0e:d4:6c:a6:c8:45:31:1a:e1:a1:1c:34:5f (ECDSA)
|   256 5b:a6:f8:e6:d2:4b:c5:c5:d4:64:78:19:d8:44:7a:92 (ED25519)
23/tcp    open  smtp      Postfix smtpd
| smtp-commands: mail.aperture-labs.org, PIPELINING, SIZE 20480000, VRFY, ETRN, STARTTLS, ENHANCEDSTATUSCODES, 8BITMIME, DSN, CHUNKING,
| ssl-cert: Subject: commonName=mail.aperture-labs.org
| Subject Alternative Name: DNS:mail.aperture-labs.org
| Not valid before: 2019-12-15T00:31:08
| Not valid after: 2020-03-14T00:31:08
| ssl-date: TLS randomness does not represent time
80/tcp    open  http      nginx
| http-title: Did not follow redirect to https://mail.aperture-labs.org/
443/tcp    open  ssl/http  nginx
| http-title: 50Go/
| Requested resource was /50Go/
| ssl-cert: Subject: commonName=mail.aperture-labs.org
| Subject Alternative Name: DNS:mail.aperture-labs.org
| Not valid before: 2019-12-15T00:31:08
| Not valid after: 2020-03-14T00:31:08
| ssl-date: TLS randomness does not represent time
| tls-align:
|   http/1.1
465/tcp    open  ssl/smtp  Postfix smtpd
| smtp-commands: mail.aperture-labs.org, PIPELINING, SIZE 20480000, VRFY, ETRN, AUTH PLAIN, AUTH=PLAIN, ENHANCEDSTATUSCODES, 8BITMIME, DSN, CHUNKING,
| ssl-cert: Subject: commonName=mail.aperture-labs.org
| Subject Alternative Name: DNS:mail.aperture-labs.org
| Not valid before: 2019-12-15T00:31:08
| Not valid after: 2020-03-14T00:31:08
| ssl-date: TLS randomness does not represent time
587/tcp    open  smtp      Postfix smtpd
| smtp-commands: mail.aperture-labs.org, PIPELINING, SIZE 20480000, VRFY, ETRN, STARTTLS, ENHANCEDSTATUSCODES, 8BITMIME, DSN, CHUNKING,
| ssl-cert: Subject: commonName=mail.aperture-labs.org
| Subject Alternative Name: DNS:mail.aperture-labs.org
| Not valid before: 2019-12-15T00:31:08
| Not valid after: 2020-03-14T00:31:08
| ssl-date: TLS randomness does not represent time
993/tcp    open  ssl/imap  Dovecot imapd
| imap-capabilities: have IDLE ID LOGIN-REFERRALS ENABLE more post-login AUTH=PLAINAUTH=PLAIN Pre-login capabilities SASL-IR listed OK LITERAL+ IMAP4rev1
| ssl-cert: Subject: commonName=mail.aperture-labs.org
| Subject Alternative Name: DNS:mail.aperture-labs.org
| Not valid before: 2019-12-15T00:31:08
| Not valid after: 2020-03-14T00:31:08
| ssl-date: TLS randomness does not represent time
2000/tcp   open  sieve     Dovecot Pigeonhole sieve 1.0
Service Info: Host: mail.aperture-labs.org
Service detection performed. Please report any incorrect results at https://nmap.org/submit/ .
Nmap done: 1 IP address (1 host up) scanned in 62.88 seconds
friebig@shells: ~ %
```

State in scanning

- If we wait for every host to reply, then check ports, then check banners, things get slow
- This works for a /24, but not for 0.0.0.0/0
- Tools like zMap abandon state and just put out packets
- If doing this, be careful what your machine/network/middleboxes can handle (see later)

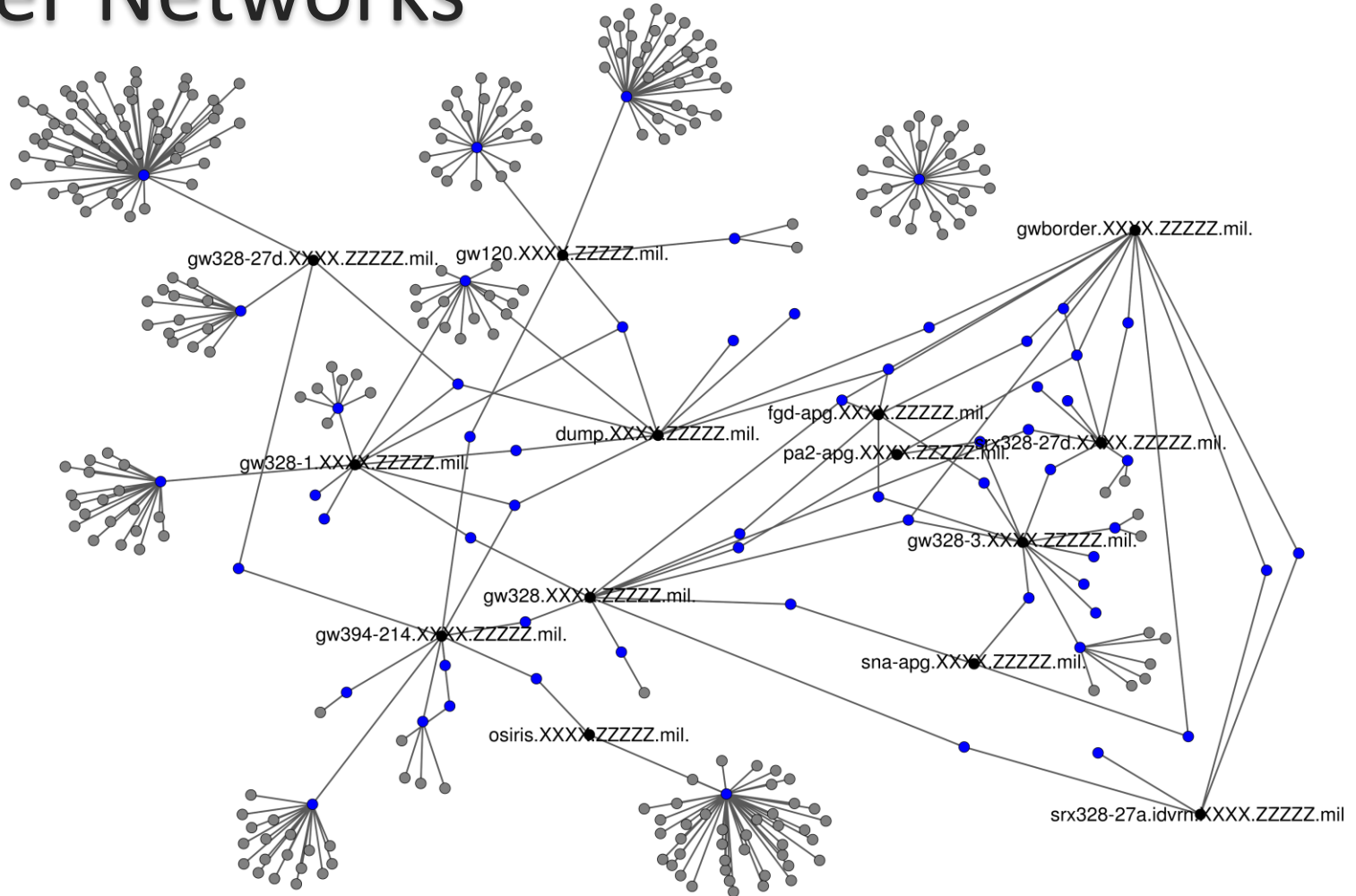
Scanning IPv6

- The 2^{32} addresses in IPv4 are easy
- The 2^{128} in IPv6 are *hard*
- Solution:
 - Generate hitlists
 - Passive measurements
 - Active measurements (DNS/rDNS)
 - Generate scan-targets from hitlists using ‘ML’

Enumerating IPv6 Reverse DNS

- RFC 8020 says 'NXDOMAIN means nothing there or anywhere thereunder'
- We start with querying, e.g.:
2.ip6.arpa.
- Get a NOERROR instead of NXDOMAIN
 - We know there is something there below, so we can continue with 0.2.ip6.arpa.
- Allows us to prune the search-tree

Computer Networks



Finding the AS to an IP

- Asking the RIRs (you should know this from TB321 (Internet governance lecture))
- Run whois services + there are some collectors
- Under Linux: whois \$IP

```
t fiebig@TUD255721 ~ % whois 195.191.196.23
% This is the RIPE Database query service.
% The objects are in RPSL format.
%
% The RIPE Database is subject to Terms and Conditions.
% See http://www.ripe.net/db/support/db-terms-conditions.pdf
%
% Note: this output has been filtered.
%       To receive output for a database update, use the "-B" flag.
%
% Information related to '195.191.196.0 - 195.191.197.255'
%
% Abuse contact for '195.191.196.0 - 195.191.197.255' is 'tobias+abuse@fiebig.nl'
%
inetnum:        195.191.196.0 - 195.191.197.255
netname:        WYBT-NET
remarks:        WYBT-NET assigned PI Space
country:        DE
org:            ORG-wA159-RIPE
admin-c:        WYBT-RIPE
tech-c:         WYBT-RIPE
status:         ASSIGNED PI
```

```
origin:      AS51827
mnt-by:      WYBT-MNT
mnt-by:      LWL-MNT
created:     2016-09-24T12:31:22Z
last-modified: 2016-09-24T12:31:22Z
source:      RIPE
```

% Information related to '195.191.196.0/23AS51827'

```
route:       195.191.196.0/23
descr:       wybt.net via AS51827/Fremaks
origin:      AS51827
mnt-by:      FRMA-MNT
mnt-by:      WYBT-MNT
created:     2013-12-09T11:55:54Z
last-modified: 2013-12-09T11:59:07Z
source:      RIPE
```

% This query was served by the RIPE Database Query Service version 1.92.6 (ANGUS)

Whois data on ASes

- Not always well maintained
- Lengthy to work with...

Looking glasses

- Famous one: <https://bgp.he.net/>
- Check which networks are announced by which AS
- (Demo)

Finding an AS for a company

- Google
- <https://apps.db.ripe.net/db-web-ui/#/fulltextsearch>
- Also: <https://bgp.he.net/>
- And: <https://ipinfo.io/countries/nl>

GRT

- The GRT (global routing table) is the ‘combined knowledge’ of all BGP routers out there
- It looks different depending on from ‘where’ you look at it
- Different projects aggregate routeviews across the Internet to give you a look at global topology.
 - RIPE-RIS, Routeviews

Practical issues of measurements

Time

- Time is a crucial element
- Deviations as small as a second might make you unable to correlate datasets
- Clocks on computers will not always be accurate

Fixing time

- Bet way forward:
 - Use NTP (Network Time Protocol) to sync time between your nodes
 - Special network measurement cards come with a dedicated time-sync feature (to ensure internal consistency)
 - If you can not NTP sync, you need an RTC

Timezones

- The world is round
- Turns out: This means we do not have the same Time-of-Day everywhere
- This is a great source for confusion
- Make sure to record the time-stamps and the time-zones in which you collected data along with it!
- Best practice: Use UTC. Everywhere.

Storage needs

- Internet measurement data is usually big
- In really common projects you will see dataset sizes of 100s of GB/day or even hour
- Plan your storage needs ahead of time
 - Prevents: Dropping data during the project
- Keep in mind that you also need space for analyzing data!

High PPS (Packets per Second)

- When you are measuring data, you will inevitable run into problems where the link is faster than you can capture
- This can also happen when sending many small packets (being unable to handle the returns; zMap)

Handling high PPS measurements

- There are dedicated measurement cards for these kinds of workloads (e.g., endace) who have their own on-card buffer to allow for, e.g., 100gE line-rate small PPS measurements
- Maybe limit your outbound PPS 😊

You building it right does not mean that it works

- Monitor your measurements
- Make sure they keep running and recording data
- Make sure they record the right data
- Make sure to also save intermediate state.
 - There is nothing more frustrating than running a 1 week measurement job, where you throw away intermediate data, and had a typo in your final aggregation.
- Have some kind of progress/state indicator

Vantage points

- In network measurement, we call the place (from) where we measure a vantage point
- In passive measurements, this is the place where we snoop data
- In active measurements, this is the place from where we originate requests

Question

Does the vantage point matter for active measurements?

Vantage points matter

- Our view on the Internet might be different, depending on from where we measure
 - BGP views/paths will differ.
 - RTT to a service in SFO is a lot quicker from LAX than from AMS
 - DNS often has location dependent replies
 - Anycast is a thing

Vantage points matter

- There will be things 'special' about our local network
- These can have an impact on our measurements

8.8.8.8 from Hetzner

```
Ubuntu 18.04 LTS
My traceroute [v0.93]
shells.aperture-labs.org (94.130.126.189) 2020-02-19T13:39:39+0100
Keys: Help Display mode Restart statistics Order of fields quit
```

Host	Packets		Pings				
	Loss%	Snt	Last	Avg	Best	Wrst	StDev
1. 94.130.126.185	0.0%	12	19.9	9.0	0.1	20.0	9.7
2. 94.130.90.65	0.0%	12	0.7	0.6	0.4	0.7	0.1
3. 213.239.245.89	0.0%	12	0.8	0.7	0.5	1.0	0.1
4. 213.239.252.29	0.0%	12	5.2	6.8	5.1	21.6	4.7
5. 72.14.218.176	0.0%	12	5.2	5.9	5.0	13.3	2.3
6. 108.170.251.193	0.0%	12	5.2	5.2	5.1	5.3	0.1
7. 108.170.235.249	0.0%	11	5.2	5.2	5.1	5.3	0.1
8. 8.8.8.8	0.0%	11	5.1	18.5	5.0	153.1	44.6

8.8.8.8 from TU Delft

```
Ubuntu 18.04 LTS
```

My traceroute [v0.92]

tud278692.tudelft.nl (131.180.98.151) 2020-02-19T13:40:42+0100

Keys: Help Display mode Restart statistics Order of fields quit

Host	Packets		Pings				
	Loss%	Snt	Last	Avg	Best	Wrst	StDev
1. 131.180.98.1	0.0%	21	0.5	3.5	0.4	23.7	7.8
2. 10.200.246.52	0.0%	21	0.7	75.3	0.4	358.7	116.2
3. 10.200.246.49	0.0%	21	0.4	0.4	0.3	0.7	0.1
4. 10.200.246.4	0.0%	21	0.5	16.3	0.5	176.4	41.4
5. 10.200.24.1	0.0%	21	0.7	0.8	0.7	1.5	0.2
6. 145.145.26.97	0.0%	21	0.9	0.9	0.6	1.9	0.3
7. 74.125.51.223	0.0%	20	10.2	7.2	1.7	20.4	6.1
8. 74.125.51.222	0.0%	20	1.8	1.9	1.7	2.8	0.2
9. 108.170.241.193	0.0%	20	1.8	1.8	1.7	2.0	0.1
10. 172.253.66.185	0.0%	20	2.2	3.0	2.1	9.7	1.6
11. 8.8.8.8	0.0%	20	1.7	1.7	1.6	1.7	0.0

DNS Interception

```
Ubuntu 18.04 LTS
tfiebig@tardis ~ % host mail.aperture-labs.org
mail.aperture-labs.org has address 94.130.126.186
tfiebig@tardis ~ % dig AAAA mail.aperture-labs.org

; <<>> DiG 9.11.3-1ubuntu1.11-Ubuntu <<>> AAAA mail.aperture-labs.org
;; global options: +cmd
;; Got answer:
;; ->>HEADER<<- opcode: QUERY, status: NOERROR, id: 61176
;; flags: qr rd ra; QUERY: 1, ANSWER: 0, AUTHORITY: 6, ADDITIONAL: 7

;; OPT PSEUDOSECTION:
; EDNS: version: 0, flags;; udp: 4096
;; QUESTION SECTION:
;mail.aperture-labs.org.                IN      AAAA

;; AUTHORITY SECTION:
org.                168674 IN      NS      d0.org.afiliast-nst.org.
org.                168674 IN      NS      a2.org.afiliast-nst.info.
org.                168674 IN      NS      a0.org.afiliast-nst.info.
org.                168674 IN      NS      b2.org.afiliast-nst.org.
org.                168674 IN      NS      c0.org.afiliast-nst.info.
org.                168674 IN      NS      b0.org.afiliast-nst.org.

;; ADDITIONAL SECTION:
a0.org.afiliast-nst.info. 168674 IN      A       199.19.56.1
a2.org.afiliast-nst.info. 168674 IN      A       199.249.112.1
b0.org.afiliast-nst.org. 168674 IN      A       199.19.54.1
b2.org.afiliast-nst.org. 168674 IN      A       199.249.120.1
c0.org.afiliast-nst.info. 168674 IN      A       199.19.53.1
d0.org.afiliast-nst.org. 168674 IN      A       199.19.57.1
```

DNS Interception

```
Ubuntu 18.04 LTS
tfiebig@tardis ~ % ssh shells.aperture-labs.org
Last login: Wed Feb 19 11:28:27 2020 from 145.94.42.199
OpenBSD 6.6 (GENERIC.MP) #372: Sat Oct 12 10:56:27 MDT 2019
tfiebig@shells ~ % host mail.aperture-labs.org
mail.aperture-labs.org has address 94.130.126.186
mail.aperture-labs.org has IPv6 address 2a01:4f8:10b:37ef::186
tfiebig@shells ~ % dig AAAA mail.aperture-labs.org

; <<>> DiG 9.4.2-P2 <<>> AAAA mail.aperture-labs.org
;; global options: printcmd
;; Got answer:
;; ->>HEADER<<- opcode: QUERY, status: NOERROR, id: 18542
;; flags: qr rd ra; QUERY: 1, ANSWER: 1, AUTHORITY: 0, ADDITIONAL: 0

;; QUESTION SECTION:
;mail.aperture-labs.org.          IN      AAAA

;; ANSWER SECTION:
mail.aperture-labs.org. 291     IN      AAAA    2a01:4f8:10b:37ef::186

;; Query time: 1 msec
;; SERVER: 2a01:4f8:10b:37ef::53#53(2a01:4f8:10b:37ef::53)
;; WHEN: Wed Feb 19 13:00:36 2020
;; MSG SIZE rcvd: 68

tfiebig@shells ~ %
```

Funny Middleboxes

```
Ubuntu 18.04 LTS
tfiebig@tardis ~ % ssh ws.tud.fiebig.nl
Last login: Wed Feb 19 11:22:23 2020 from 131.180.123.197
tfiebig@tud278692 ~ % nmap -T insane www.aperture-labs.org
Starting Nmap 7.70 ( https://nmap.org ) at 2020-02-19 13:32 CET
Warning: 94.130.126.187 giving up on port because retransmission cap hit (2).
Nmap scan report for www.aperture-labs.org (94.130.126.187)
Host is up (0.012s latency).
Not shown: 527 filtered ports, 469 closed ports
PORT      STATE SERVICE
22/tcp    open  ssh
80/tcp    open  http
443/tcp    open  https
8008/tcp   open  http

Nmap done: 1 IP address (1 host up) scanned in 4.23 seconds
tfiebig@tud278692 ~ %
```

Funny Middleboxes

```
Ubuntu 18.04 LTS
tfiebig@tardis ~ % ssh shells.aperture-labs.org
Last login: Wed Feb 19 12:00:22 2020 from 145.94.42.199
OpenBSD 6.6 (GENERIC.MP) #372: Sat Oct 12 10:56:27 MDT 2019
tfiebig@shells ~ % nmap -T insane www.aperture-labs.org
Starting Nmap 7.80 ( https://nmap.org ) at 2020-02-19 13:32 CET
Nmap scan report for www.aperture-labs.org (94.130.126.187)
Host is up (0.00051s latency).
Other addresses for www.aperture-labs.org (not scanned): 2a01:4f8:10b:37ef::187
Not shown: 997 closed ports
PORT      STATE SERVICE
22/tcp    open  ssh
80/tcp    open  http
443/tcp    open  https

Nmap done: 1 IP address (1 host up) scanned in 37.57 seconds
tfiebig@shells ~ %
```

Funny Middleboxes

```
Ubuntu 18.04 LTS
tfiebig@tardis ~ % ssh ws.tud.fiebig.nl
Last login: Wed Feb 19 11:22:23 2020 from 131.180.123.197
tfiebig@tud278692 ~ % nmap -T insane www.aperture-labs.org
Starting Nmap 7.70 ( https://nmap.org ) at 2020-02-19 13:32 CET
Warning: 94.130.126.187 giving up on port because retransmission cap hit (2).
Nmap scan report for www.aperture-labs.org (94.130.126.187)
Host is up (0.012s latency).
Not shown: 527 filtered ports, 469 closed ports
PORT      STATE SERVICE
22/tcp    open  ssh
80/tcp    open  http
443/tcp    open  https
3008/tcp  open  http

Nmap done: 1 IP address (1 host up) scanned in 4.23 seconds
tfiebig@tud278692 ~ %
```

Funny Middleboxes

- First problem is TU Delft's DNS servers filtering all AAAA/IPv6 related queries
- Second problem is TU Delft using a standard censorship/user-monitoring system to observe users' web-access for security reasons
- Middleboxes/firewalls may also filter things

Blowing up state

- Firewalls/Middleboxes are often stateful
- Means: For each packet/connection they record state
- These state-tables have a limited size
- If one client, e.g., zMap's the Internet, the state table blows up and the network goes down

In general: Watch out for bottlenecks

- Your local DNS cache might get overloaded by your queries
- An upstream network provider might get congested
- The site you are measuring (when web-scraping) might go down
- Etc.

Being a good netizen

Six rules to less hassle...

1. Rule: If you do measurements, involve you HREC

- Collecting data (active or passive) always has ethical constraints with it
- Involve your Human Research Ethics Council and get their approval
- If you do not have one, follow the Menlo report
 - https://en.wikipedia.org/wiki/Menlo_Report

2. Rule: Ensure data is protected

- FDE on research devices (especially if mobile)
- Proper anonymization if you share data
- Proper wiping of old disks etc.

3. Rule: Involve your local IT

- Internet measurements may impact your local network
- Tell your IT what you do, and get approval from them
- If they know what you are doing, they will be less pissed if you break something
- Also: They are the best ones to tell you what kind of funny middle boxes they run

4. Rule: Provide information on your measurement systems

- Set an informative reverse DNS entry on the involved hosts
- Have a website on these hosts explaining your research, how to opt out (see next slide), whom to contact in case of problems etc. Have the website on what the name resolves to as well

5. Rule: Provide an opt-out mechanism

- People will not like it, if you scan their network
- Integrate a mechanism in your process, which allows them to 'opt-out', i.e., be no longer visited by your scans in the future
- Make sure that mechanism works

6. Rule: Be nice to the network

- You will cause load on networks (local/remote)
- Make sure you do not hurt smaller networks due to too many packets/bandwidth being inbound
- If you are doing higher-level protocol scans, other load considerations and bottlenecks apply!