The Opaque Internet

Jon Crowcroft, 17.8.2023

https://www.cst.cam.ac.uk/people/jac22

TCP/IP Headers as we used to see them

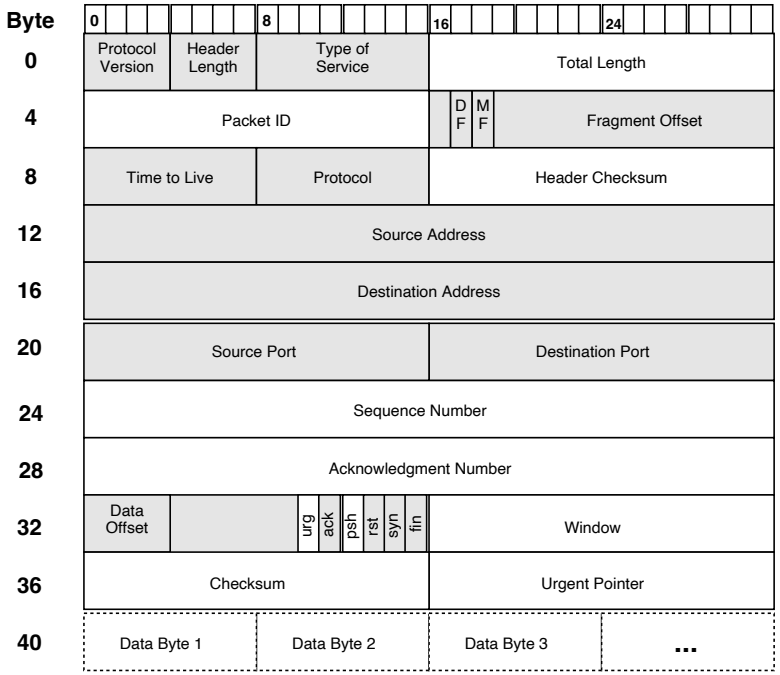


Figure 3: Fields that change during a TCP connection

Old school

- Nats (port NATs), firewalls, Carrier Grad Nats
- ToR
- Proxies/caches
- Middle Boxes (“accelerators”) (early ack, muck with window etc)
- CDNs, Load Balancers

- And of course HTTPs (TLS)….means we only see IP and ports maybe

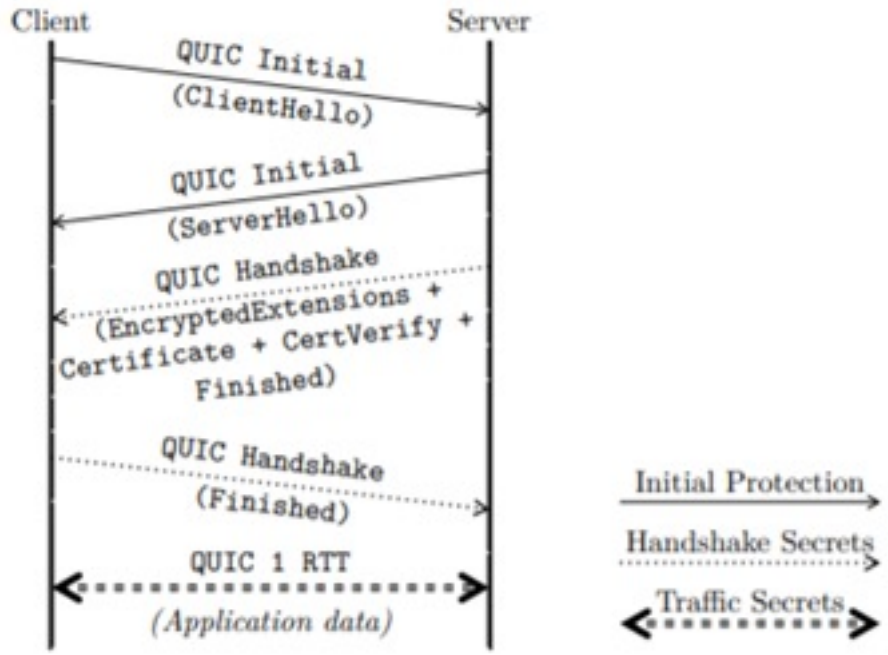
“new” kids

- Akamai, Cloudflare
- E.g. see

  https://radar.cloudflare.com/

- Books are out of date☺
  - E.g. lots of QUIC, Masque, one-hop relay etc etc
  - TCP is only about half of it these days
  - IPv6 addr alloc is weird
  - Before you even get there you have DOT
    - https://www.cloudflare.com/learning/dns/dns-over-tls/

QUIC (on UDP…) exchanges..

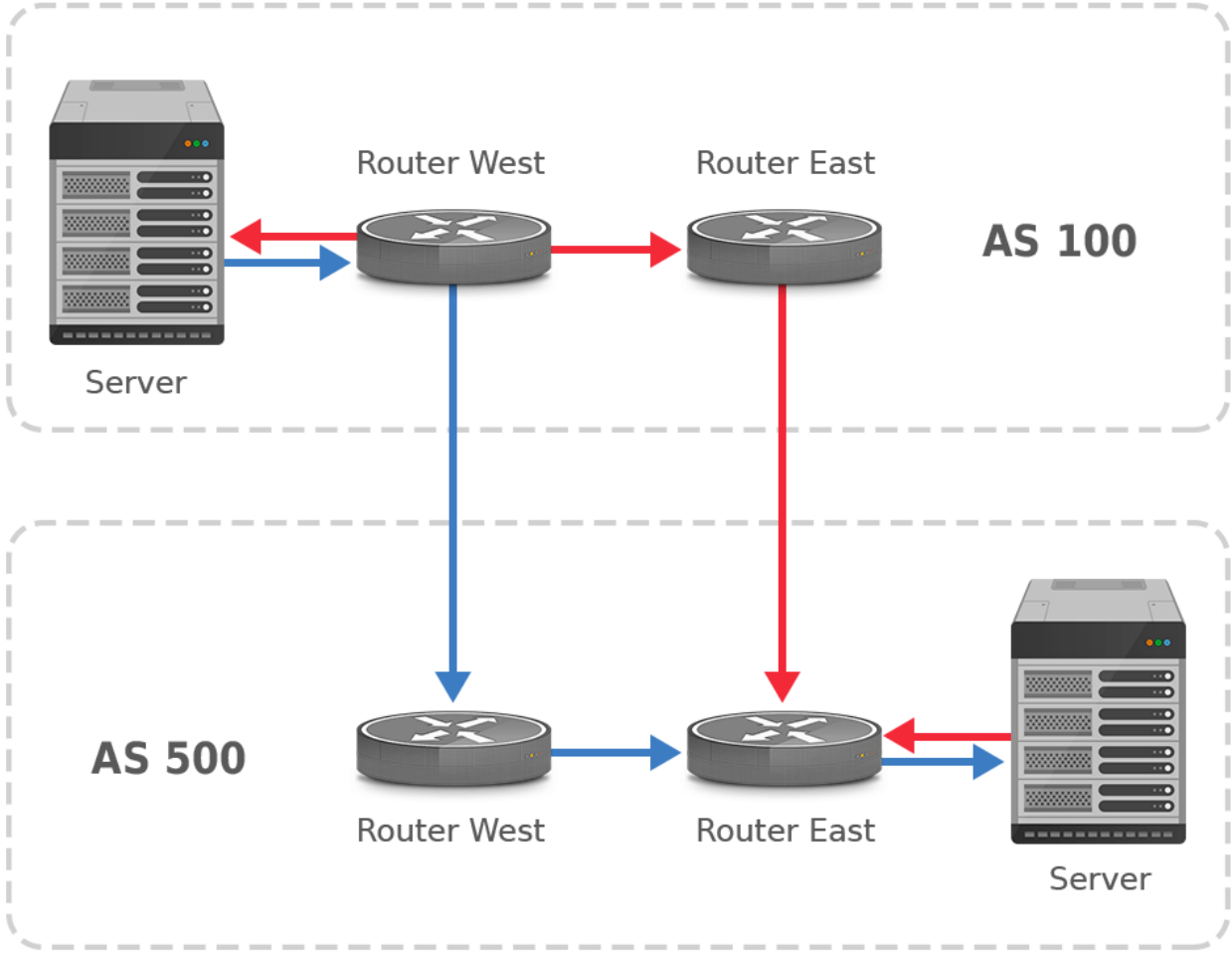


Challenges #1

- BGP – routeviews
- Can’t do from single vantage point…
- Lots of tools to detect hijacks etc, but depend on
  - Deploying collectors is a Big Effort
  - Luckily, lots of people have done this
  - Look for their resources/repositories of data too!

Asymmetric routes are v. common

- So outbound isn’t same as
- Inbound☺



Challenges #2

- Performance measurement is hard
  - packet trains, etc – need to be cleverer
  - ping mesh won’t detect topology
  - Embeddings, layer 2 segments etc

Challenges #3

- Censorship isn’t binary (any more)
  - Partly as DPI doesn’t work when most stuff is crypted (TLS/QUIC)
  - Lots of in-flight packet modification near edge
  - Where people go via cache/proxy/loadbalancer
  - Or where service has a plain (IM/zoom/teams/skype/jitsi mixers) hub

Challenges #4

- Adversaries to measurement
- Inject false responses
- Deep six your measurement traffic
- How do they know?
  - Flow signatures….subtle (ML based sometimes)
  - Sometimes just look at very simple things is best
  - Packet header fields (if plain)

Challenges #5

- Care of ethics (negative impact on performance&privacy)
- Some viewpoints are not safe to use…and may endanger others –
  - see this essential reading:-
  Ethical Concerns for Censorship Measurement
  https://conferences.sigcomm.org/sigcomm/2015/pdf/papers/nsethics/p17.pdf
  - Your measurement may reduce the performance someone needs (and paid for) – these folks take extreme care not to do that
  https://availability.samknows.com/broadband/

Re-decentralised

- Federated stuff is hard to measure
- Tor, Mastodon, Matrix etc
- Intentionally so (for good reasons)
    - Avoid censorship
    - Avoid state surveillance (e.g. whistleblower or NGO)
    - Avoid untrustworthy systems at all

Conclusions

- See
  https://conferences.sigcomm.org/imc/2022/paper-access/
    (or any other year☺
  https://www.codebgp.com/about/
    (recent Greek startup bought by cisco☺
  https://www.routeviews.org/routeviews/