Ollscoil na Gaillimhe
University of Galway

CT 420
Real-Time Systems

Logging, Debugging and Visualization of QUIC Traffic

Dr. Jawad Manzoor
Assistant Professor
School of Computer Science

Contents

- Traffic Analysis using Wireshark
- qlog and qviz
- Visualization Case Studies

Motivation

Ollscoil na Gaillimhe
University of Galway

- There are many things that can go wrong during network communication that can lead to sub-optimal performance of your web application.
- Logging, debugging and visualizations are used to analyze the protocols and find root cause of the problems.
- For TCP, the most commonly used method is packet capture.
  - Analyze pcap files in Wireshark
- For QUIC, newer methods are developed recently.
  - qlog
  - qviz

Case Study 1

- Client experiencing slower speeds on QUIC as compared to TCP.
  - Analyze the network traffic to find the root cause.
  - Use cURL to download the webpage and capture the network traffic

- cURL is a command line tool that developers use to transfer data to and from a server.
  - It is compatible with almost every operating system and connected device.
  - It is useful for testing endpoints.
  - It has HTTP3 support

Demo

Ollscoil na Gaillimhe
University of Galway

- Prerequisites
  - Install docker desktop
  - Get curl-http3 docker file from the GitHub repo
  - https://github.com/rmarx/curl-http3
  - Build docker image

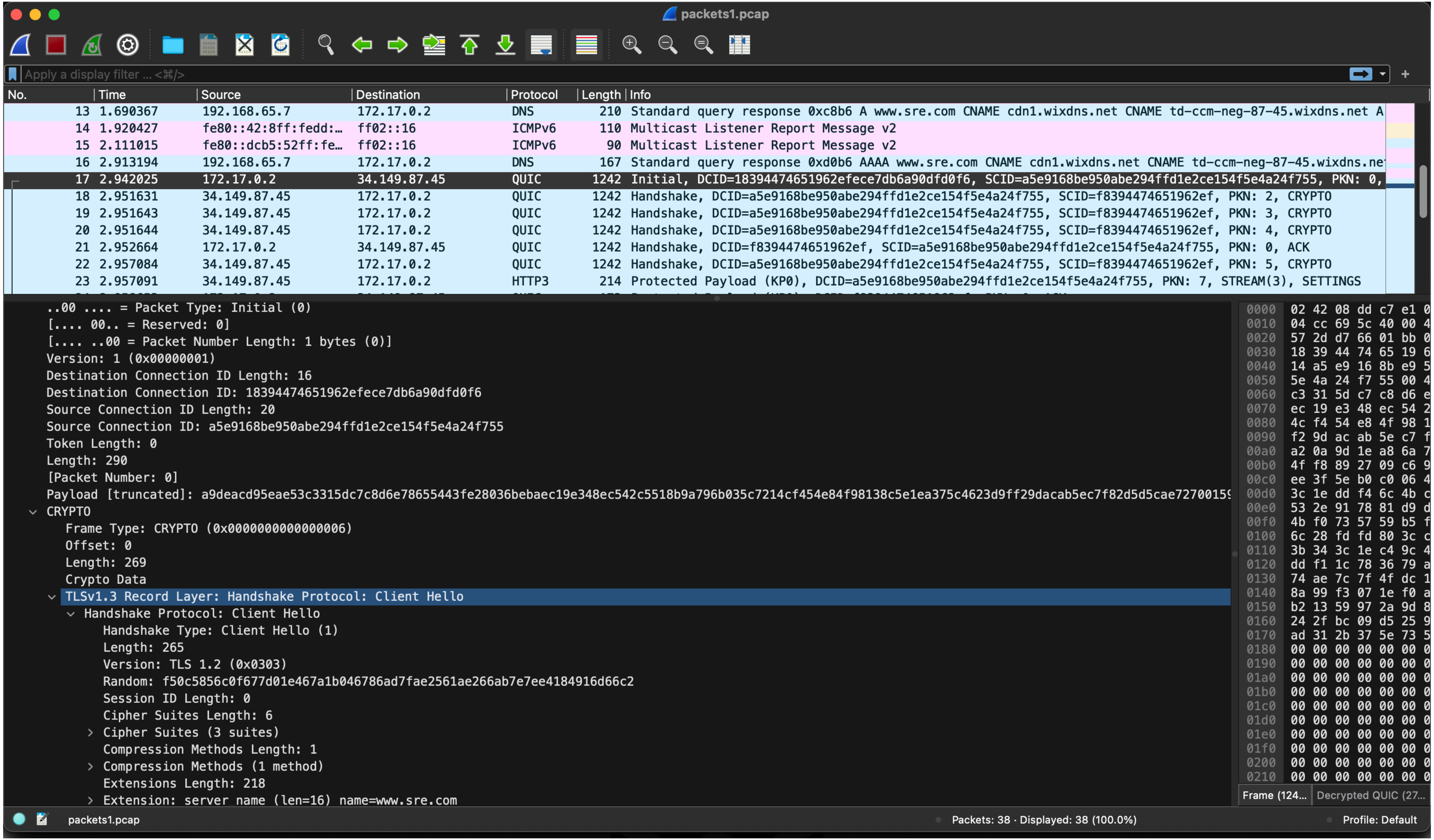
- Runs a container in interactive mode (-it).
- Mounts a directory (pcaps_on_host) on the host to /srv in the container.
- Logs QUIC events (qlog) to /srv.
- Logs TLS keys to /srv/tls_keys.txt for decrypting HTTPS traffic.

```
docker run -it --rm
   --volume $(pwd)/pcaps_on_host:/srv
   --env QLOGDIR=/srv
   --env SSLKEYLOGFILE=/srv/tls_keys.txt

rmarx/curl-http3

   bash -c "tcpdump -w /srv/packets.pcap -i eth0 & sleep 1; curl -IL
   https://www.sre.com --http3;
   sleep 2; pkill tcpdump; sleep 2"
```

- Captures HTTP/3 (QUIC) network traffic for analysis.
- The .pcap file can be opened in Wireshark to inspect HTTP/3 behavior.
- Useful for debugging HTTP/3 connectivity issues.


packets1.pcap
Apply a display filter ... <⌘/>
No. | Time | Source | Destination | Protocol | Length | Info
13 1.690367 192.168.65.7 172.17.0.2 DNS 210 Standard query response 0xc8b6 A www.sre.com CNAME cdn1.wixdns.net CNAME td-ccm-neg-87-45.wixdns.net A
14 1.920427 fe80::42:8ff:fedd:… ff02::16 ICMPv6 110 Multicast Listener Report Message v2
15 2.111015 fe80::dcb5:52ff:fe… ff02::16 ICMPv6 90 Multicast Listener Report Message v2
16 2.913194 192.168.65.7 172.17.0.2 DNS 167 Standard query response 0xd0b6 AAAA www.sre.com CNAME cdn1.wixdns.net CNAME td-ccm-neg-87-45.wixdns.ne
17 2.942025 172.17.0.2 34.149.87.45 QUIC 1242 Initial, DCID=18394474651962efece7db6a90dfd0f6, SCID=a5e9168be950abe294ffd1e2ce154f5e4a24f755, PKN: 0,
18 2.951631 34.149.87.45 172.17.0.2 QUIC 1242 Handshake, DCID=a5e9168be950abe294ffd1e2ce154f5e4a24f755, SCID=f8394474651962ef, PKN: 2, CRYPTO
19 2.951643 34.149.87.45 172.17.0.2 QUIC 1242 Handshake, DCID=a5e9168be950abe294ffd1e2ce154f5e4a24f755, SCID=f8394474651962ef, PKN: 3, CRYPTO
20 2.951644 34.149.87.45 172.17.0.2 QUIC 1242 Handshake, DCID=a5e9168be950abe294ffd1e2ce154f5e4a24f755, SCID=f8394474651962ef, PKN: 4, CRYPTO
21 2.952664 172.17.0.2 34.149.87.45 QUIC 1242 Handshake, DCID=f8394474651962ef, SCID=a5e9168be950abe294ffd1e2ce154f5e4a24f755, PKN: 0, ACK
22 2.957084 34.149.87.45 172.17.0.2 QUIC 1242 Handshake, DCID=a5e9168be950abe294ffd1e2ce154f5e4a24f755, SCID=f8394474651962ef, PKN: 5, CRYPTO
23 2.957091 34.149.87.45 172.17.0.2 HTTP3 214 Protected Payload (KP0), DCID=a5e9168be950abe294ffd1e2ce154f5e4a24f755, PKN: 7, STREAM(3), SETTINGS
..00 .... = Packet Type: Initial (0)
[.... 00.. = Reserved: 0]
[.... ..00 = Packet Number Length: 1 bytes (0)]
Version: 1 (0x00000001)
Destination Connection ID Length: 16
Destination Connection ID: 18394474651962efece7db6a90dfd0f6
Source Connection ID Length: 20
Source Connection ID: a5e9168be950abe294ffd1e2ce154f5e4a24f755
Token Length: 0
Length: 290
[Packet Number: 0]
Payload [truncated]: a9deacd95eae53c3315dc7c8d6e78655443fe28036bebaec19e348ec542c5518b9a796b035c7214cf454e84f98138c5e1ea375c4623d9ff29dacab5ec7f82d5d5cae7270015
CRYPTO
Frame Type: CRYPTO (0x0000000000000006)
Offset: 0
Length: 269
Crypto Data
TLSv1.3 Record Layer: Handshake Protocol: Client Hello
Handshake Protocol: Client Hello
Handshake Type: Client Hello (1)
Length: 265
Version: TLS 1.2 (0x0303)
Random: f50c5856c0f677d01e467a1b046786ad7fae2561ae266ab7e7ee4184916d66c2
Session ID Length: 0
Cipher Suites Length: 6
Cipher Suites (3 suites)
Compression Methods Length: 1
Compression Methods (1 method)
Extensions Length: 218
Extension: server_name (len=16) name=www.sre.com
Frame (124… Decrypted QUIC (27…
packets1.pcap
Packets: 38 · Displayed: 38 (100.0%)
Profile: Default

packets1.pcap

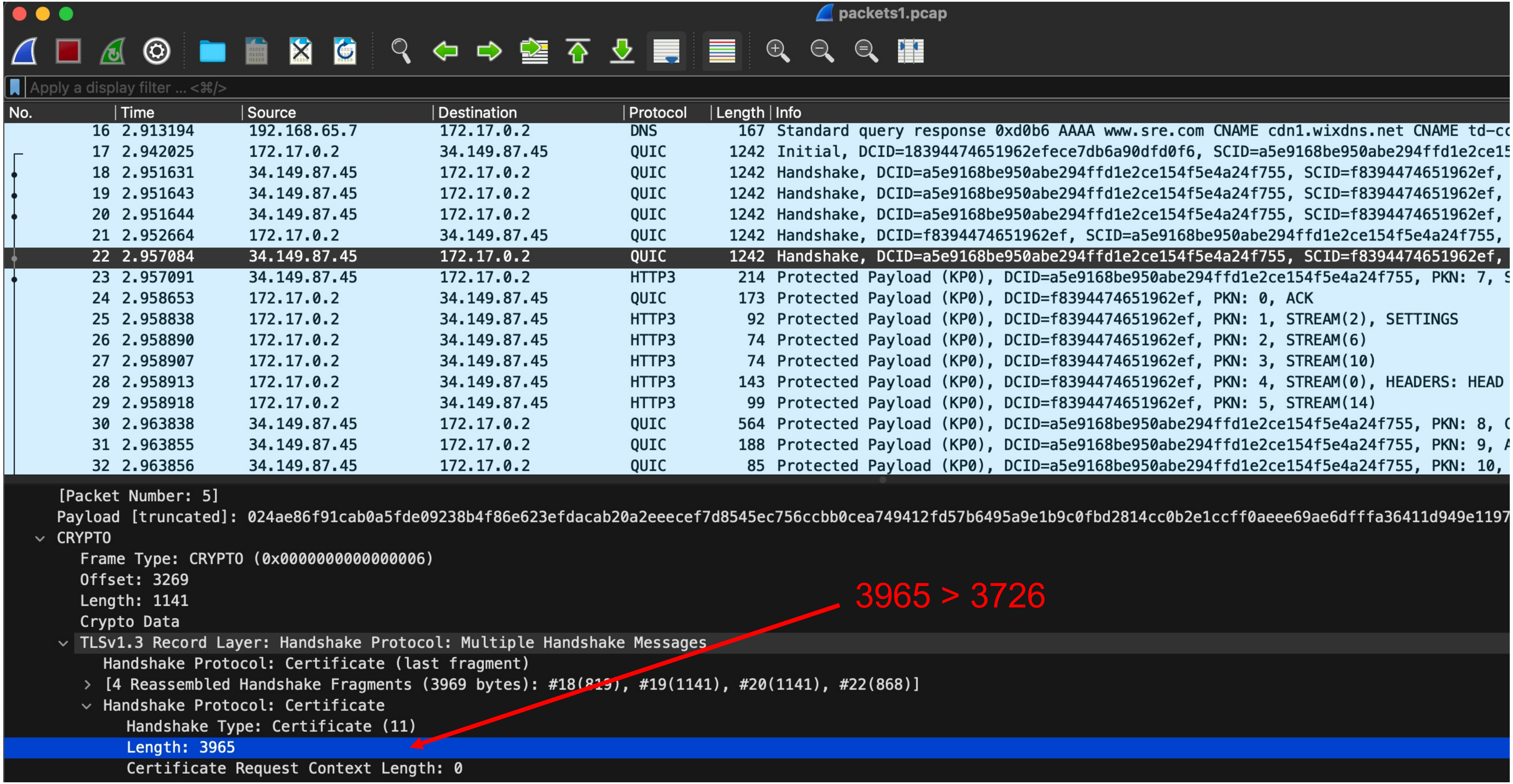
Apply a display filter ... <⌘/>

No.	Time	Source	Destination	Protocol	Length	Info
16	2.913194	192.168.65.7	172.17.0.2	DNS	167	Standard query response 0xd0b6 AAAA www.sre.com CNAME cdn1.wixdns.net CNAME td-ccm-neg-87-45.wixdns.ne
17	2.942025	172.17.0.2	34.149.87.45	QUIC	1242	Initial, DCID=18394474651962efece7db6a90dfd0f6, SCID=a5e9168be950abe294ffd1e2ce154f5e4a24f755, PKN: 0,
18	2.951631	34.149.87.45	172.17.0.2	QUIC	1242	Handshake, DCID=a5e9168be950abe294ffd1e2ce154f5e4a24f755, SCID=f8394474651962ef, PKN: 2, CRYPTO
19	2.951643	34.149.87.45	172.17.0.2	QUIC	1242	Handshake, DCID=a5e9168be950abe294ffd1e2ce154f5e4a24f755, SCID=f8394474651962ef, PKN: 3, CRYPTO
20	2.951644	34.149.87.45	172.17.0.2	QUIC	1242	Handshake, DCID=a5e9168be950abe294ffd1e2ce154f5e4a24f755, SCID=f8394474651962ef, PKN: 4, CRYPTO
21	2.952664	172.17.0.2	34.149.87.45	QUIC	1242	Handshake, DCID=f8394474651962ef, SCID=a5e9168be950abe294ffd1e2ce154f5e4a24f755, PKN: 0, ACK
22	2.957084	34.149.87.45	172.17.0.2	QUIC	1242	Handshake, DCID=a5e9168be950abe294ffd1e2ce154f5e4a24f755, SCID=f8394474651962ef, PKN: 5, CRYPTO
23	2.957091	34.149.87.45	172.17.0.2	HTTP3	214	Protected Payload (KP0), DCID=a5e9168be950abe294ffd1e2ce154f5e4a24f755, PKN: 7, STREAM(3), SETTINGS
24	2.958653	172.17.0.2	34.149.87.45	QUIC	173	Protected Payload (KP0), DCID=f8394474651962ef, PKN: 0, ACK
25	2.958838	172.17.0.2	34.149.87.45	HTTP3	92	Protected Payload (KP0), DCID=f8394474651962ef, PKN: 1, STREAM(2), SETTINGS
26	2.958890	172.17.0.2	34.149.87.45	HTTP3	74	Protected Payload (KP0), DCID=f8394474651962ef, PKN: 2, STREAM(6)
27	2.958907	172.17.0.2	34.149.87.45	HTTP3	74	Protected Payload (KP0), DCID=f8394474651962ef, PKN: 3, STREAM(10)
28	2.958913	172.17.0.2	34.149.87.45	HTTP3	143	Protected Payload (KP0), DCID=f8394474651962ef, PKN: 4, STREAM(0), HEADERS: HEAD /: HEAD /
29	2.958918	172.17.0.2	34.149.87.45	HTTP3	99	Protected Payload (KP0), DCID=f8394474651962ef, PKN: 5, STREAM(14)
30	2.963838	34.149.87.45	172.17.0.2	QUIC	564	Protected Payload (KP0), DCID=a5e9168be950abe294ffd1e2ce154f5e4a24f755, PKN: 8, CRYPTO
31	2.963855	34.149.87.45	172.17.0.2	QUIC	188	Protected Payload (KP0), DCID=a5e9168be950abe294ffd1e2ce154f5e4a24f755, PKN: 9, ACK, DONE, NT, NCI
32	2.963856	34.149.87.45	172.17.0.2	QUIC	85	Protected Payload (KP0), DCID=a5e9168be950abe294ffd1e2ce154f5e4a24f755, PKN: 10, ACK

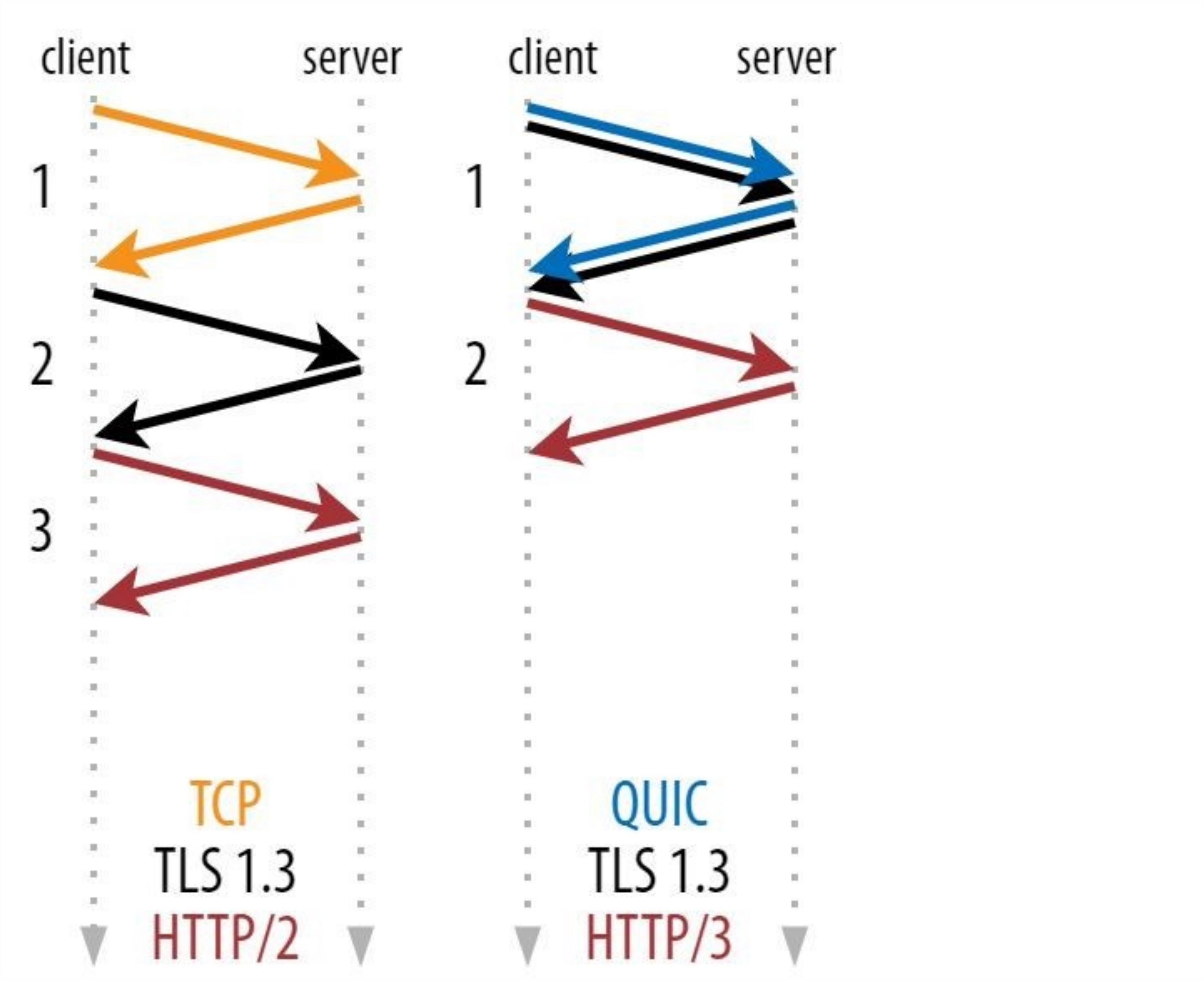
Client sent 1x -> 1242 bytes

Server is limited to 3x -> 1242 * 3 = 3726 bytes

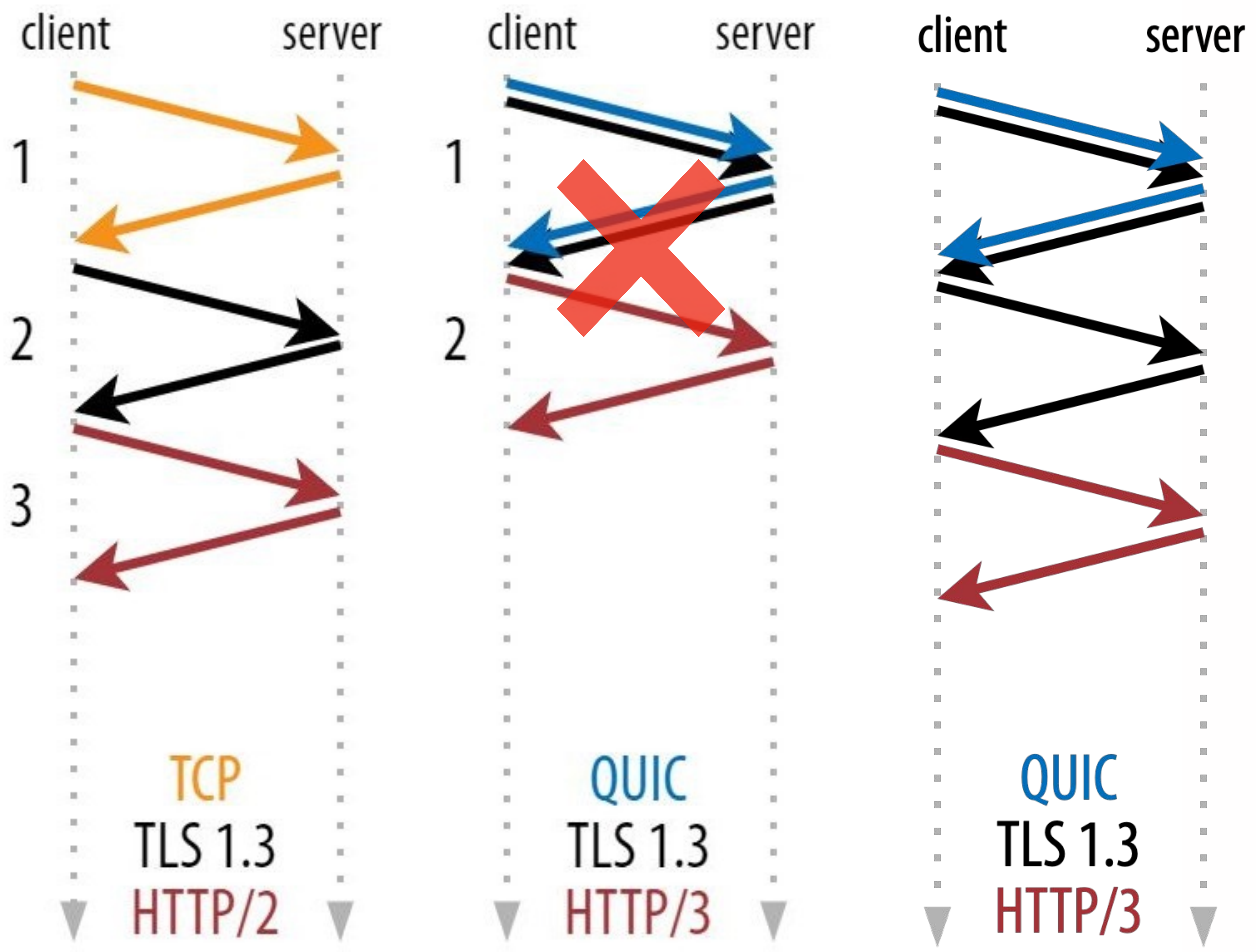
packets1.pcap Packets: 38 · Displayed: 38 (100.0%) · Selected: 3 (7.9%) Profile: Default


packets1.pcap
Apply a display filter ... <⌘/>
No. | Time | Source | Destination | Protocol | Length | Info
16 2.913194 192.168.65.7 172.17.0.2 DNS 167 Standard query response 0xd0b6 AAAA www.sre.com CNAME cdn1.wixdns.net CNAME td-cc
17 2.942025 172.17.0.2 34.149.87.45 QUIC 1242 Initial, DCID=18394474651962efece7db6a90dfd0f6, SCID=a5e9168be950abe294ffd1e2ce15
18 2.951631 34.149.87.45 172.17.0.2 QUIC 1242 Handshake, DCID=a5e9168be950abe294ffd1e2ce154f5e4a24f755, SCID=f8394474651962ef,
19 2.951643 34.149.87.45 172.17.0.2 QUIC 1242 Handshake, DCID=a5e9168be950abe294ffd1e2ce154f5e4a24f755, SCID=f8394474651962ef,
20 2.951644 34.149.87.45 172.17.0.2 QUIC 1242 Handshake, DCID=a5e9168be950abe294ffd1e2ce154f5e4a24f755, SCID=f8394474651962ef,
21 2.952664 172.17.0.2 34.149.87.45 QUIC 1242 Handshake, DCID=f8394474651962ef, SCID=a5e9168be950abe294ffd1e2ce154f5e4a24f755,
22 2.957084 34.149.87.45 172.17.0.2 QUIC 1242 Handshake, DCID=a5e9168be950abe294ffd1e2ce154f5e4a24f755, SCID=f8394474651962ef,
23 2.957091 34.149.87.45 172.17.0.2 HTTP3 214 Protected Payload (KP0), DCID=a5e9168be950abe294ffd1e2ce154f5e4a24f755, PKN: 7,
24 2.958653 172.17.0.2 34.149.87.45 QUIC 173 Protected Payload (KP0), DCID=f8394474651962ef, PKN: 0, ACK
25 2.958838 172.17.0.2 34.149.87.45 HTTP3 92 Protected Payload (KP0), DCID=f8394474651962ef, PKN: 1, STREAM(2), SETTINGS
26 2.958890 172.17.0.2 34.149.87.45 HTTP3 74 Protected Payload (KP0), DCID=f8394474651962ef, PKN: 2, STREAM(6)
27 2.958907 172.17.0.2 34.149.87.45 HTTP3 74 Protected Payload (KP0), DCID=f8394474651962ef, PKN: 3, STREAM(10)
28 2.958913 172.17.0.2 34.149.87.45 HTTP3 143 Protected Payload (KP0), DCID=f8394474651962ef, PKN: 4, STREAM(0), HEADERS: HEAD
29 2.958918 172.17.0.2 34.149.87.45 HTTP3 99 Protected Payload (KP0), DCID=f8394474651962ef, PKN: 5, STREAM(14)
30 2.963838 34.149.87.45 172.17.0.2 QUIC 564 Protected Payload (KP0), DCID=a5e9168be950abe294ffd1e2ce154f5e4a24f755, PKN: 8,
31 2.963855 34.149.87.45 172.17.0.2 QUIC 188 Protected Payload (KP0), DCID=a5e9168be950abe294ffd1e2ce154f5e4a24f755, PKN: 9,
32 2.963856 34.149.87.45 172.17.0.2 QUIC 85 Protected Payload (KP0), DCID=a5e9168be950abe294ffd1e2ce154f5e4a24f755, PKN: 10,
[Packet Number: 5]
Payload [truncated]: 024ae86f91cab0a5fde09238b4f86e623efdacab20a2eeecef7d8545ec756ccbb0cea749412fd57b6495a9e1b9c0fbd2814cc0b2e1ccff0aeee69ae6dfffa36411d949e1197
CRYPTO
Frame Type: CRYPTO (0x0000000000000006)
Offset: 3269
Length: 1141
Crypto Data
TLSv1.3 Record Layer: Handshake Protocol: Multiple Handshake Messages
Handshake Protocol: Certificate (last fragment)
[4 Reassembled Handshake Fragments (3969 bytes): #18(819), #19(1141), #20(1141), #22(868)]
Handshake Protocol: Certificate
Handshake Type: Certificate (11)
Length: 3965
Certificate Request Context Length: 0
3965 > 3726

Handshake time: *Theory*



Handshake time: *Practice*

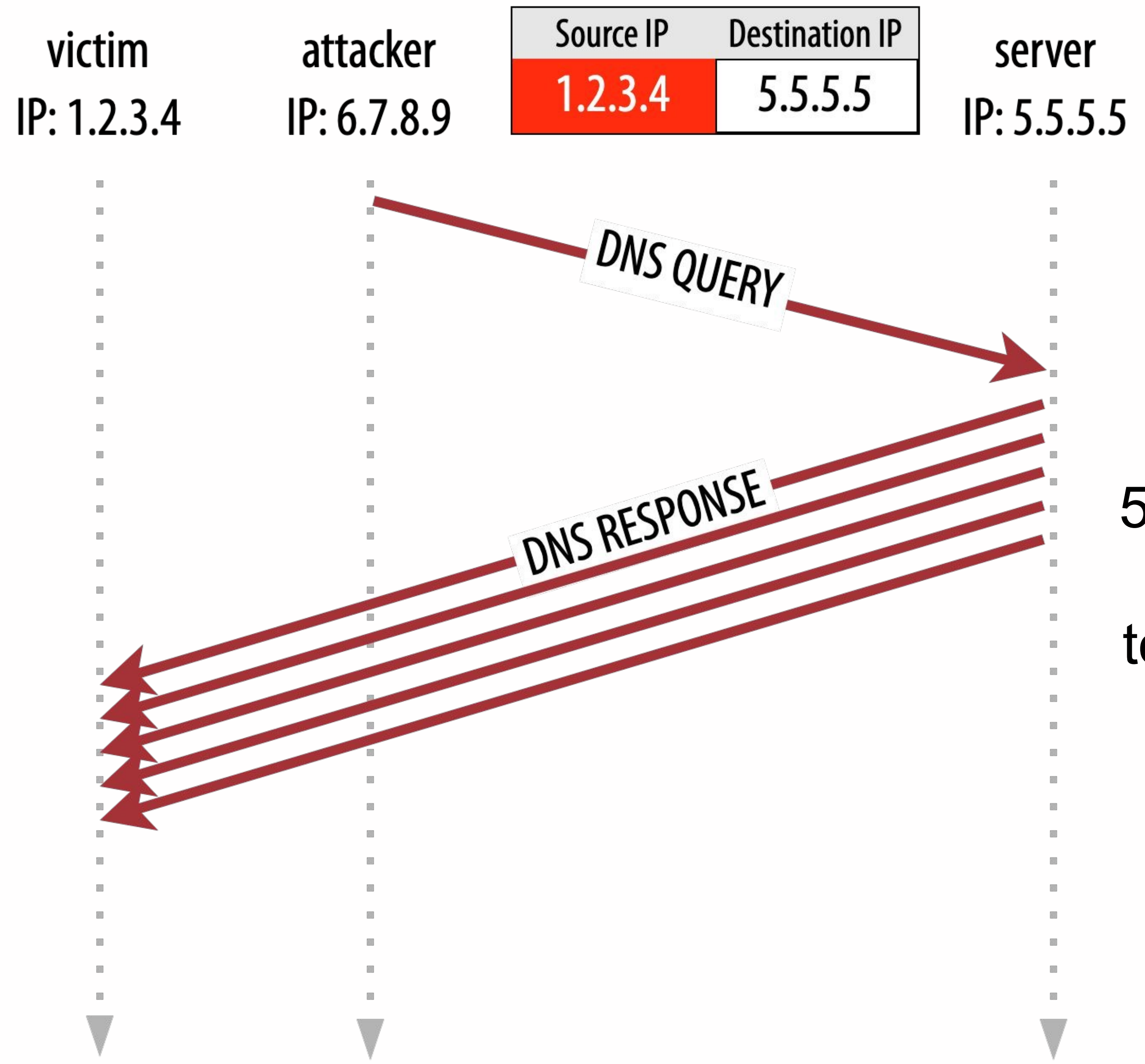


Why limit response to 3x?

- To prevent UDP reflection / amplification attack

Memcached: **51000x** amplification

https://blog.cloudflare.com/memcrashed-major-amplification-attacks-from-port-11211

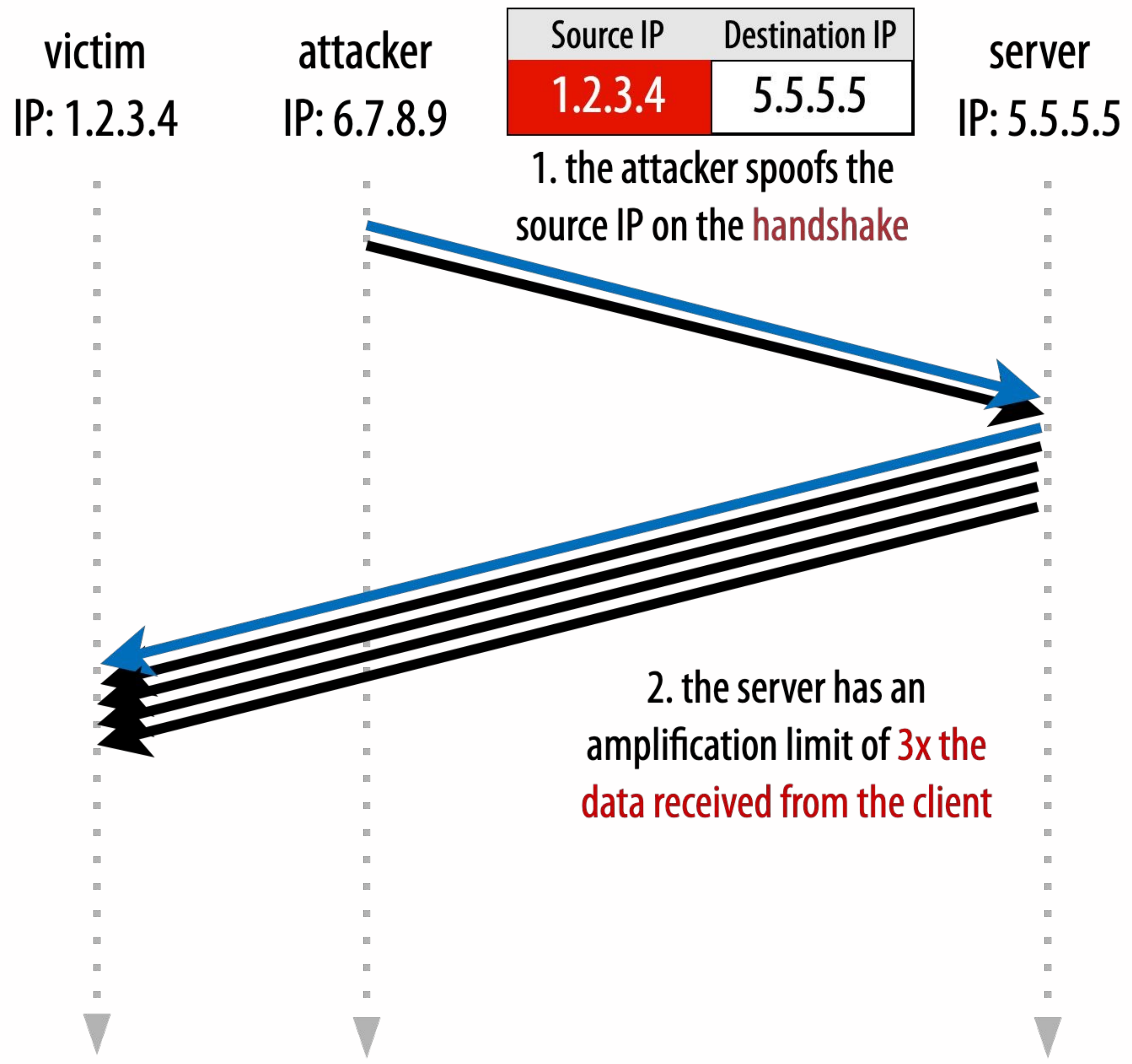


DNS

50x bandwidth amplification towards victim

Impact of 3x limit

Benefit

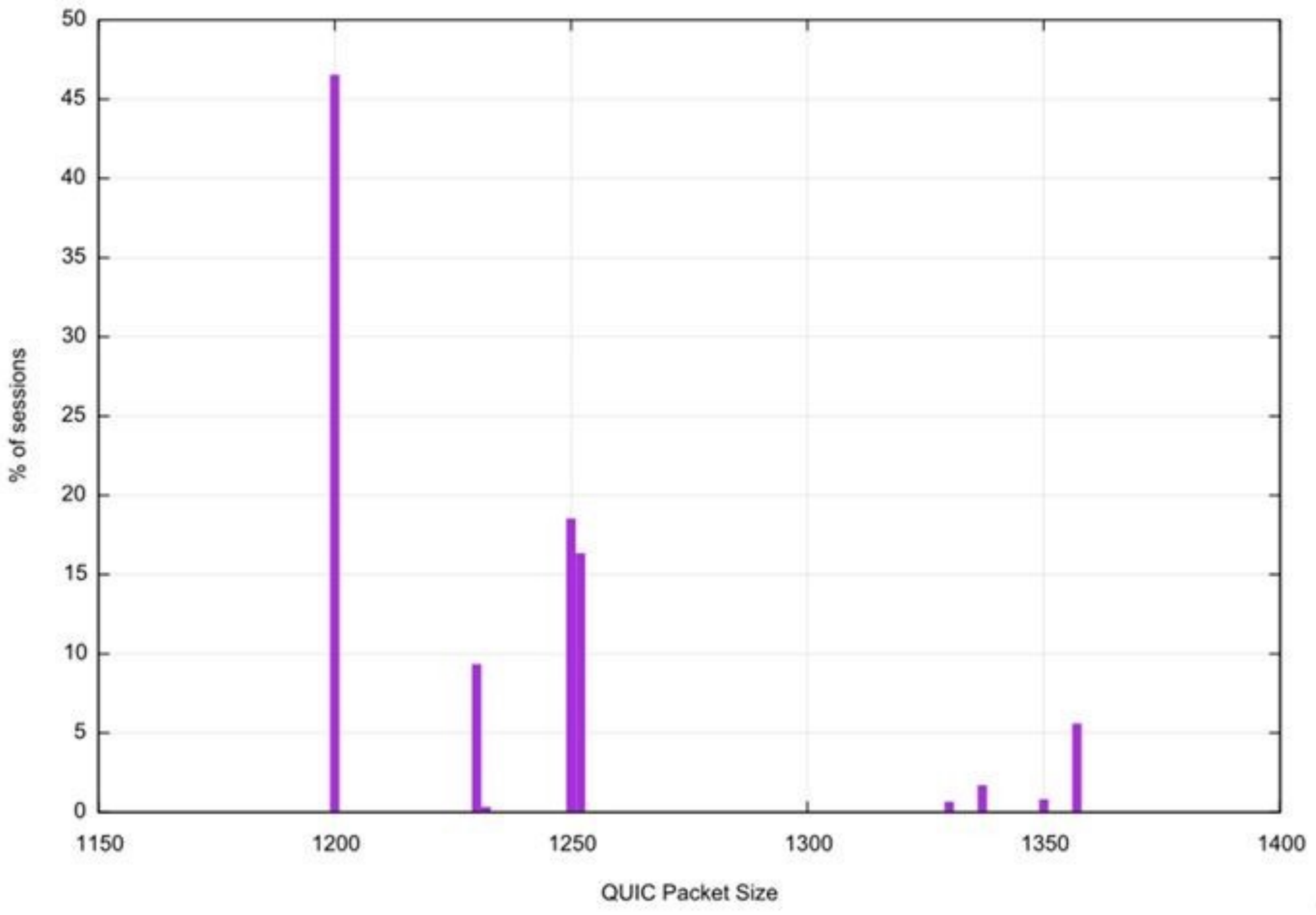


Drawback

- TLS certificate size can sometimes be over the 3x limit
- Multiple round trips are required to complete the handshake
- Large TLS certificates impede QUIC performance

Countermeasures

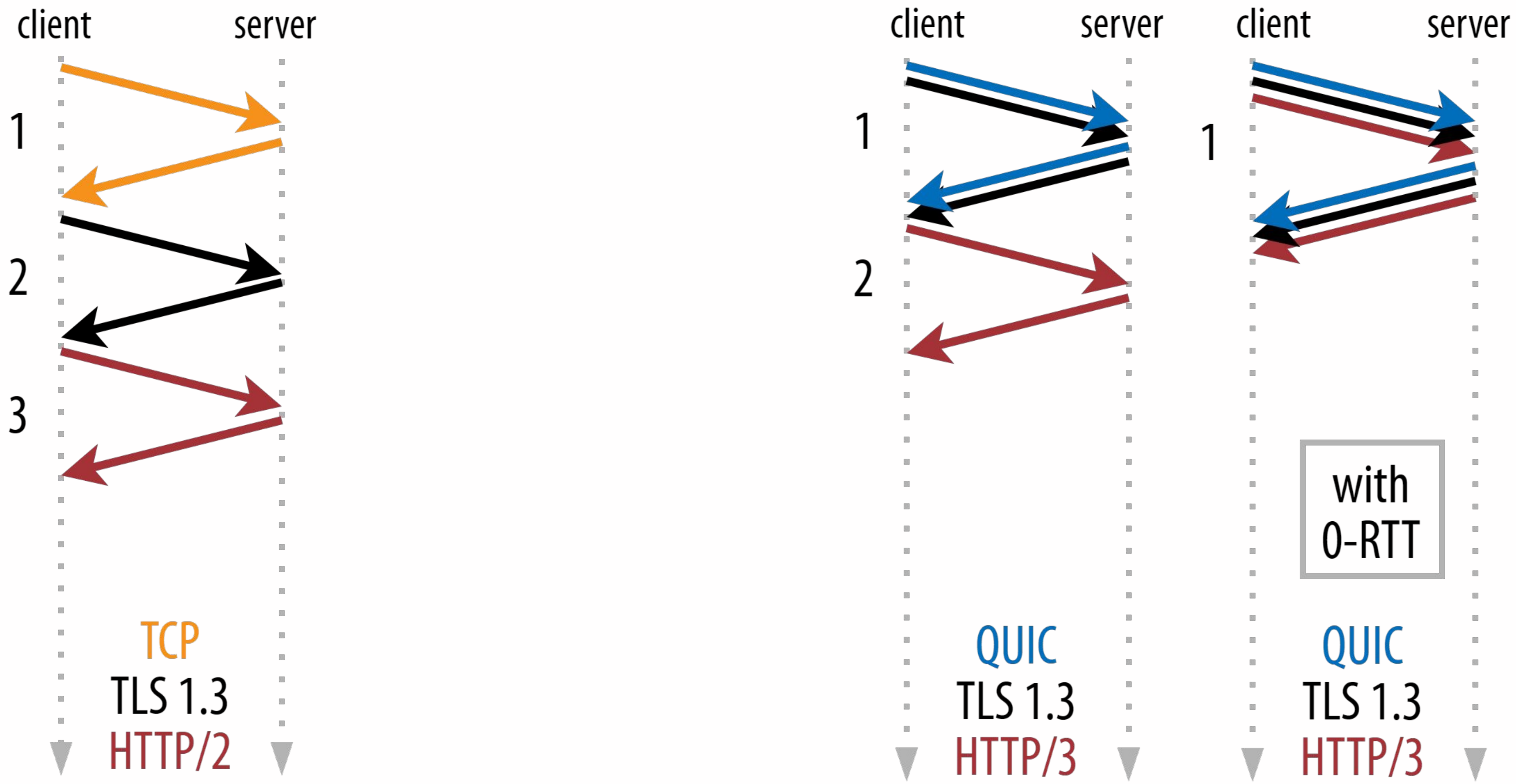
- cURL uses initial packet size of 1240 bytes
- Different clients use different sizes, so the performance can vary.
- Many deployments ignore 3X and go to 4, 5 or 6X to get handshake done in 1 RTT



Case Study 2

Ollscoil na Gaillimhe
University of Galway

- A research experiment shows HTTP/3 is around 50% faster than HTTP/2 in Time To First Byte(TTFB), but it should be 33% (or 66% if using 0-RTT)

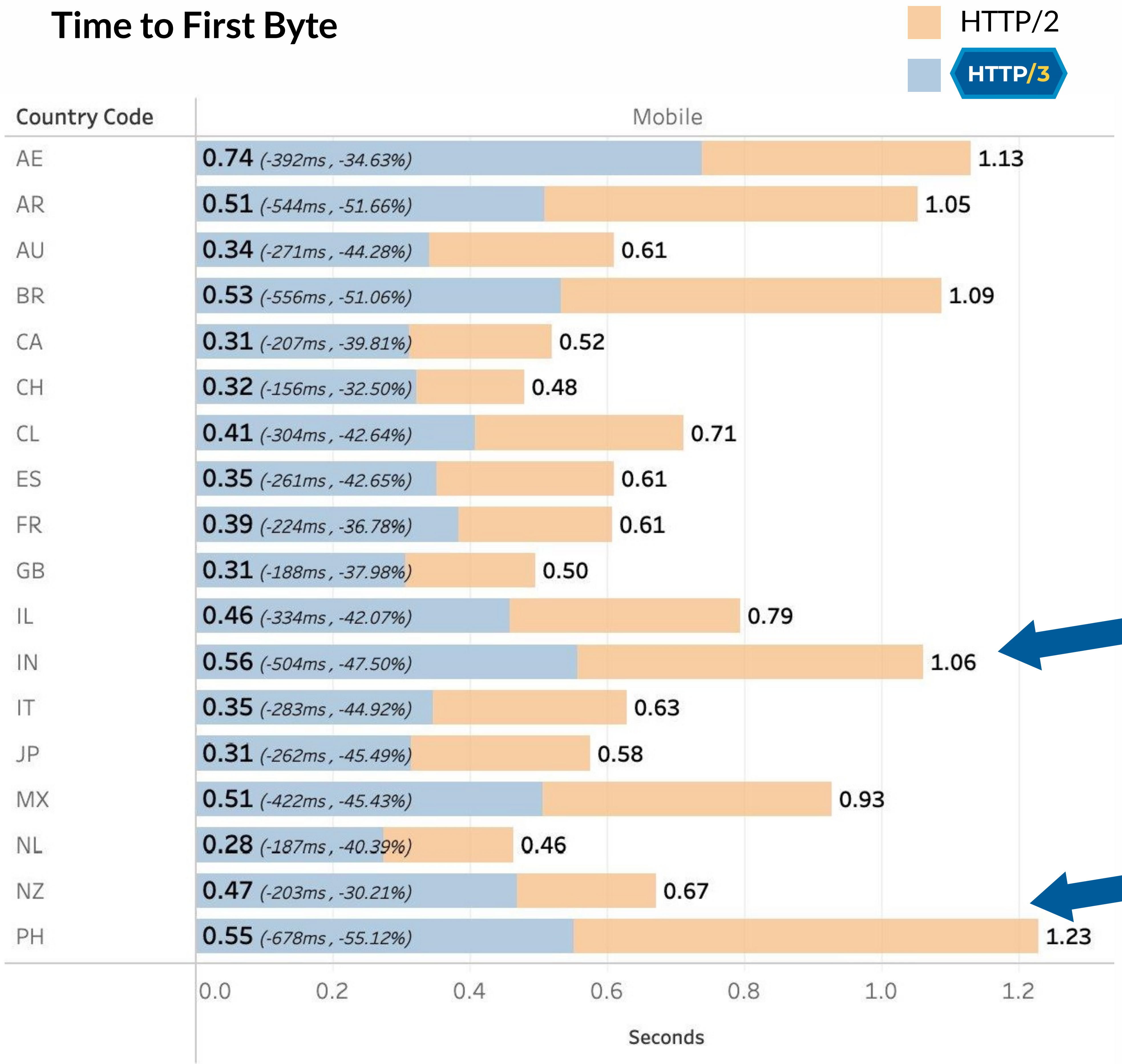


~33% faster

~66% faster

Yet we see ~**50%**?

Time to First Byte

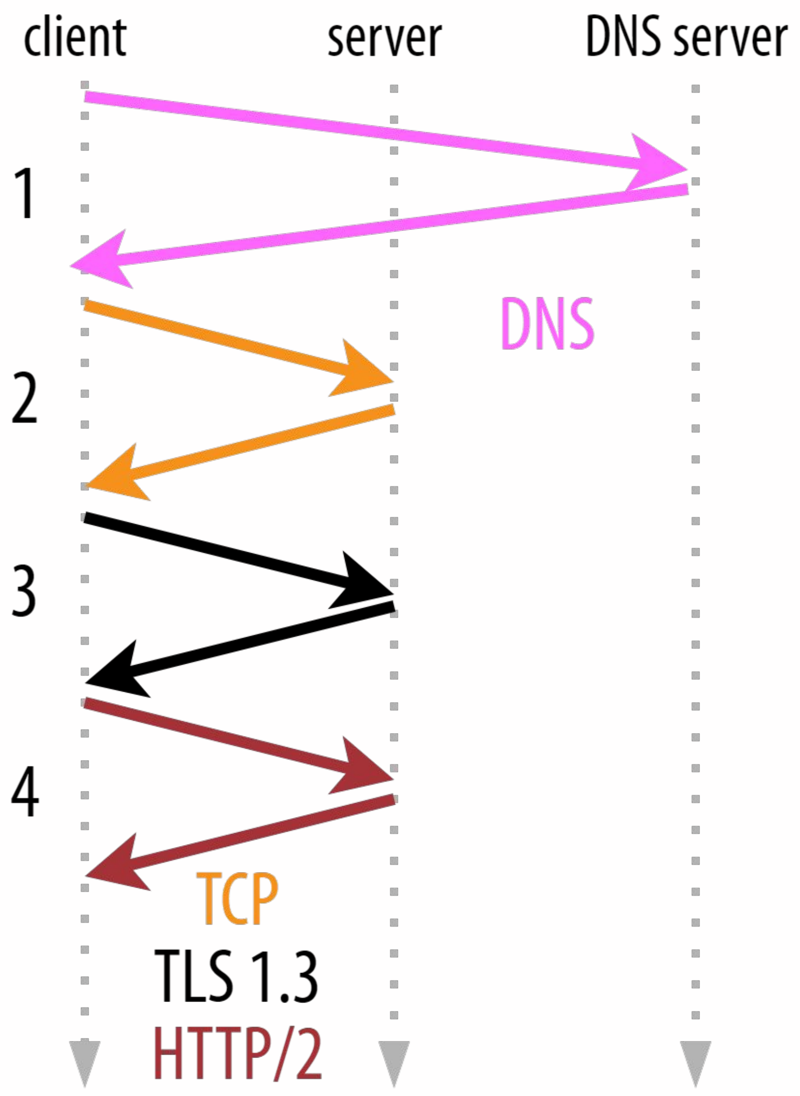


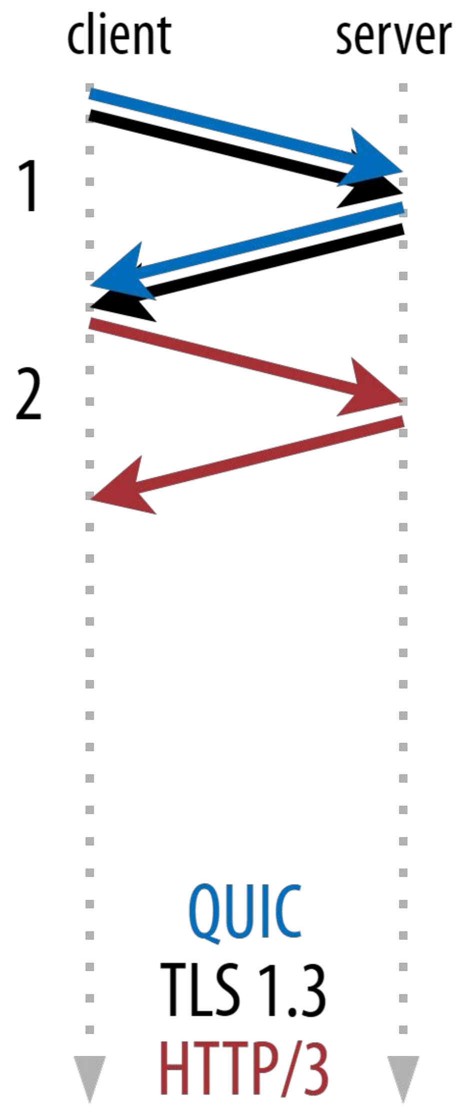
India:
47.50% faster!
1000ms vs 560ms!

Philippines:
55% faster!
1230ms vs 550ms!

Traffic Analysis

- To find the root cause we capture traffic and analyze it
- Through traffic analysis we discover that DNS time in included in HTTP2
- But why is DNS time not present in HTTP3?





~50% faster

= unfair comparison

Browser only do HTTP/3 after *discovery*

Ollscoil na Gaillimhe
University of Galway

- When talking to a new domain, the browser does not start with HTTP3 because its not sure if the server supports it
- For a new hostname browser performs the following:

1. Browser requests the page over HTTP/1 or 2
2. Server sends back "**alternative services**" header

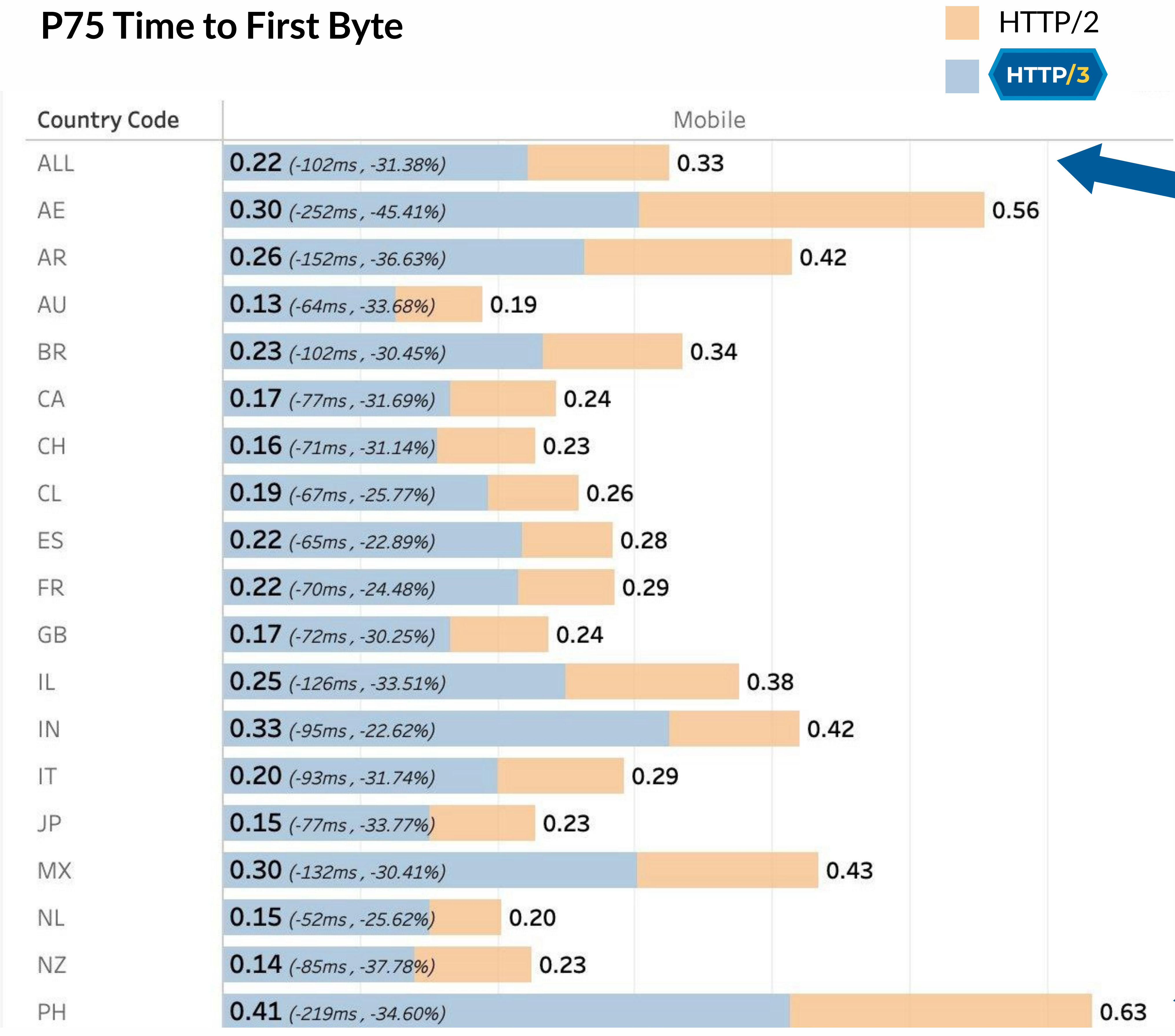
   ▼ **Response Headers**

   alt-svc: `h3=":443"; ma=86400`

   cache-control: `private, no-cache, no-store, must-revalidate`

3. Browser stores alt-svc info in alt-svc "cache"
4. Browser *also* tries HTTP/3 from now on, *in parallel* with HTTP/1 or 2 ("free" fallback)

P75 Time to First Byte



Mean:
31% faster! 330ms vs 220ms!

Philippines:
34% faster! 630ms vs 410ms!

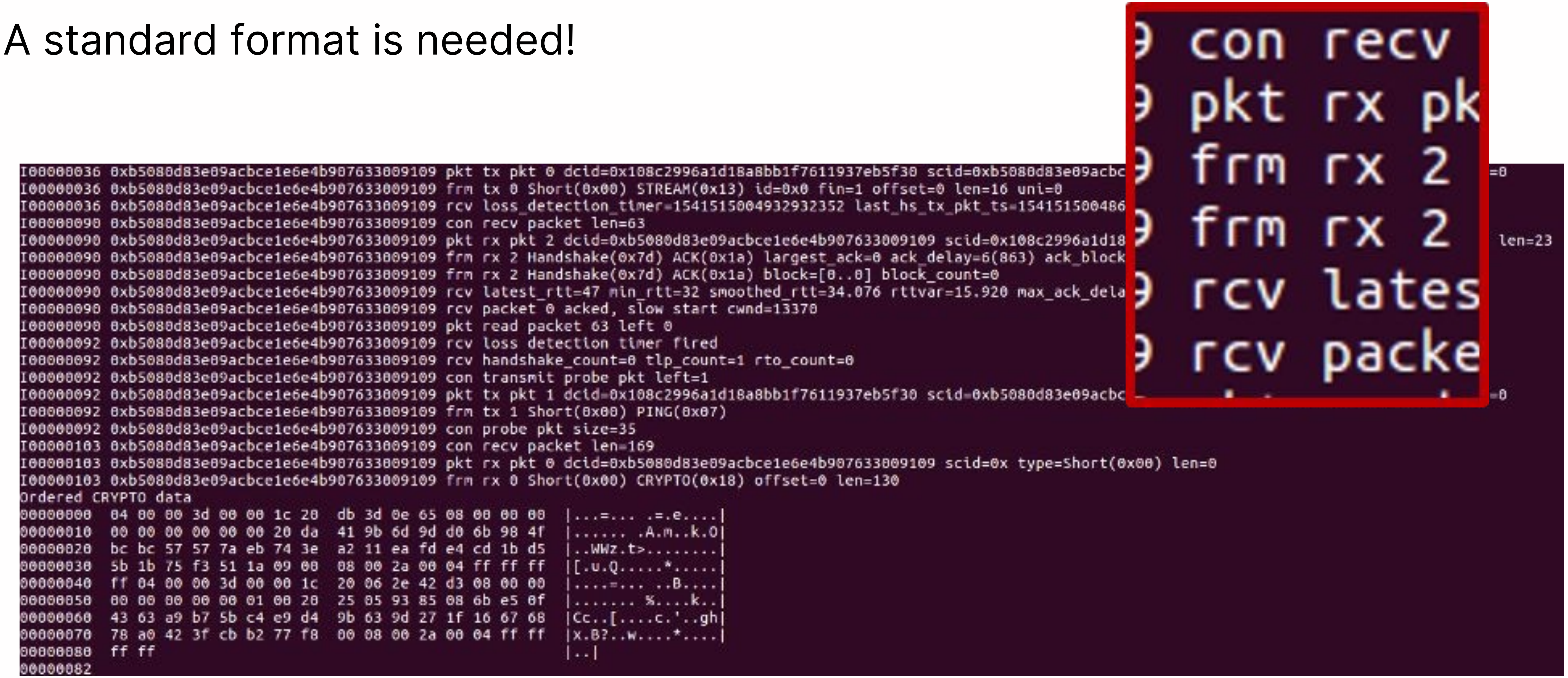
Why do we need other tools

Ollscoil na Gaillimhe
University of Galway

- Why can't we just use Wireshark?
  - QUIC is heavily encrypted and very little information is visible in Wireshark without decryption keys
    - Don't always have TLS decryption keys.
  - A lot of core performance information is not sent on the wire, it is only available at the end points
  - Some features not fully supported
    - HTTP/3 QPACK header decoding was added just a few months ago.
  - Wireshark JSON/XML output isn't easy to use by default.
  - Wire image does not contain all info
    - Internal state information is missing, e.g. no congestion window

Log Information

- There is a lot of useful information in the application log
- However, parsing random application logs is not fun!
- A standard format is needed!



[qlog]

OLLSCOIL NA GAILLIMHE
UNIVERSITY OF GALWAY

- Structured endpoint logs
- Log metadata and state in the endpoints (client and server) in the QUIC implementations.
- qlog is a schema for JSON describing QUIC events
- Each qlog event is defined by a timestamp, a category (e.g., “transport”), an event type (e.g., “packet_sent”) and some type specific data (e.g., the size of the sent packet and its header fields).
- qlog is flexible
  - New event categories, types and metadata can trivially be added, modified and extended

[qlog]

Ollscoil na Gaillimhe
University of Galway

- qlog examples


```json
{
"time": 15000,
"name": "transport:packet_received",
"data": {
    "header": {
        "packet_type": "1rtt",
        "packet_number": 25
    },
    "frames": [
    {
        "frame_type": "ack",
        "acked_ranges": [
            [10,15],
            [17,20]
        ]
     }]
}}
```

```json
{
"time": 15001,
"name": "recovery:metrics_updated",
"data": {
  "min_rtt": 25,
  "smoothed_rtt": 30,
  "latest_rtt": 25,

  "congestion_window": 60,
  "bytes_in_flight": 77000,
}
```

qlog adoption

Ollscoil na Gaillimhe
University of Galway

>70% of QUIC implementations have (partial) support:

- aioquic
- quic-go
- quiche
- mvfst
- picoquic
- haskell
- ngtcp2
- ...

Others do something similar:

- msquic
- google quiche

Facebook has deployed it in production

Store over 30 billion qlog events daily

IETF standardization in-progress
https://datatracker.ietf.org/doc/html/draft-ietf-quic-qlog-main-schema-11

<qvis>

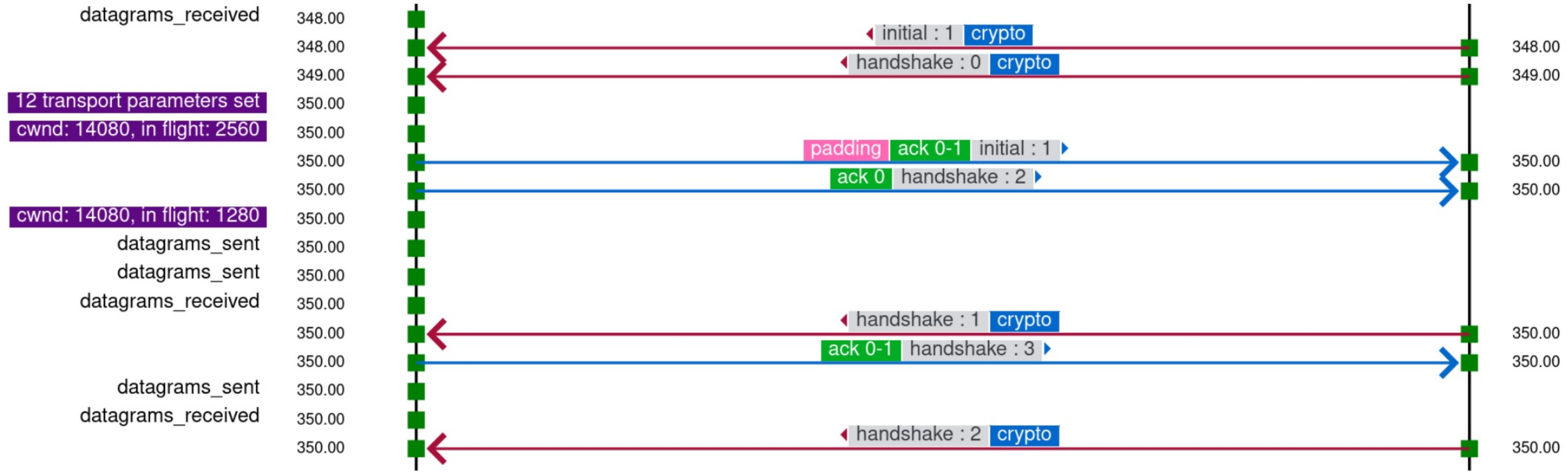
- qviz is open-source toolsuite that can directly ingest and visualize qlog files
- It provide a number of tools:
  - Sequence diagram
  - High-level statistics overview
  - Congestion control
  - Multiplexing
  - Packetization

Ollscoil na Gaillimhe
University of Galway

Visualization Case Studies

Sequence Diagram

- The sequence tool generates a sequence diagram.
- The green squares on both sides represent events.
- All the green boxes, event names and packet information can be clicked which brings up the corresponding qlog file in plaintext, allowing for further, more detailed packet inspection

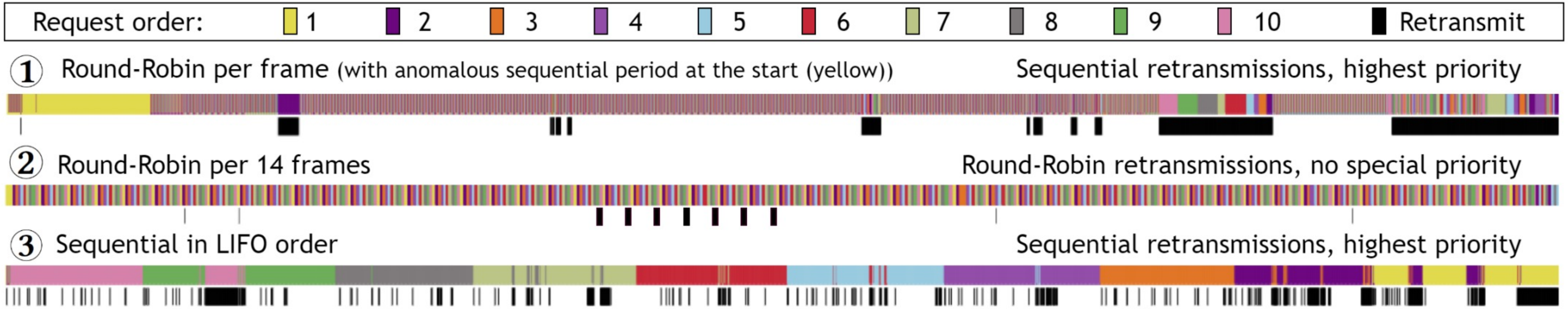


Stream Multiplexing and Prioritization

- Modern protocol stacks often multiplex data from several parallel “streams” onto one connection (e.g., HTML, CSS and image files when loading a web page).
- This multiplexing can happen in various ways
- (e.g., files are sent sequentially as a whole or are scheduled via Round-Robin (RR) after being subdivided in chunks) and is typically steered using a prioritization system
- qvis multiplexing diagram can be used to verify and debug an implementation.
- It shows the response payload carrying frames, displayed on a horizontal line with different colors to distinguish the stream each frame belongs to.

Stream Multiplexing and Prioritization

- This example shows multiplexing behavior across three different QUIC stacks when downloading 10 MB files in parallel
  - Each small colored rectangle is one payload frame belonging to a file.
  - Black areas indicate which frames above them contain retransmitted data.
  - Data arrives from left to right.



Debugging QUIC and HTTP/3 with qlog and qvis, RobinMarx et. al, AppliedNetworking Research Workshop (ANRW '20)

Stream Multiplexing and Prioritization

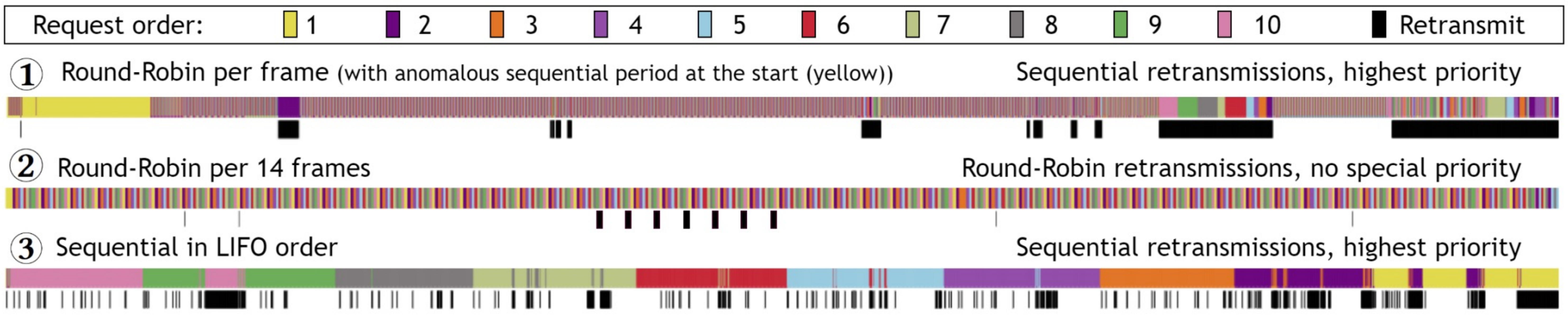
Ollscoil na Gaillimhe
University of Galway

Observations

- RR schemes show frequent color changes(1 , 2)
- Long contiguous swaths(3) mean sequential transfers
- In (3) later streams are interrupted with retransmissions of earlier ones
- (2) interleaves retransmissions with new data
- (1) changes its multiplexing behavior from RR to sequential for lost data

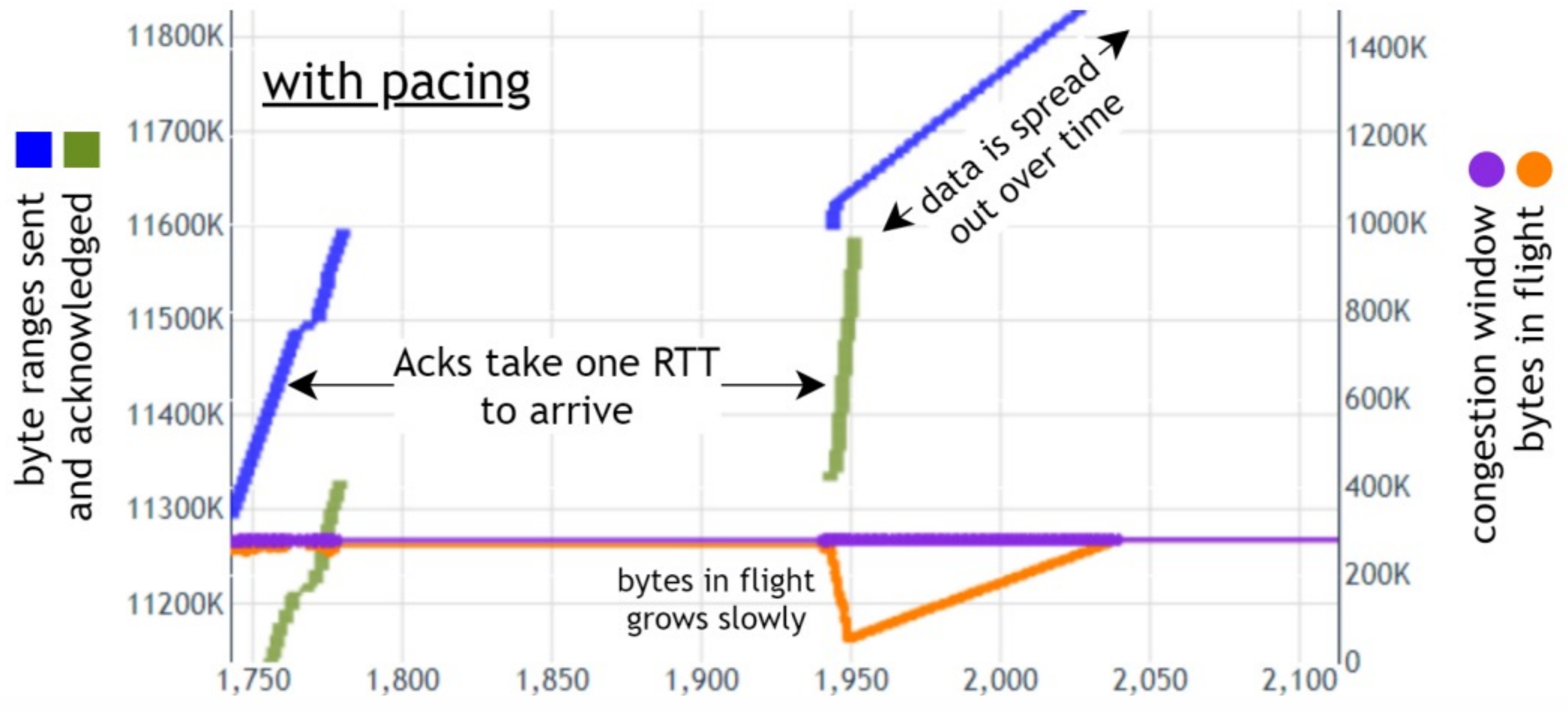
Abnormalities

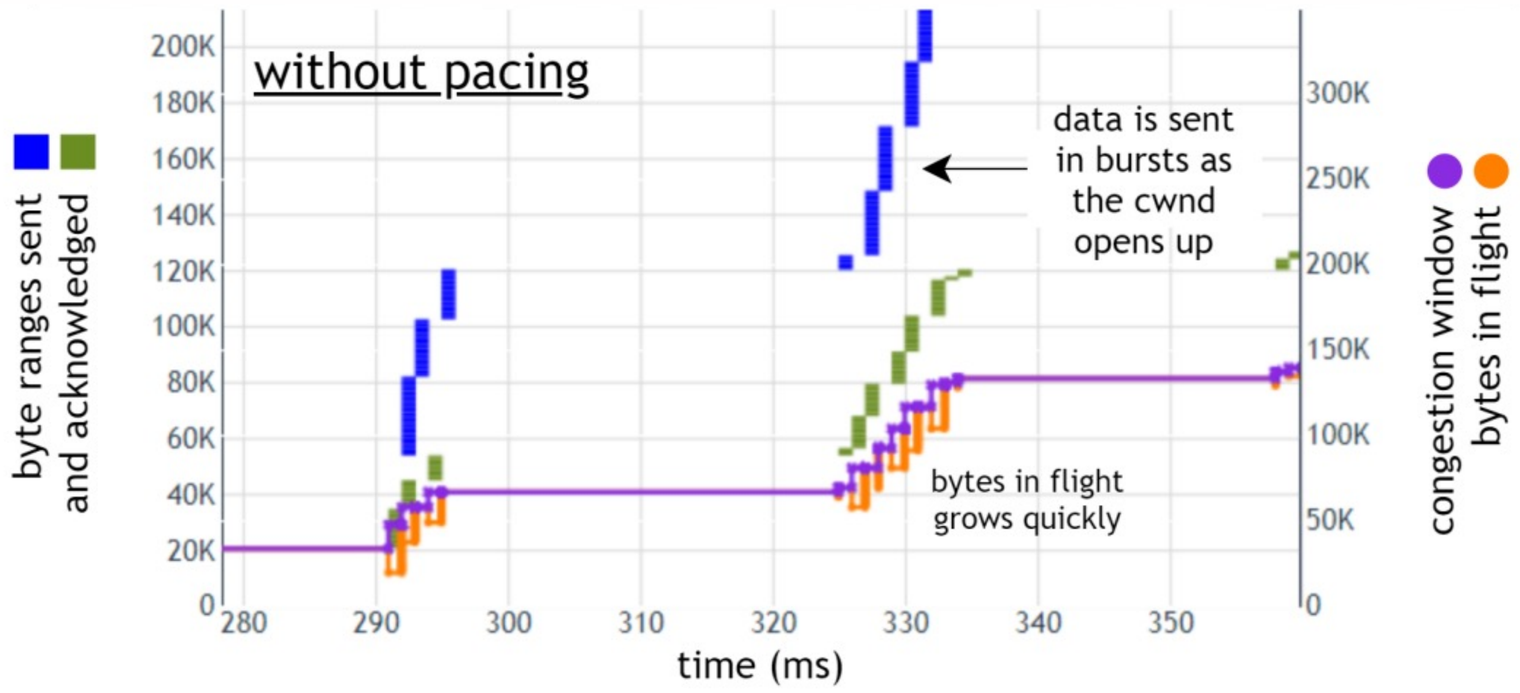
- (1) normally uses RR but has a long sequential period at the start
- (3) unintentionally sent data in Last-In First-Out order, the worst-case for web performance



Congestion Control (CC)

- CC is topic of active research which is more open to experimentation in QUIC.
- Debugging CCs is a major reason for create custom visualizations.
- qvis suite includes a comprehensive congestion control graph.
- It plots data sent, acknowledgements received, flow control limits, congestion window, bytes in flight,and employed RTT measurements on a timeline.

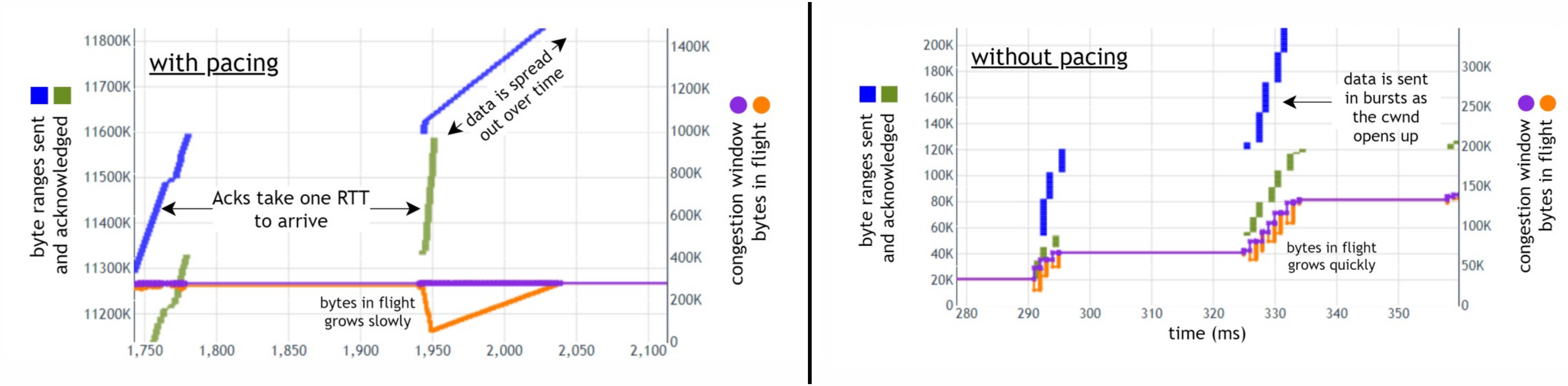




Congestion Control (CC)

Ollscoil na Gaillimhe
University of Galway

- Observations
  - With pacing, the bytes in flight grow slowly over time as data is spread out, while without pacing, it jumps up quickly.
  - Pacing is the practice of spreading out packets across an RTT instead of sending them in short bursts, and is thought to reduce packet loss.



Debugging QUIC and HTTP/3 with qlog and qvis - ANRW '20: Proceedings of the 2020 Applied Networking Research Workshop

Congestion Control (CC)

- Practical uses
  - Facebook diagnosed their BBR code not entering the probeRTT state at the right time.
  - They also identified large-scale pacing issues between their transatlantic data centers due to errors in RTT measurement.
  - Cloudflare used qvis to debug their Cubic CC with ‘hystart’ implementation.
  - Bugs were found in QUIC’s retransmission logic during its complex handshake.

Demo

- https://qvis.quictools.info/

Acknowledgement

- The content is adapted from Dr. Robin Marx's presentation at SREcon23
- https://www.usenix.org/conference/srecon23emea/presentation/marx

OLLSCOIL NA GAILLIMHE
UNIVERSITY OF GALWAY

Thank you for your attention!

University
ofGalway.ie