Modern Web Protocols Part 2

cs249i

Where we left off…

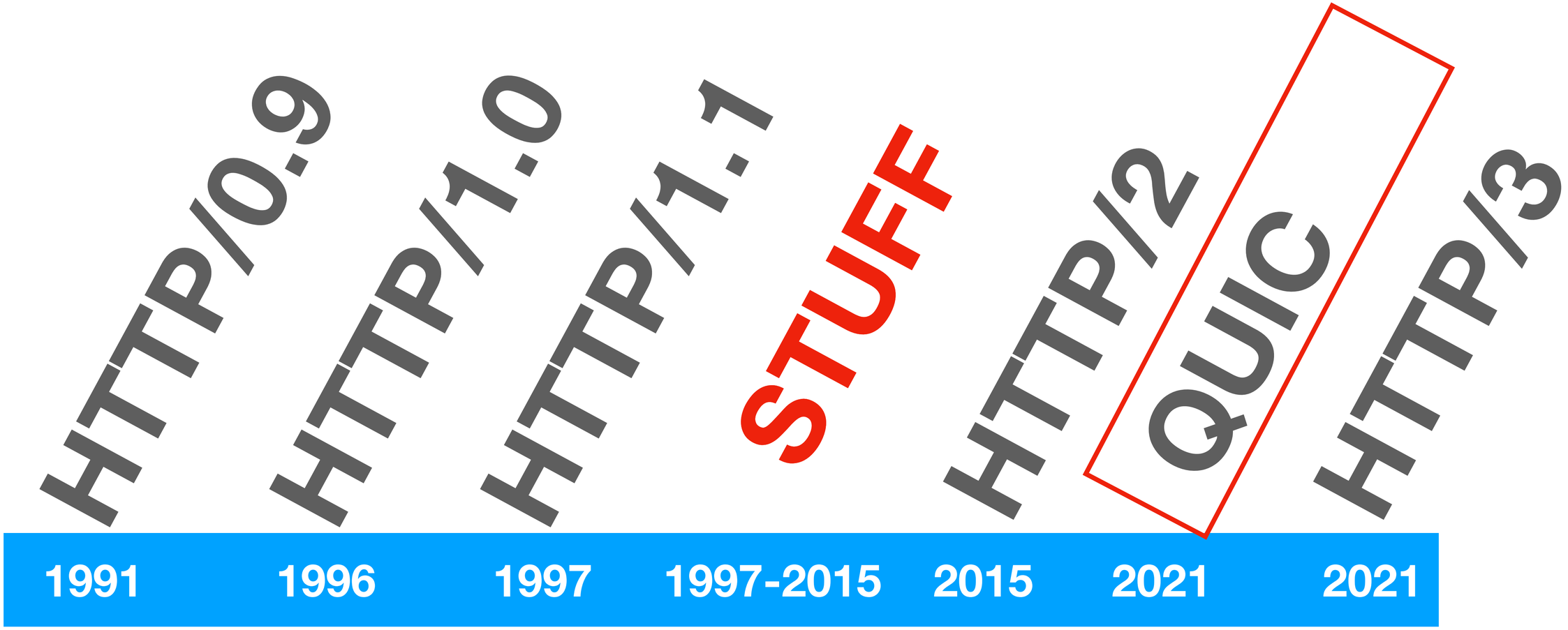
A History of Web Protocols

- HTTP/1.x has lots of problems
  - Standard request / response paradigm + HoL blocking made it difficult to scale to support increasingly complex modern websites
  - Solutions to this problem (e.g., HTTP pipelining) ran into serious deployment challenges which never let it take off
- HTTP/2 was created to solve lots of these problems
  - Implemented a new abstraction (e.g., byte streams, frames, messages) and fixed lots of challenges
  - But still suffers from HoL blocking, just at the TCP layer instead of the HTTP layer

Lingering Questions

- Q1: How many HPACK tables are there at once?
  - There is a pair of static / dynamic tables generated *per HTTP/2 connection*.
  - Size restraints make this feasible (but are tunable by client / server in settings frames)
  - If two mutually distrustful clients are using the same HTTP/2 connection, they can probe dynamic table state (and potentially leak client information)
- Q2: Can Server Push be used as a notification system?
  - No, the browser doesn’t expose server push in JavaScript, see Push API instead

A History of Web Protocols



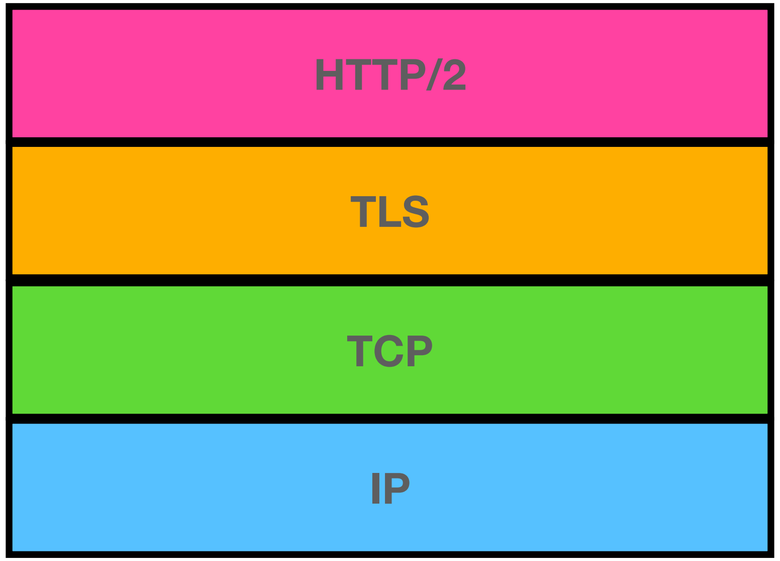
QUIC

A New Way Forward

- A core problem with HTTP up to this point is a fundamental limitation of *reliable transport over TCP*
  - We want to have reliability guarantees, but the way this is implemented in the layering model (e.g., in TCP) makes it such that applications don't have flexibility to define what reliability means!
- We could try to change TCP?
  - But that requires updating every router in the world. Way too hard.
- QUIC idea: What if we re-envisioned what we needed from lower network layers?

QUIC

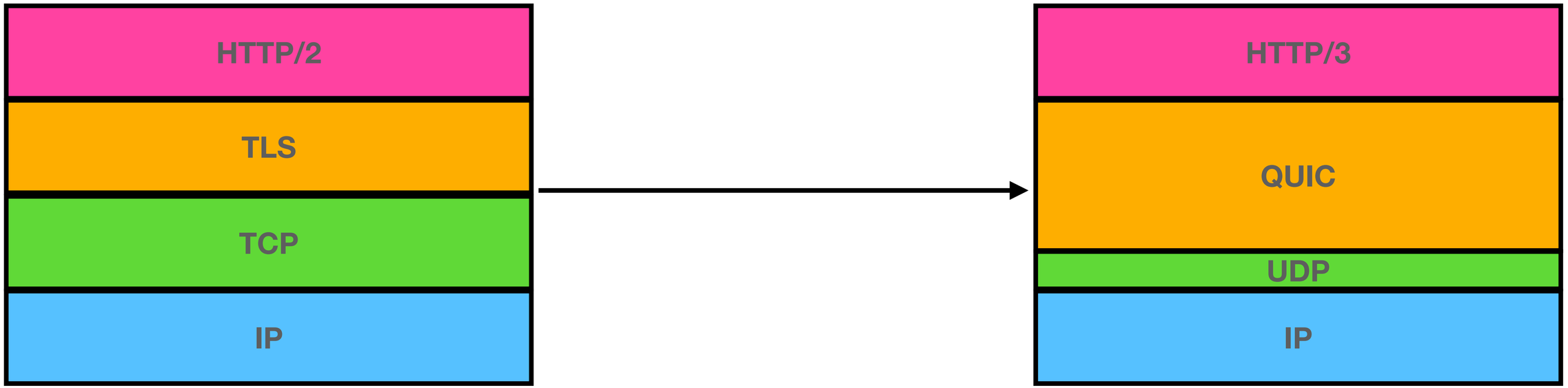
A New Transport Layer



The current world

QUIC

A New Transport Layer

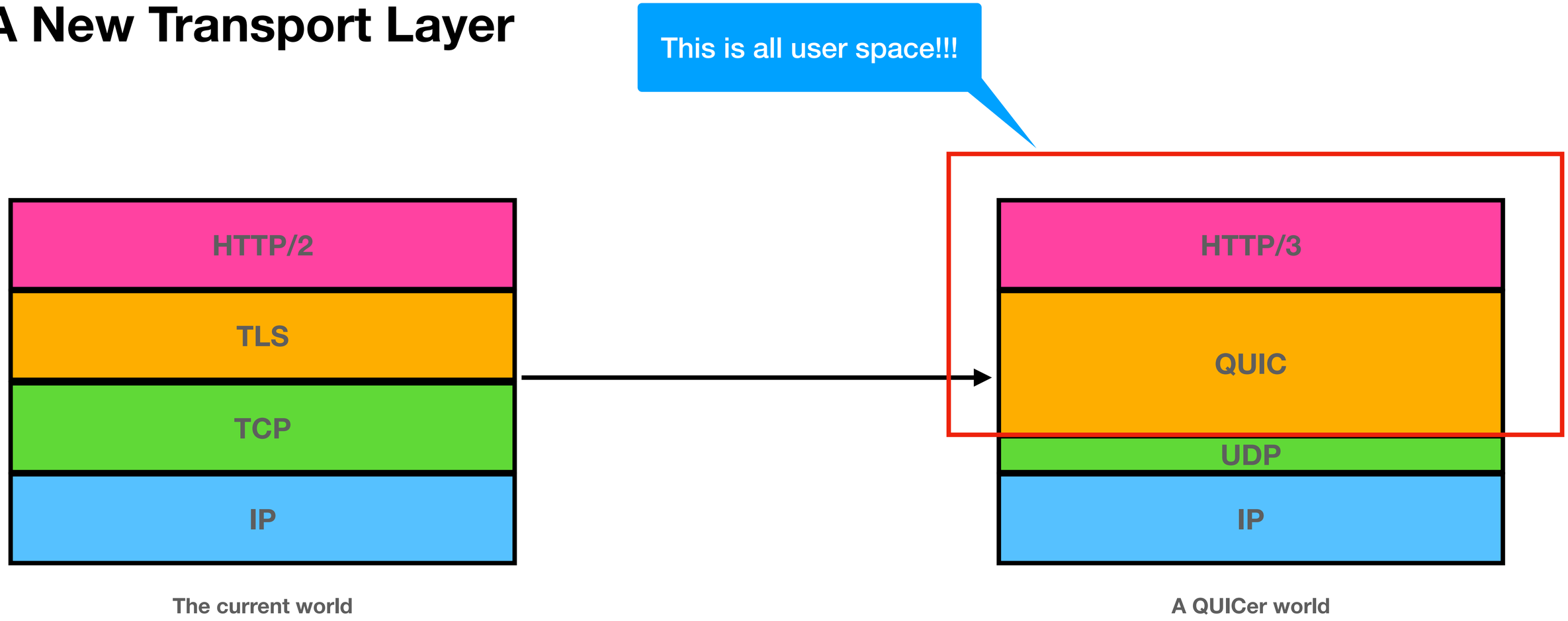


The current world

A QUICer world

QUIC

A New Transport Layer



QUIC

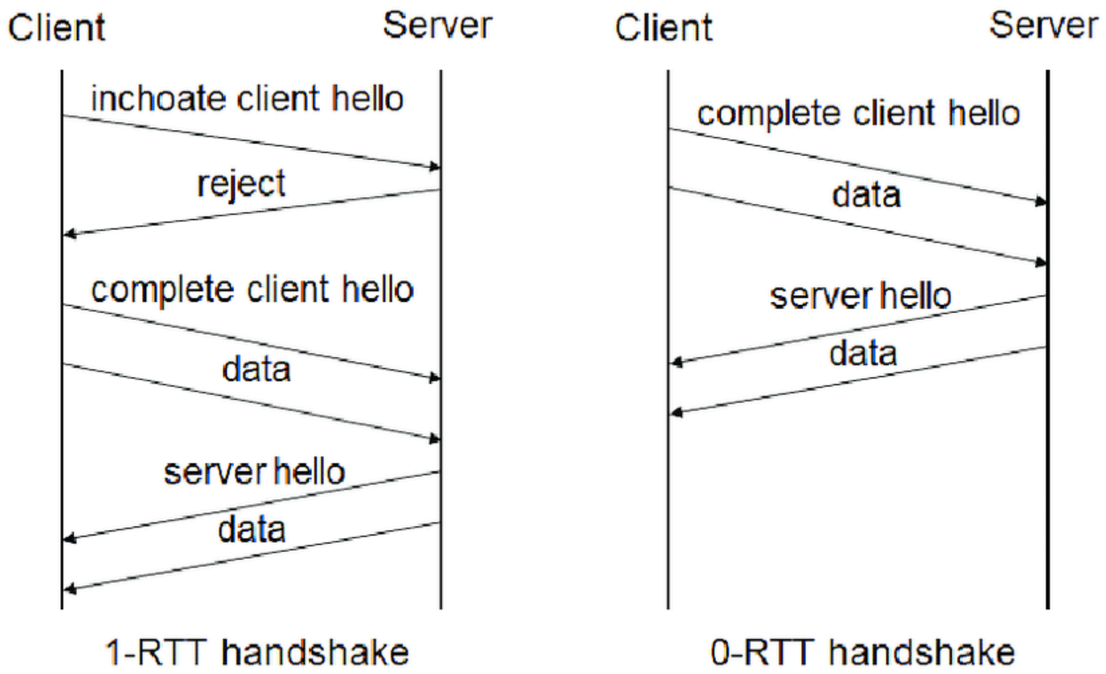
Design Goals

- A new, reliable transport layer
- Easily deployable and evolvable
  - Make this something that exists in userspace and something that doesn’t require us to update every router ever
- Security by default
  - Build in encryption, integrity checks, and authentication into the transport layer itself
- Reduce unnecessary delays imposed by strict layering
  - Handshake delays (e.g., TLS handshake), HoL blocking (HTTP, TCP)

QUIC

Establishing a Connection

- The first time a client wants to communicate with a server, it send an *inchoate client hello* in cleartext, which will initiate a REJ (reject) from the server
    - The server will send back a number of details, including a certificate chain (for server authentication), long term keying materials, and other server metadata
- The client will then use the server information provided to send a *complete client hello*, and immediately start sending encrypted data with non forward-secure keys
- Server sends back *server hello*, with ephemeral forward-secure public keying material
- Client *caches* server details (based on origin), so for any future connection, the client can simply use the server block data to send encrypted messages moving forward. This is known as a **0-RTT protocol.**



QUIC

Two Types of Headers

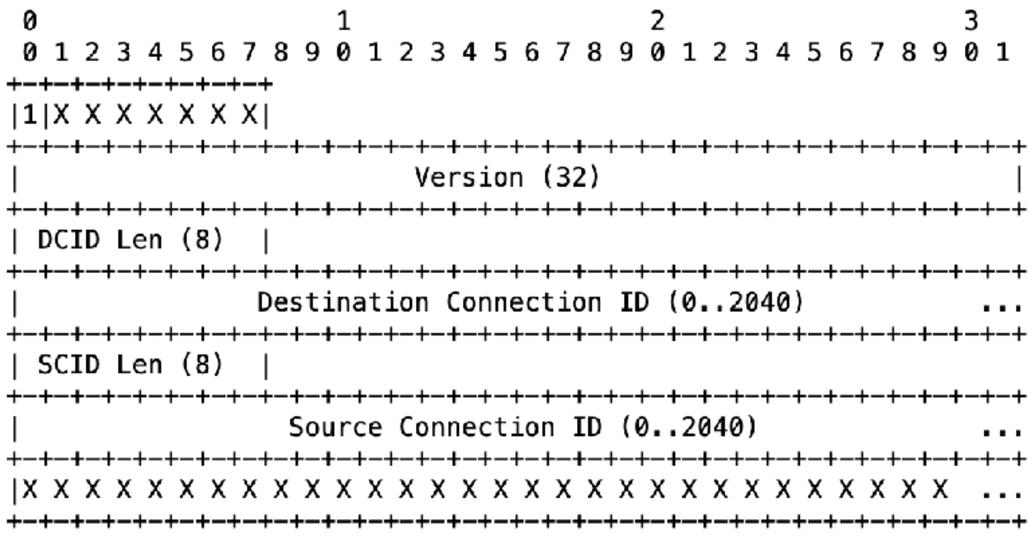
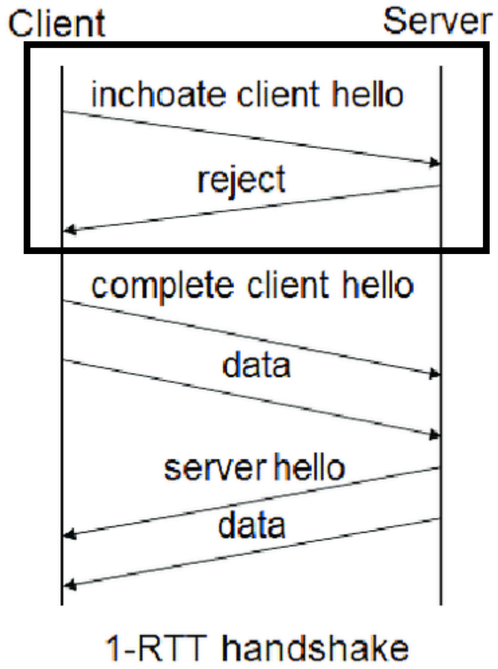
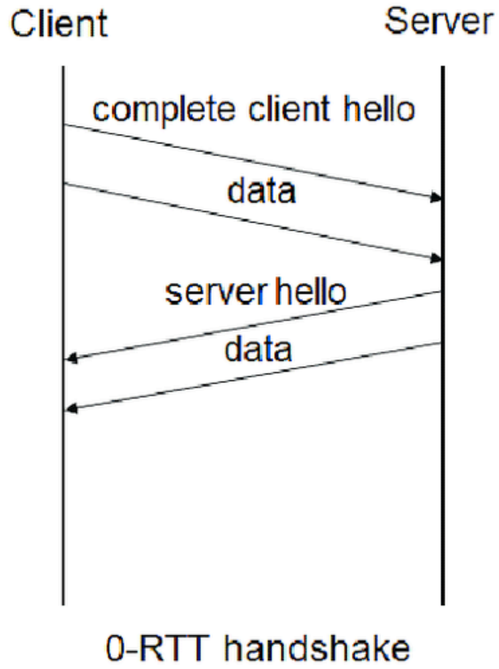


Figure 1: QUIC Long Header



1-RTT handshake



0-RTT handshake

https://datatracker.ietf.org/doc/html/draft-ietf-quic-transport-19#section-17.2

QUIC

Two Types of Headers

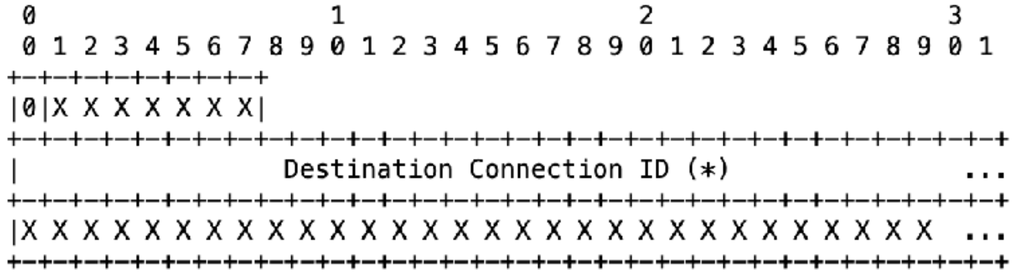
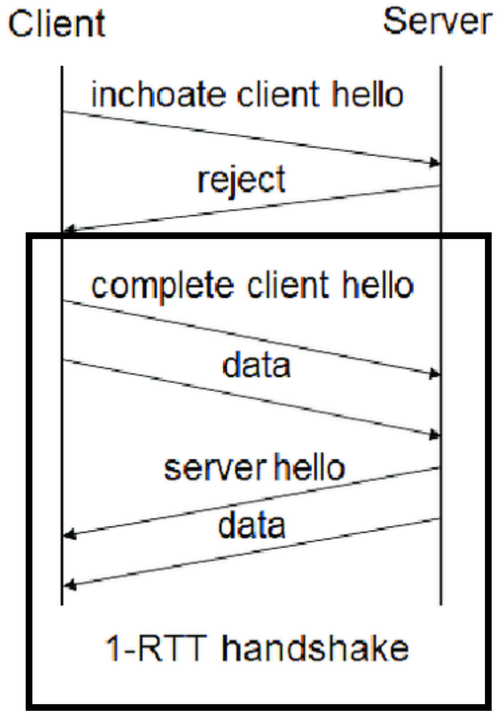
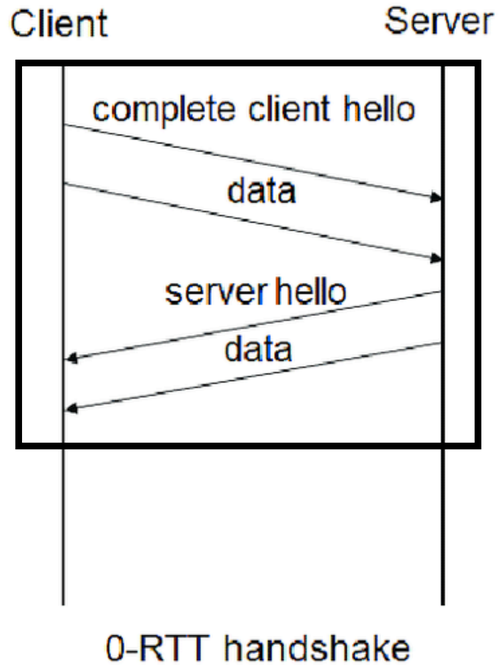


Figure 2: QUIC Short Header



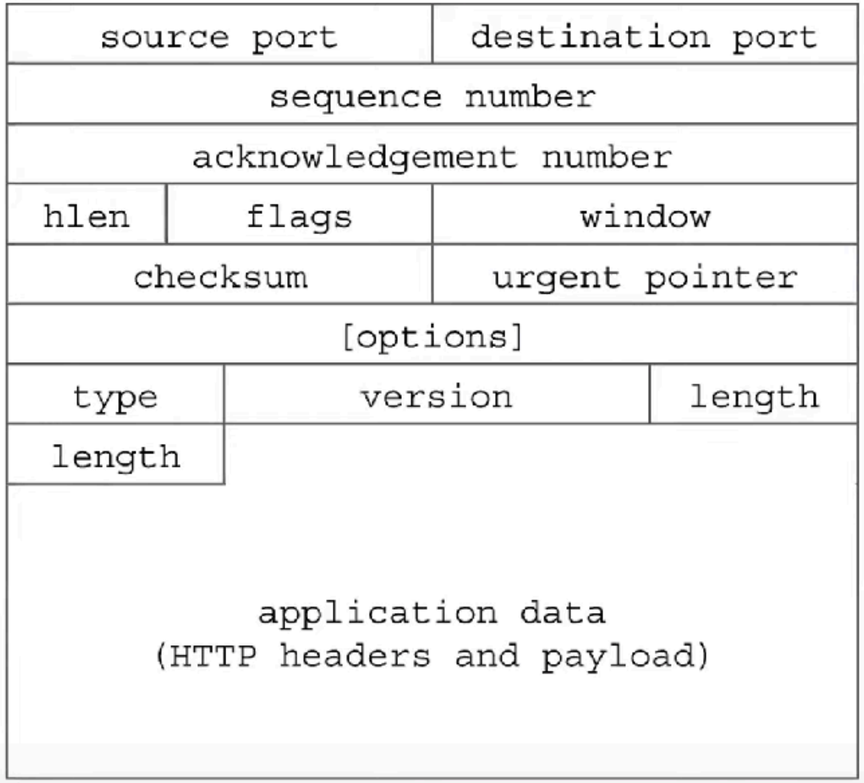


https://datatracker.ietf.org/doc/html/draft-ietf-quic-transport-19#section-17.2

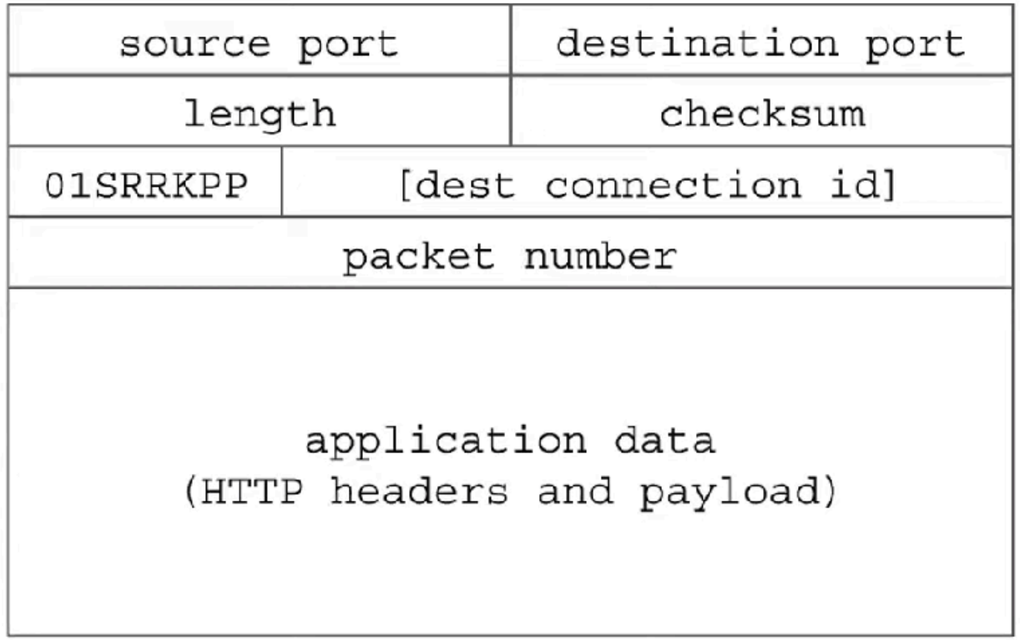
QUIC

Encrypt as much as possible

HTTP w/ TLS + TCP



HTTP w/ QUIC

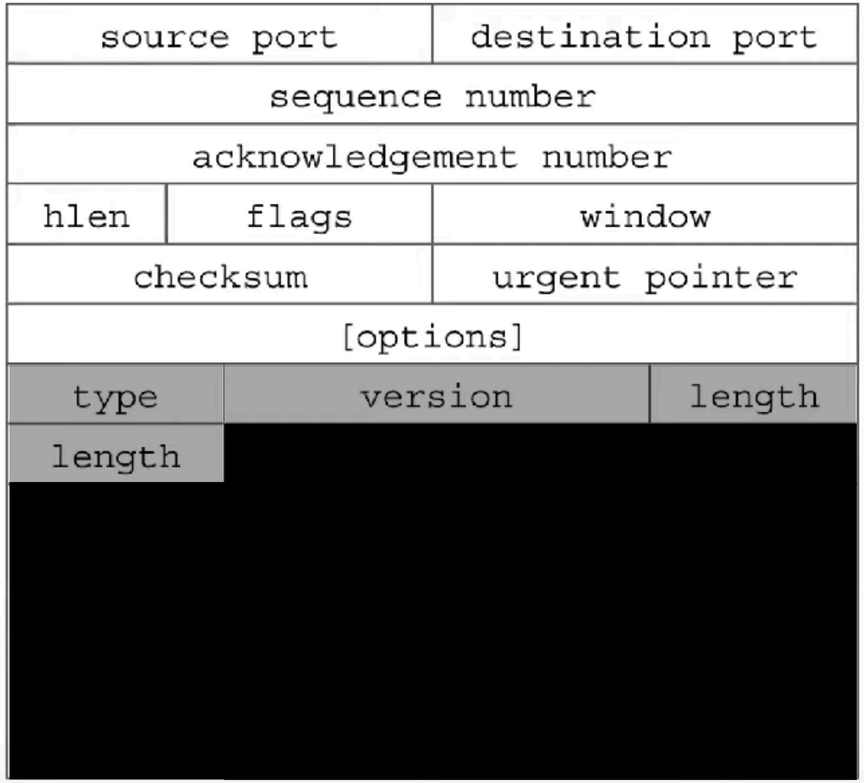


Slide stolen from: https://www.youtube.com/watch?v=31J8PoLW9iM&t=9104s

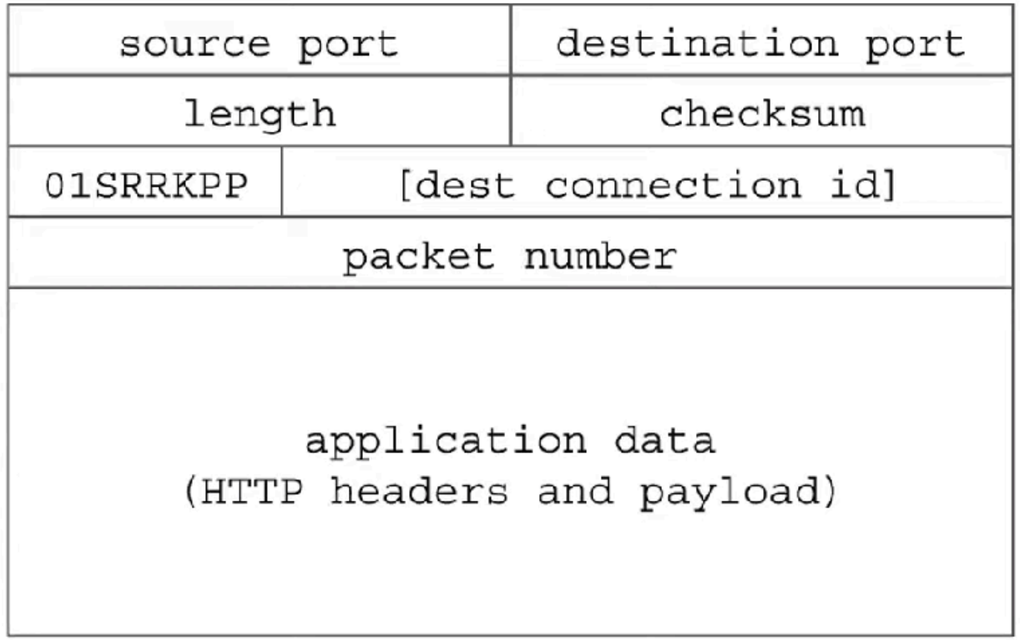
QUIC

Encrypt as much as possible

HTTP w/ TLS + TCP



HTTP w/ QUIC

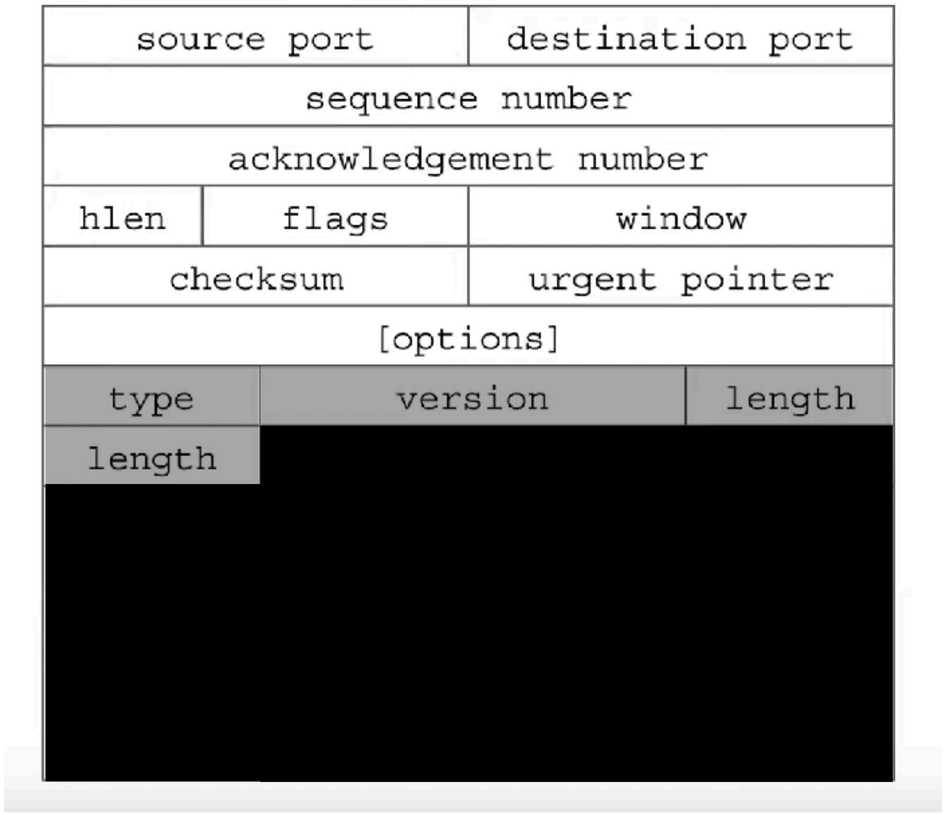


Slide stolen from: https://www.youtube.com/watch?v=31J8PoLW9iM&t=9104s

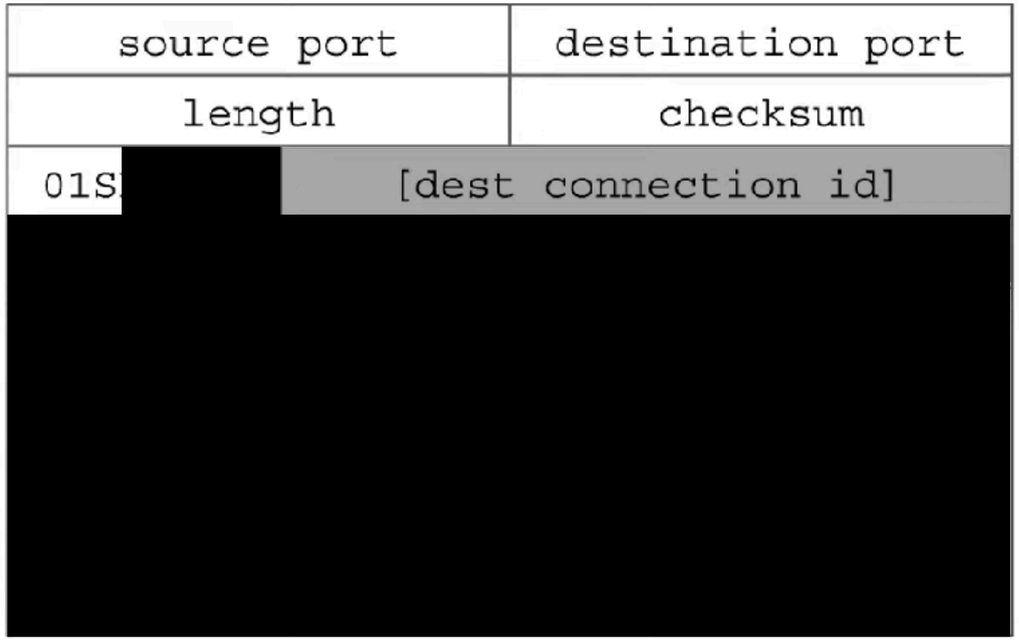
QUIC

Encrypt as much as possible

HTTP w/ TLS + TCP



HTTP w/ QUIC

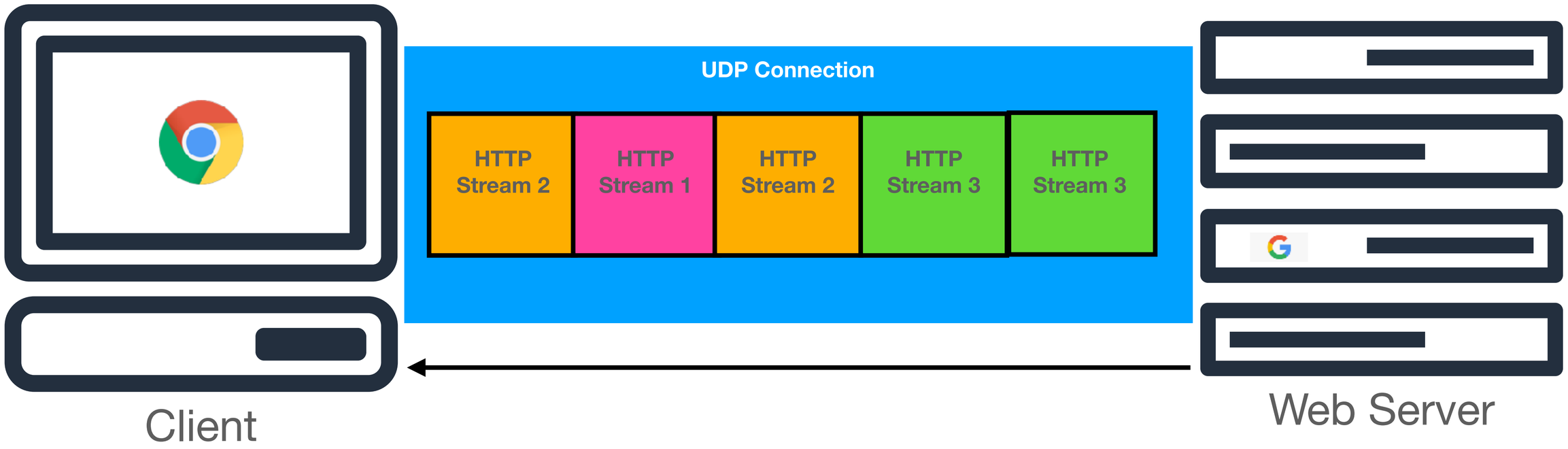


Slide stolen from: https://www.youtube.com/watch?v=31J8PoLW9iM&t=9104s

QUIC

Maintaining the Stream Abstraction

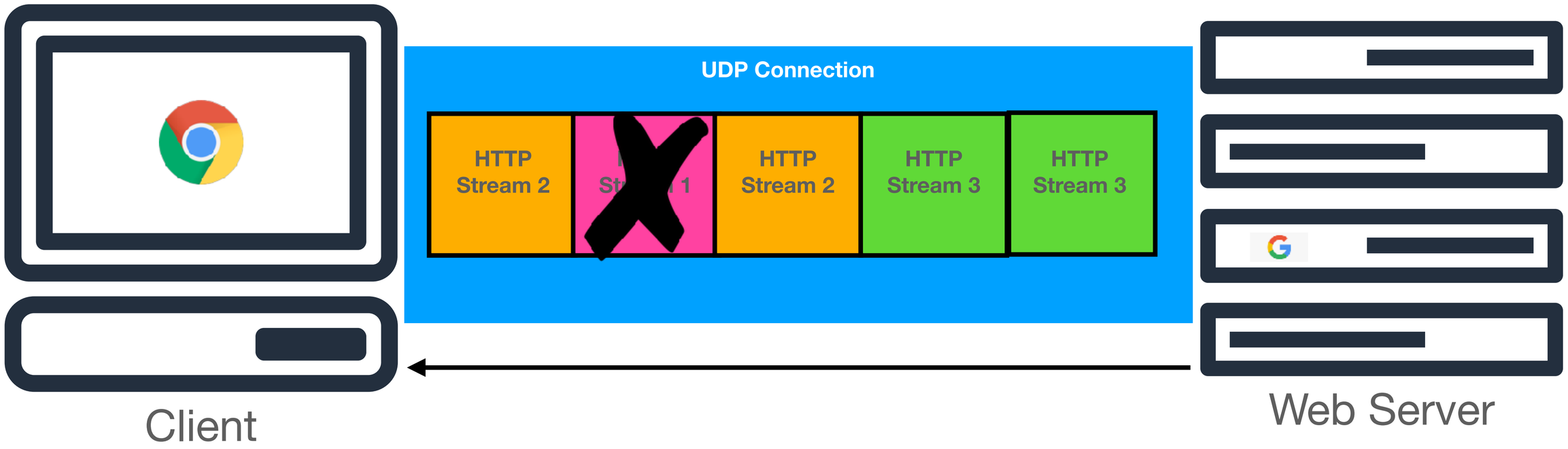
- QUIC uses the idea of a stream (with a stream_id) as a baseline abstraction for sending data between two endpoints, similar to HTTP/2



QUIC

Maintaining the Stream Abstraction

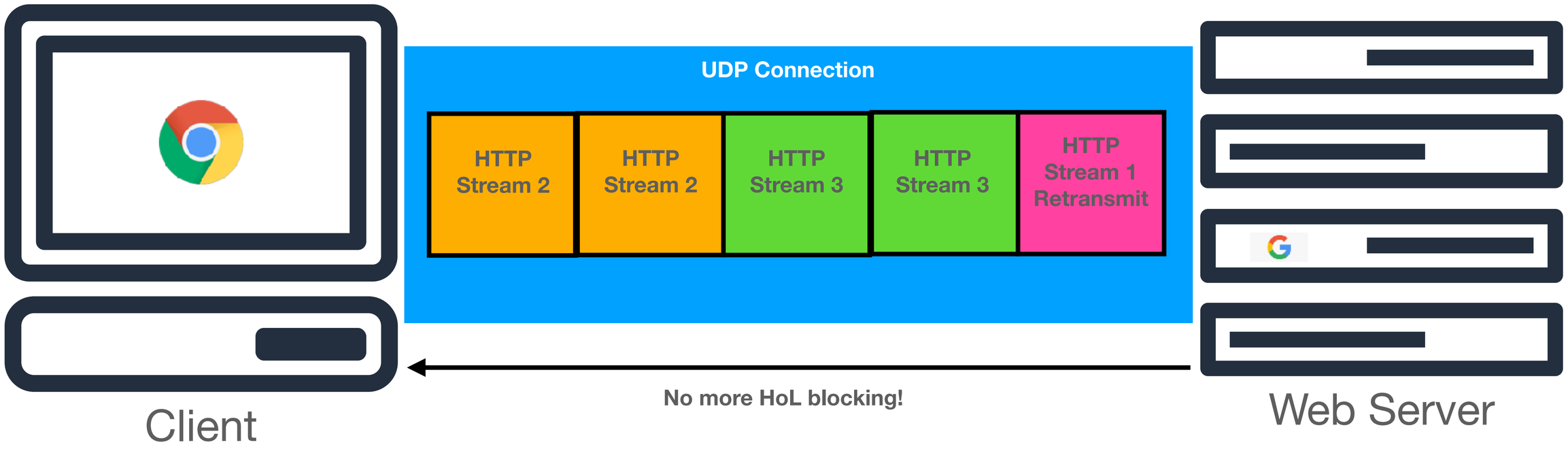
- QUIC uses the idea of a stream (with a stream_id) as a baseline abstraction for sending data between two endpoints, similar to HTTP/2



QUIC

Maintaining the Stream Abstraction

- QUIC uses the idea of a stream (with a stream_id) as a baseline abstraction for sending data between two endpoints, similar to HTTP/2



TCP vs. QUIC

Recovering from Losses

- TCP uses sequence numbers + acknowledgement numbers to identify whether or not a packet has been lost, and needs to be retransmitted
  - Unfortunately, sequence numbers mean two things: reliability **and** the order at which the bytes are supposed to be delivered to the receiver
  - On top of this, TCP retransmissions use the *same* sequence number, so it becomes very hard to know whether an ACK was sent for first transmission or a retransmission
  - TCP conflates transmission ordering AND delivery ordering in one number

TCP vs. QUIC

Recovering from Losses

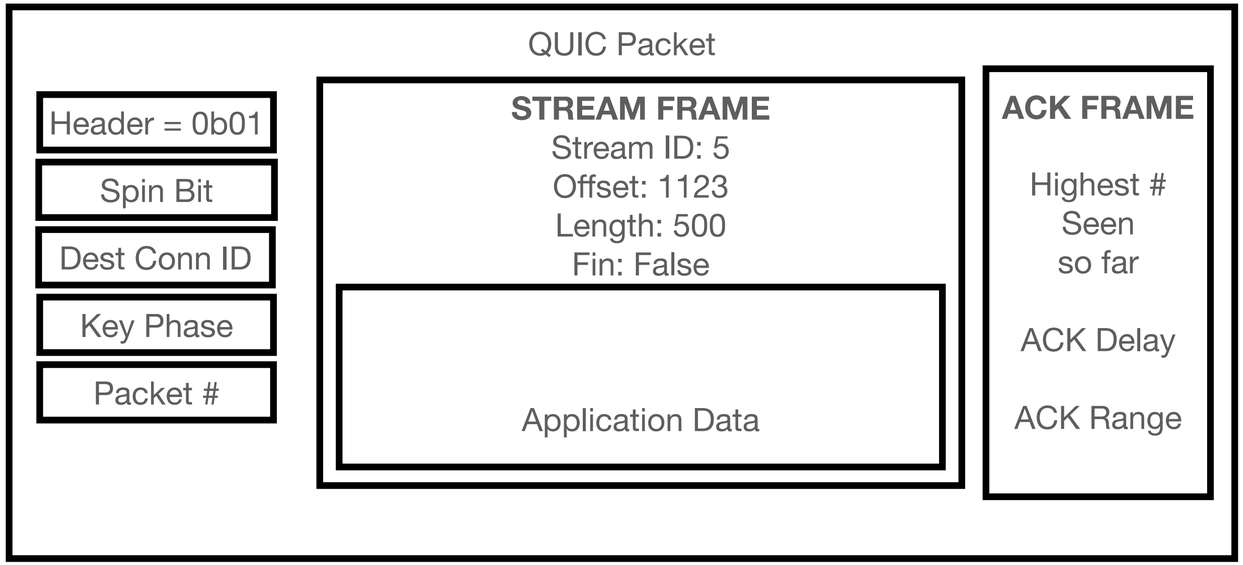
- QUIC decouples transmission and delivery ordering through its use of *streams*
  - Each packet contains a packet number, which is **unique and monotonically increasing, even on retransmission**
  - Clients will ACKNOWLEDGE packet numbers, and the server can identify if an outstanding packet has not been acknowledged… you can find the details at the link below
  - Each frame in a stream contains a *stream offset*, which alerts the client of how to properly reorder the packets on the delivery side
- Enables simpler loss detection than TCP

https://tools.ietf.org/id/draft-ietf-quic-recovery-27.html

QUIC

Packetization

- Packets can contain multiple types of frames (e.g., Stream frames, ACK frames, crypto frames)
- Stream frames contain stream IDs and *offsets* for the receiver to re-order out-of-order packets
- ACK frames contain acknowledgements for the highest packet number we've seen so far, and a range for what packets we've acked so far



QUIC

Connection Rebinding

- Because QUIC connections are over UDP, they can persist *beyond traditional network boundaries*, like your home NAT
  - No more resetting connection when your underlying network changes
- QUIC does this through the use of several unique variable length Connection IDs to identify the connection, with a protocol in place to verify the connection through a network change
- See RFC for notes on address spoofing + off-path packet attackers (something they've considered!)

QUIC

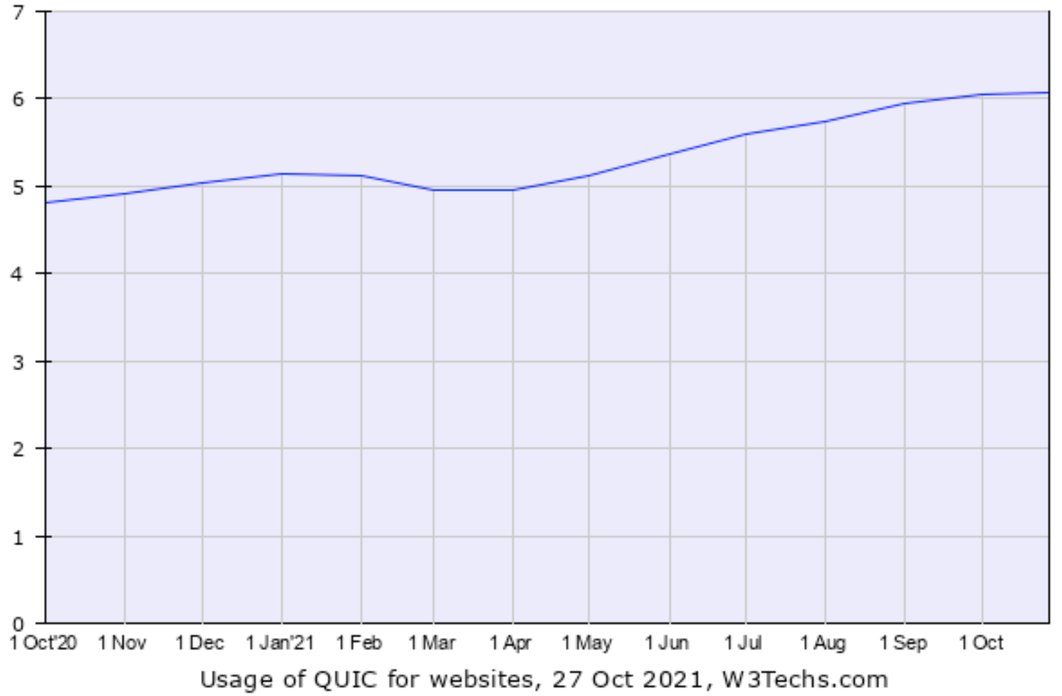
NATs, Middleboxes, Deployment Challenges

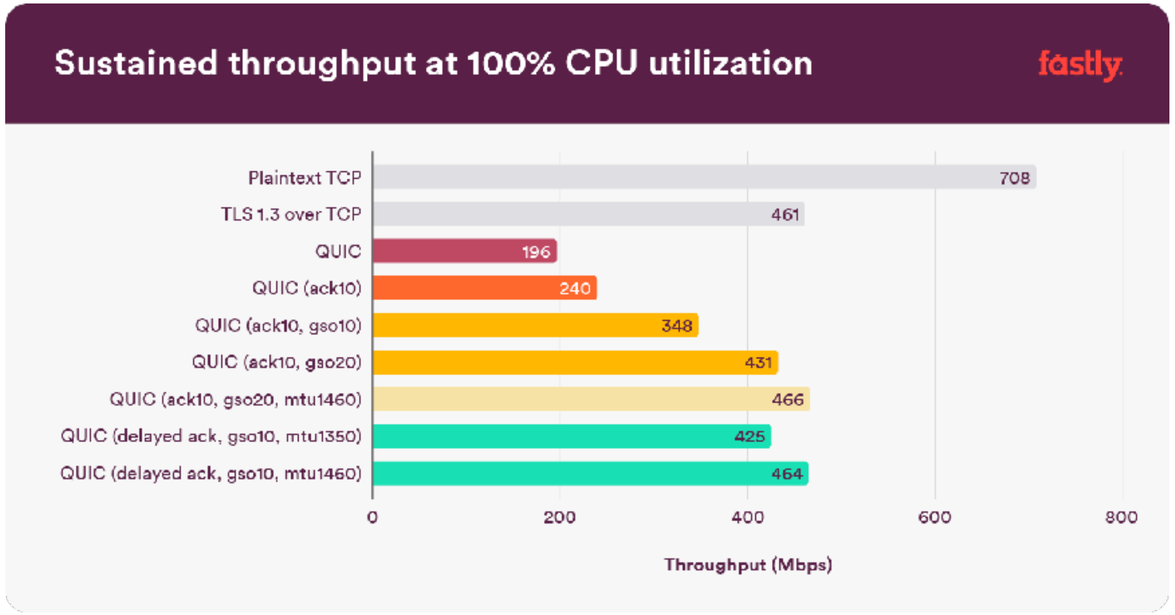
- Typically, NATs keep track of TCP connections by using a 5-tuple (src_port, src_ip, dst_port, dst_ip, protocol), and can maintain state because they have access to TCP headers
- Not all NATs speak QUIC yet, and even if they did, header information is encrypted, so they default to processing UDP packets, which could cause short timeouts and routing issues
- UDP-based protocols are susceptible to *reflection attacks,* where attackers use UDP servers with spoofed source ports to amplify their attack, and QUIC can be asymmetric on *inchoate client hello*
    - This is why QUIC has a REJ packet to start, but this increases the number of round trips required on initial connection. Probably a decent trade off.

QUIC Deployment

QUICly eating the world

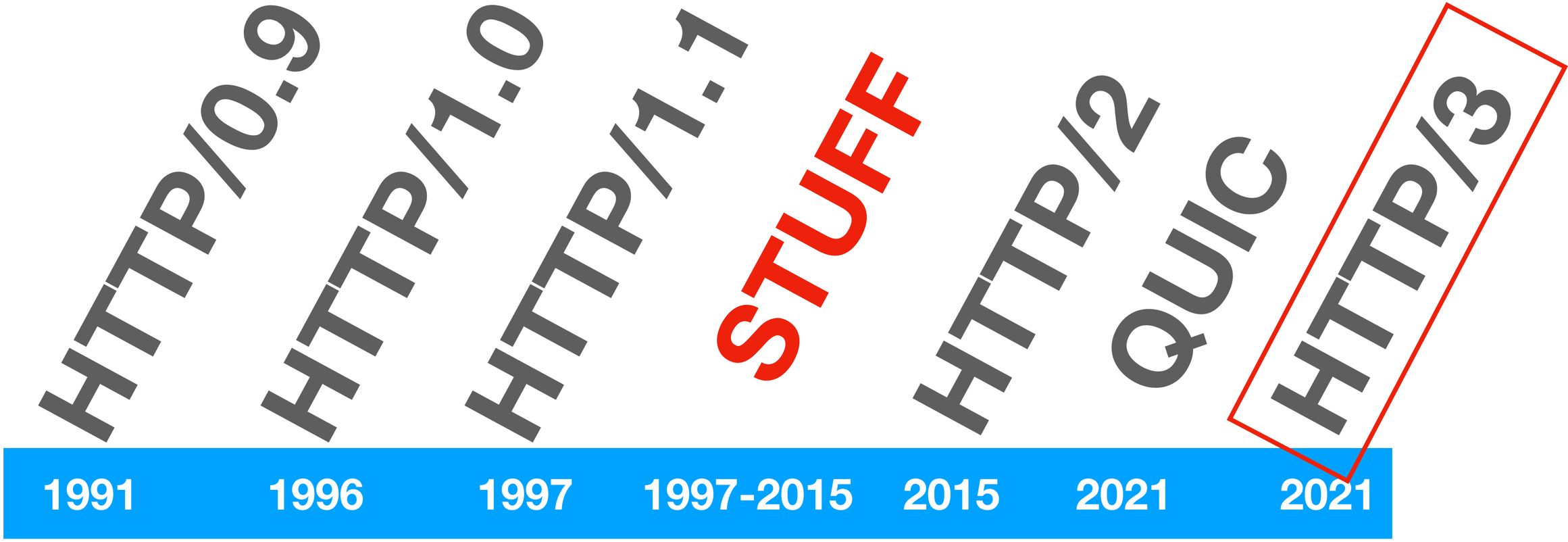
- QUIC was officially ratified by the IETF in May 2021 (RFC 9000)
- QUIC support already existed in Chrome for a while, but is now available in Firefox as well
- QUIC is being deployed everywhere
  - 6% of websites use QUIC, but will grow post RFC ratification
  - Google apps all use QUIC, 75% of Facebook uses QUIC
  - Some ISPs have reported that **20% of their packets were over QUIC**
- With appropriate tuning in high performance benchmarks, QUIC is so far as good as TLS 1.3 over TCP





https://w3techs.com/technologies/details/ce-quic
https://www.fastly.com/blog/measuring-quic-vs-tcp-computational-efficiency

A History of Web Protocols



HTTP/3 is HTTP over QUIC!

HTTP/3

Building HTTP over QUIC

- Still being iterated on by IETF (no RFC number yet)
- HTTP/3 uses the same abstraction as HTTP/2 (e.g., streams, frames, etc.), except it utilizes these streams as supported by QUIC rather than implementing on top of TCP
- This causes some notable new challenges:
  - HPACK, the clever header encryption scheme, cannot be enforced anymore without causing HoL blocking (recall that headers MUST appear before response data in HTTP/2)
  - HTTP/2 enjoyed stream prioritization, which is hard to implement in the transport layer on top of everything else

HTTP/3 vs. HTTP/2

Notable Changes?

- HPACK is updated to QPACK, which is designed to allow for out-of-order header data (and updating dynamic tables accordingly)
  - Essentially, adds more ability for client to control when to use a dynamic table entry – no need to wait to update an entry or read a table entry before processing a request
- Removed stream prioritization altogether!
  - Deemed too challenging to use for clients and offered little guarantees anyway, so it is being discussed independently

Recap

- The web has drastically changed over time, with developers doing more than ever before and websites becoming increasingly complex
  - But for a long time, our protocols didn’t match the growing complexity of the world
- New protocols like SPDY, HTTP/2 were useful in working within our paradigm, but there is **change** afoot!
  - People are not liking TCP as much, and companies like Google are starting to throw their weight around in envisioning a new future for layering requirements
- We are redefining “end-to-end” abstractions… let’s see how it goes :)

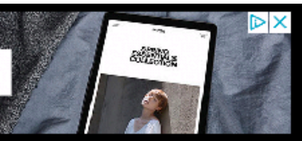
Web Content

cs249i

Modern Websites

Third Party Resources

- Modern websites rely on many different types of *third-party resources* to provide services to keep their websites functional
  - Third party resources are ones served by external parties – so for example, if you are on cnn.com, any resource served from a domain that is NOT cnn.com (e.g., doubleclick.com, google-analytics.com)
  - These resources could be anything from static images to JavaScript libraries to analytics, advertising, the list goes on…



Trump escalates January 6 cover-up

ANALYSIS

The former President is trying to keep the House select committee probing January 6 from seeing a list of documents as he ramps up his political comeback

KFILE Trump lawyer said 'courage and the spine' would help Pence send election to the House in comments before January 6

Brian Stelter's ominous prediction: Imagine it's 2022 and ...

January 6 committee is losing patience with Trump's former chief of staff Mark Meadows as it seeks his testimony

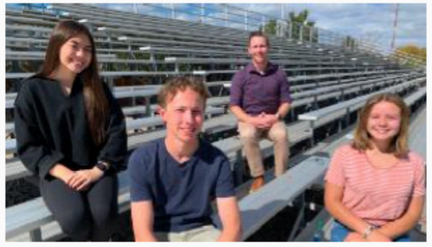
Washington Post report rebuts the January 6 alt-reality that Tucker Carlson promotes

Biden says US 'continuing to suffer' from Trump's decision to pull out of Iran nuclear deal

LIVE UPDATES

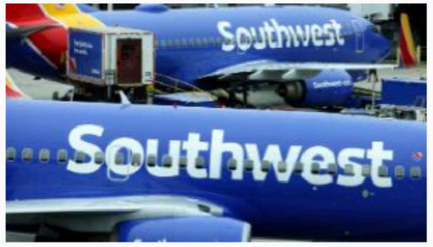
Astros top Braves 9-5 in World Series Game 5

• **Trivia:** Can you name the only player to play in all 3 cities that the Braves have called home?
• **Analysis:** The Braves may win the World Series. But they're striking out with some fans



Students are fed up with raging adults at school board meetings

• A Texas lawmaker is investigating 850 books on race and gender that could cause 'discomfort' to students
• **Opinion:** When parents scream at school board meetings, how can I teach their children?



Southwest launches investigation into pilot reportedly using anti-Biden phrase on flight

Reporter reveals what Lindsey Graham said during January 6 riot

White House press secretary tests positive for Covid, last saw Biden Tuesday

BREAKING Japan's Fumio Kishida defies expectations as ruling party keeps majority

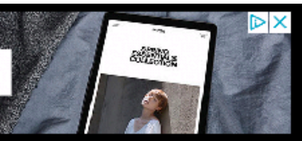
Aurora borealis puts on a gorgeous show

'Step up or step out': Lawmaker calls out attorney general

Police investigating desecration of Torah scroll at fraternity

COP26 climate talks off to an ominous start after weak G20 leaders' meeting

Video shows passengers fleeing knife attack on train



PODCAST: Tug of War | TRENDING: World Series | Pope and Biden | Tuesday elections | Halloween train attack | Economic bills | G20 summit | Box office |

Trump escalates January 6 cover-up

The former President is trying to keep the House select committee probing January 6 from seeing a list of documents as he ramps up his political comeback

KFILE Trump lawyer said 'courage and the spine' would help Pence send election to the House in comments before January 6

Brian Stelter's ominous prediction: Imagine it's 2022 and ...

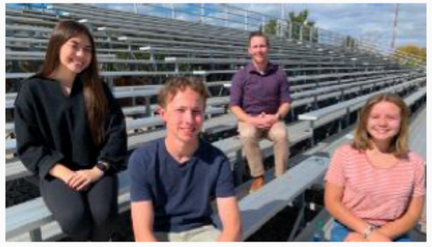
January 6 committee is losing patience with Trump's former chief of staff Mark Meadows as it seeks his testimony

Washington Post report rebuts the January 6 alt-reality that Tucker Carlson promotes

Biden says US 'continuing to suffer' from Trump's decision to pull out of Iran nuclear deal

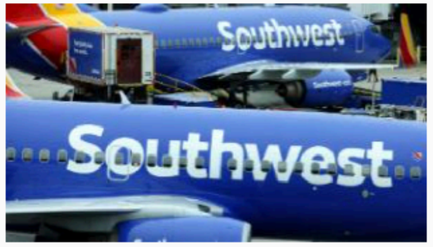
Astros top Braves 9-5 in World Series Game 5

• **Trivia:** Can you name the only player to play in all 3 cities that the Braves have called home?
• **Analysis:** The Braves may win the World Series. But they're striking out with some fans



Students are fed up with raging adults at school board meetings

• A Texas lawmaker is investigating 850 books on race and gender that could cause 'discomfort' to students
• **Opinion:** When parents scream at school board meetings, how can I teach their children?



Southwest launches investigation into pilot reportedly using anti-Biden phrase on flight

Reporter reveals what Lindsey Graham said during January 6 riot

White House press secretary tests positive for Covid, last saw Biden Tuesday

BREAKING Japan's Fumio Kishida defies expectations as ruling party keeps majority

Aurora borealis puts on a gorgeous show

'Step up or step out': Lawmaker calls out attorney general

Police investigating desecration of Torah scroll at fraternity

COP26 climate talks off to an ominous start after weak G20 leaders' meeting

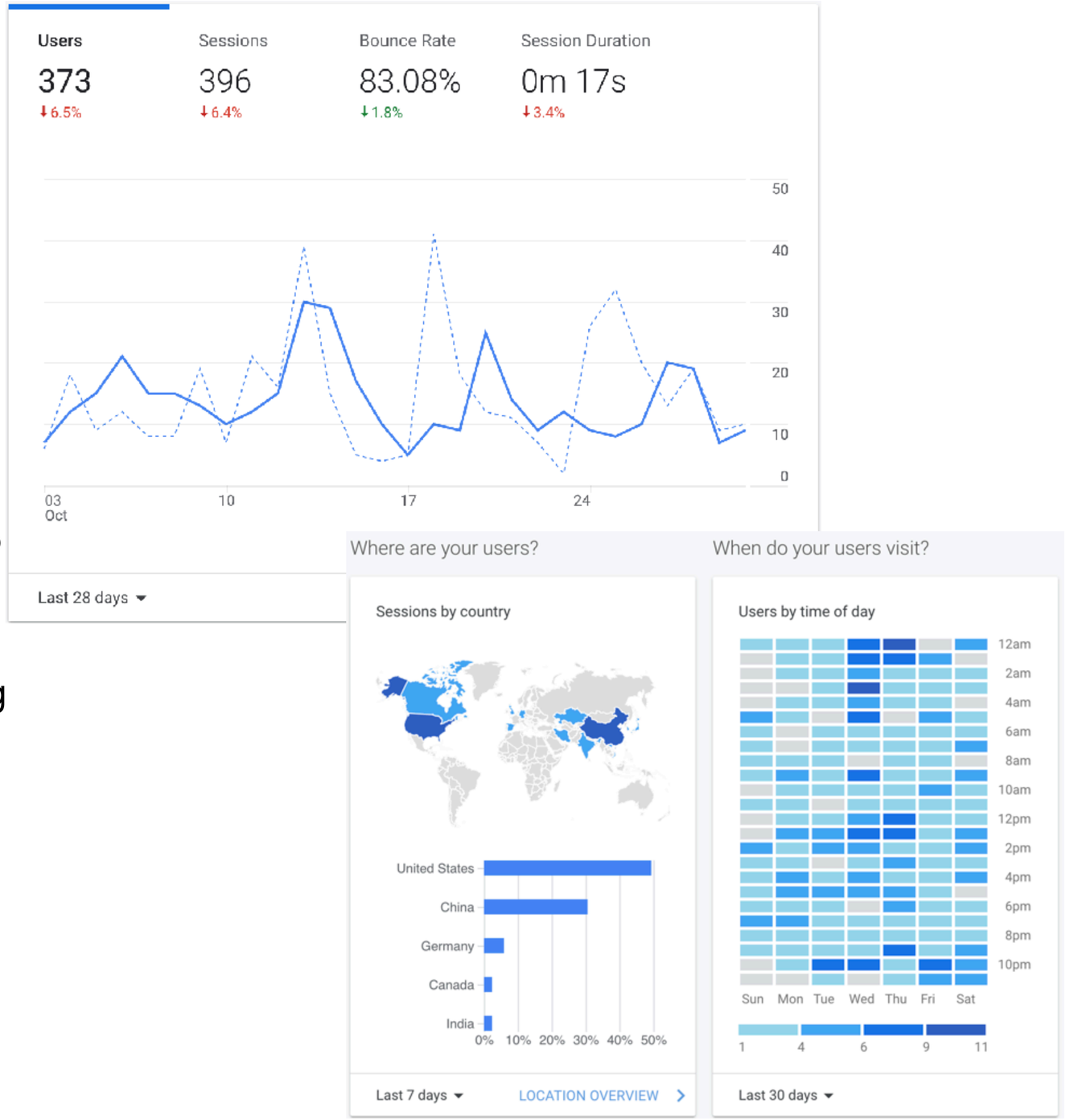
Video shows passengers fleeing knife attack on train




Modern Websites

Analytics

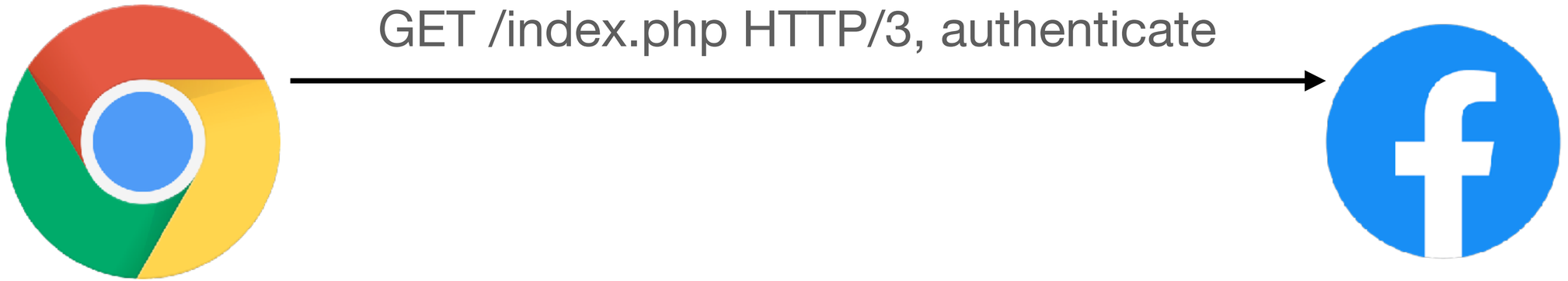
- Many websites rely on analytics on their users to continue to improve their services
  - For example, Google provides Google Analytics, which appears on an estimated 70% of the top websites
- As an analytics user, you can see where your clients are connecting from, you can see how long they spent on the page, what devices they're connecting from, and a ton of other interesting details
  - These are typically scoped to a single request, but in recent years, companies have been expanding the scope of what they know about users…



Web Tracking

Cookies and Code

- Major companies typically use *cookies* to offer extended functionality for websites (e.g., keeping you logged in, keeping certain settings stored in your browser, etc.)



Web Tracking

Cookies and Code

- Major companies typically use *cookies* to offer extended functionality for websites (e.g., keeping you logged in, keeping certain settings stored in your browser, etc.)

Web Tracking

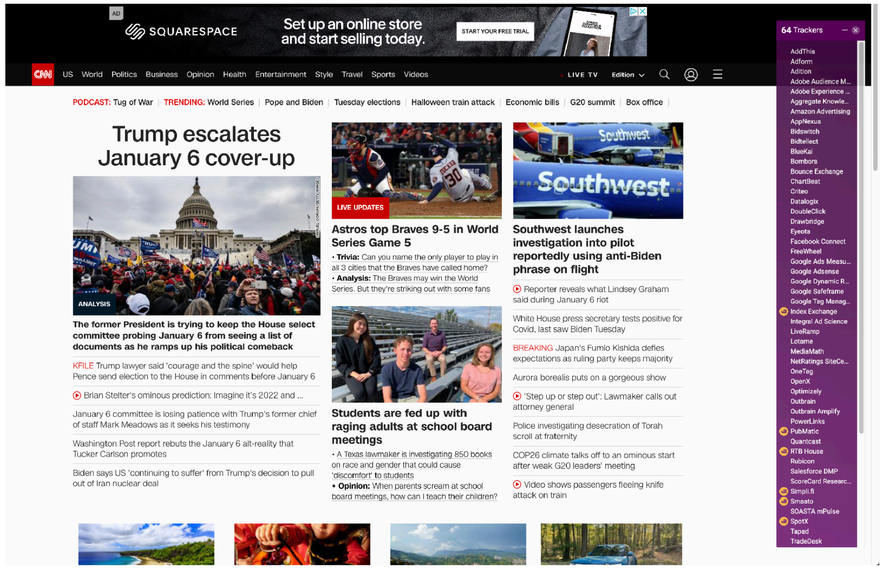
Cookies and Code

- Major companies typically use *cookies* to offer extended functionality for websites (e.g., keeping you logged in, keeping certain settings stored in your browser, etc.)
- Once a cookie is set, the browser attaches a cookie to every subsequent request sent out for that particular domain
  - Cookies are by default scoped to the first-party domain that set the cookie
  - No other domains can read the cookie value!
- …then how does web tracking work?

Web Tracking

Cookies and Code

GET / HTTP/3

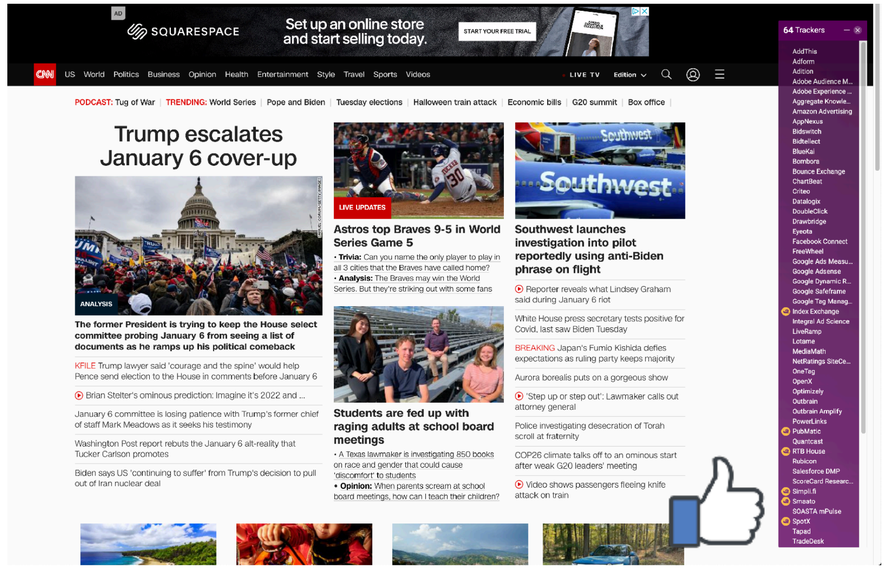


Web Tracking

Cookies and Code

GET / HTTP/3

GET /facebook-like.js HTTP/3



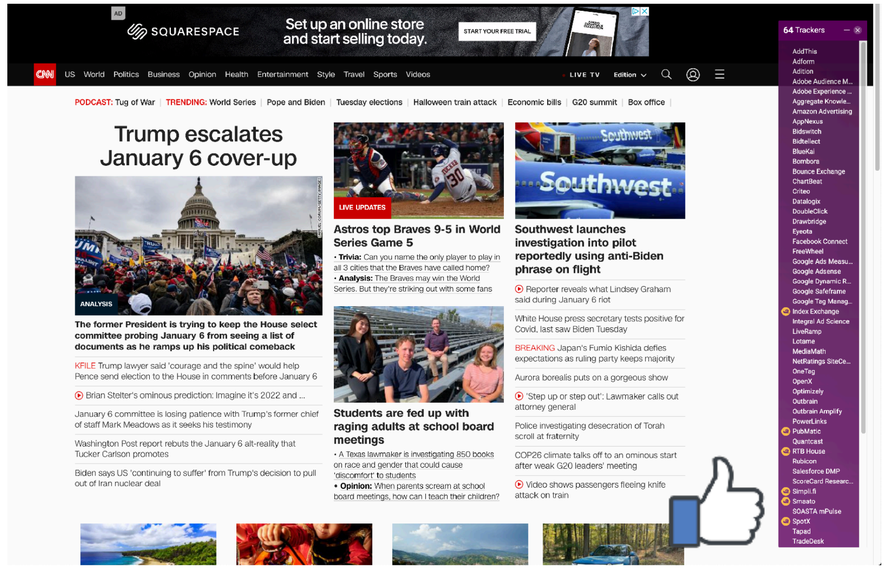
Web Tracking

Cookies and Code

GET / HTTP/3

GET /facebook-like.js HTTP/3

Cookie: User=Deepak, Referer=cnn.com



- With this request, companies can link your cookie to your browsing data (e.g., through Referer header, Host headers, Origin, or just JavaScript)

Web Tracking

Browser Fingerprinting

- Websites can also fingerprint you effectively with *browser fingerprinting*, which is a technique that leverages all your settings to identify you, and stores this in a cookie on your browser
  - https://iamunique.org
- So long as JavaScript can run (by third-parties), you run the risk of being “followed” on the web


```
{
 "user-agent": "Mozilla/5.0 (Macintosh; Intel Mac OS X 10.15; rv:93.0) Gecko/20100101 Firefox/93.0",
 "accept": "text/html,application/xhtml+xml,application/xml;q=0.9,image/avif,image/webp,*/*;q=0.8",
 "accept-encoding": "gzip, deflate, br",
 "accept-language": "en-US,en;q=0.5",
 "upgrade-insecure-requests": "1",
 "referer": "https://amiunique.org/",
 "userAgent-js": "Mozilla/5.0 (Macintosh; Intel Mac OS X 10.15; rv:93.0) Gecko/20100101 Firefox/93.0",
 "platform": "MacIntel",
 "cookies": "yes",
 "timezone": 420,
 "languages-js": "en-US,en",
 "ad": "no",
 "doNotTrack": "NC",
 "navigator_properties": [
  "vibrate",
  "javaEnabled",
  "getGamepads",
  "getVRDisplays",
  "mozGetUserMedia",
  "sendBeacon",
  "requestMediaKeySystemAccess",
  "registerProtocolHandler",
  "taintEnabled",
```

Web Tracking

Browser Fingerprinting

- Websites can also fingerprint you effectively with *browser fingerprinting*, which is a technique that leverages all your settings to identify you, and stores this in a cookie on your browser
  - https://iamunique.org
- So long as JavaScript can run (by third-parties), you run the risk of being “followed” on the web


```
{
 "user-agent": "Mozilla/5.0 (Macintosh; Intel Mac OS X 10.15; rv:93.0) Gecko/20100101 Firefox/93.0",
 "accept": "text/html,application/xhtml+xml,application/xml;q=0.9,image/avif,image/webp,*/*;q=0.8",
 "accept-encoding": "gzip, deflate, br",
 "accept-language": "en-US,en;q=0.5",
 "upgrade-insecure-requests": "1",
 "referer": "https://amiunique.org/",
 "userAgent-js": "Mozilla/5.0 (Macintosh; Intel Mac OS X 10.15; rv:93.0) Gecko/20100101 Firefox/93.0",
 "platform": "MacIntel",
 "cookies": "yes",
 "timezone": 420,
 "languages-js": "en-US,en",
 "ad": "no",
 "doNotTrack": "NC",
 "navigator_properties": [
  "vibrate",
  "javaEnabled",
  "getGamepads",
  "getVRDisplays",
  "mozGetUserMedia",
  "sendBeacon",
  "requestMediaKeySystemAccess",
  "registerProtocolHandler",
  "taintEnabled",
  "permissions"
```

Web Tracking

Prevalence of Major Companies

- Major companies have large presences on the web, and as a result, can see the majority of websites that you visit
  - Google appears on 82.2% of the Top 1M (by AS), because of analytics and advertising services
  - Facebook appears on 34.1%, to enable social sharing + tracking

Company	Prevalence on Top 1M
Google	82.2%
Facebook	34.1%
Amazon	32.6%
Cloudflare	30.7%
Akamai	20.3%
MaxCDN	19.0%
Edgecast	17.9%
Fastly	15.5%
SoftLayer	11.8%
Twitter	11.2%

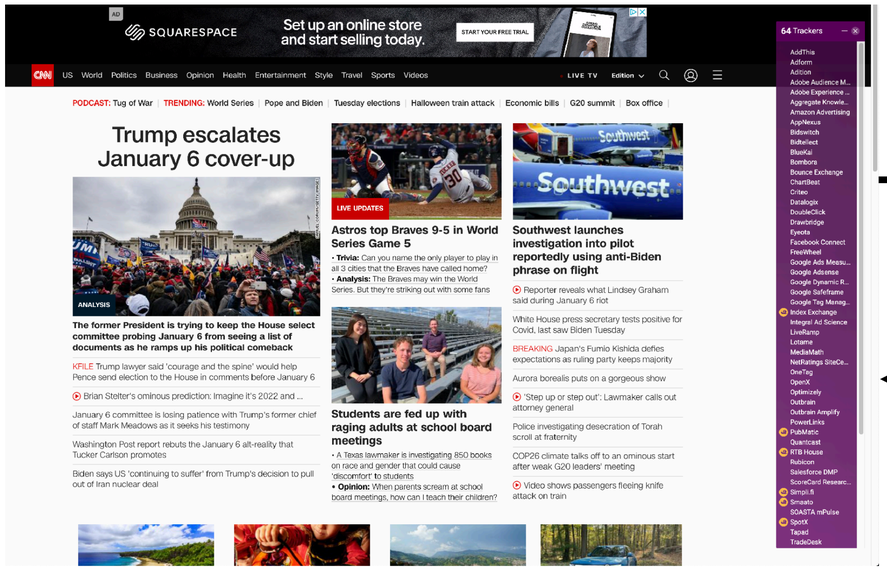
Web Tracking

Cookie Syncing

- Even if a company is not available on every website, companies often times *share* cookie information
  - “Cookie Synchronization: Everything You Always Wanted to know but were afraid to ask” – WebConf 2019
- Core idea is simple: If you have a collaboration agreement with another third-party, you simply *redirect* requests to them upon receiving requests

Web Tracking

Cookie Syncing



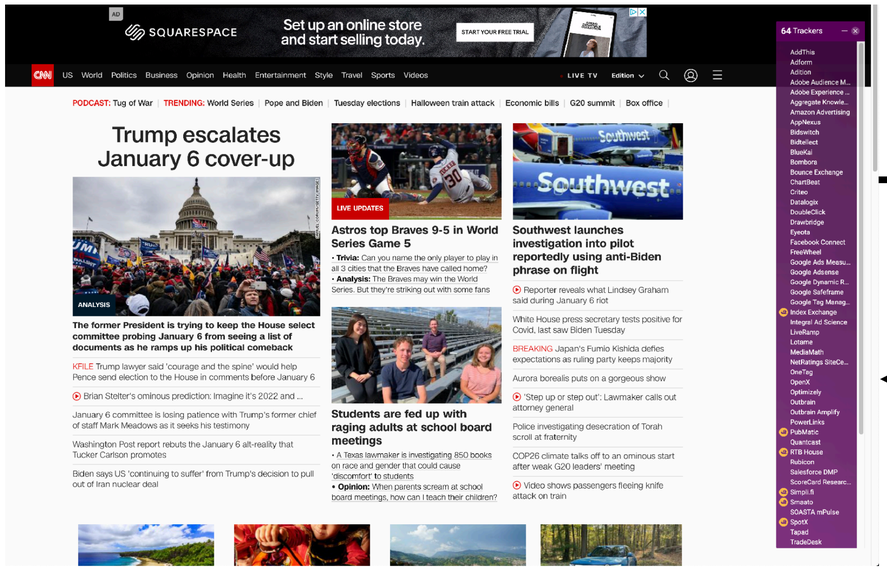
GET tracker.com/pixel.jpg

Response, Set-Cookie: User=user123

tracker.com

Web Tracking

Cookie Syncing



GET advertiser.com/pixel.jpg

Response, Set-Cookie: User=userABC

advertiser.com

Web Tracking

Cookie Syncing

GET tracker.com/pixel.jpg, cookie=user123

tracker.com

advertiser.com

Web Tracking

Cookie Syncing

GET tracker.com/pixel.jpg, cookie=user123

REDIRECT, advertiser.com?syncID=user123&publisher=nytimes.com

tracker.com

advertiser.com

Web Tracking

Cookie Syncing

GET tracker.com/pixel.jpg, cookie=user123

REDIRECT, advertiser.com?syncID=user123&publisher=nytimes.com

GET syncID=user123, cookie=userABC

tracker.com

advertiser.com

Web Tracking

Cookie Syncing

GET tracker.com/pixel.jpg, cookie=user123

REDIRECT, advertiser.com?syncID=user123&publisher=nytimes.com

GET syncID=user123, cookie=userABC

tracker.com

advertiser.com

- Third-parties with cookie syncing is enabled on 78% of modern websites :(

Web Tracking
Cookie Ghostwriting

- Not all first-party cookies *should* be treated the same!

Web Tracking
Cookie Ghostwriting

- Not all first-party cookies *should* be treated the same!

GET tracker.com/script.js

tracker.com

Web Tracking
Cookie Ghostwriting

- Not all first-party cookies *should* be treated the same!

GET tracker.com/script.js

tracker.com

script.js

document.cookie = "user=userABC"

Web Tracking
Cookie Ghostwriting

- 42% of identifier cookies are *ghostwritten* in modern websites

GET tracker.com/script.js

tracker.com

script.js

document.cookie = "user=userABC"

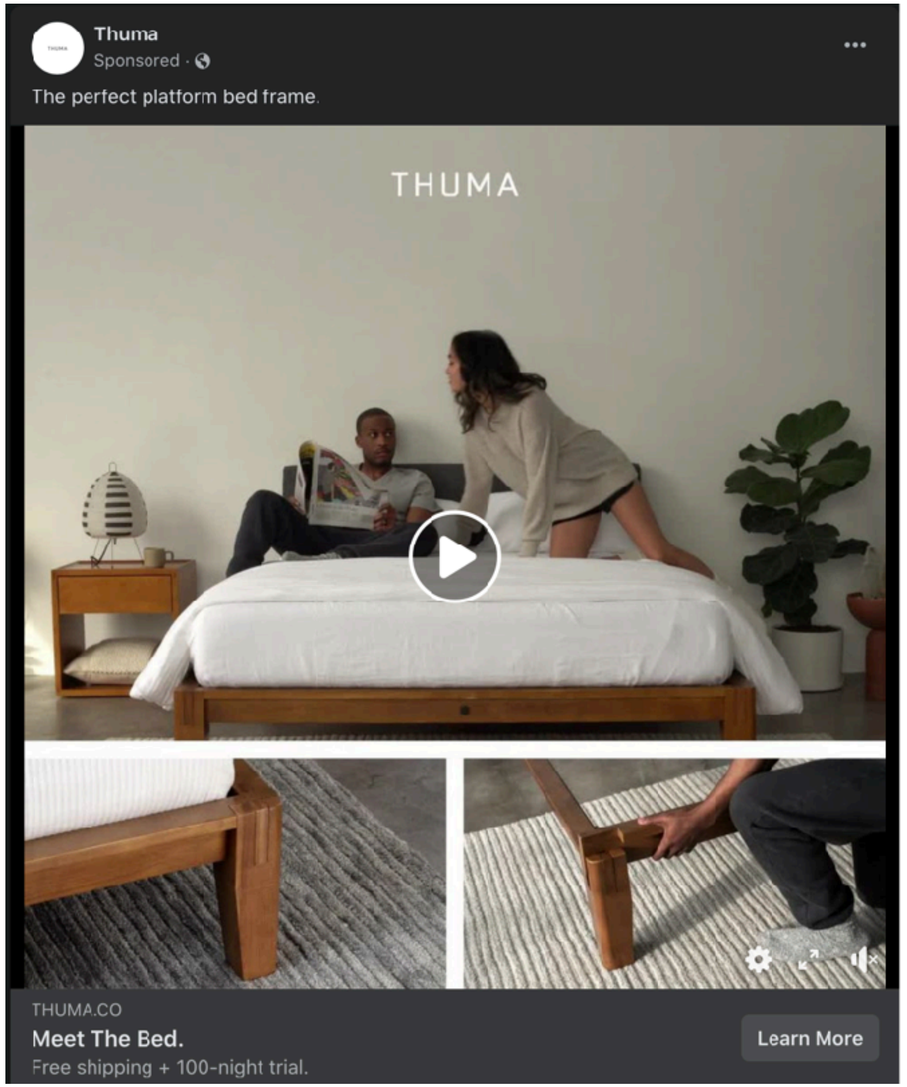
advertiser.com

Why is there so much tracking?

Online Advertising

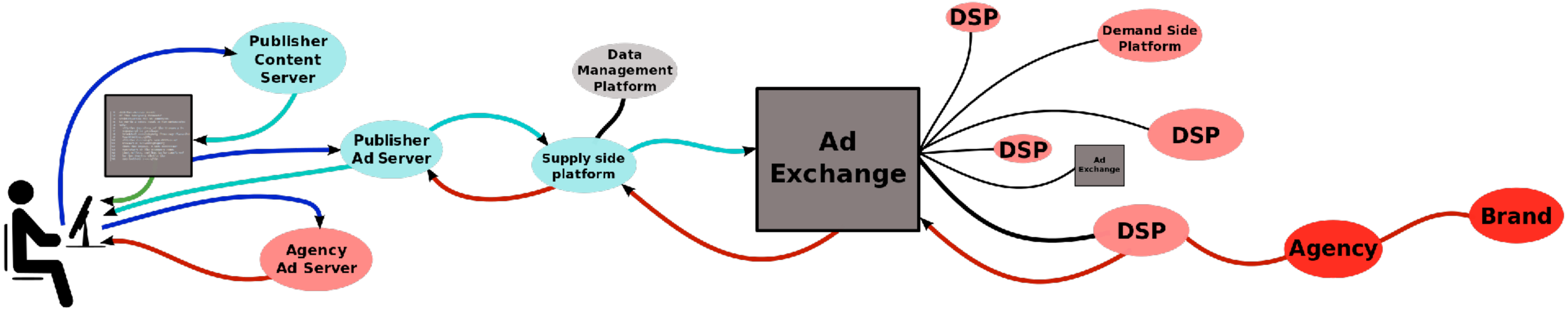
The Best Thing Since Sliced Bread! Available for $4.99 at your local Costco.

- Companies typically track you around the web to build profiles for *targeted advertising*
  - The more targeted your advertising, the more revenue you can make from advertisers who are potentially willing to give you more money to sell the ad spot
  - Useful for advertisers to know if people with your browsing habits, your properties, your whatever are browsing on the web



Online Advertising

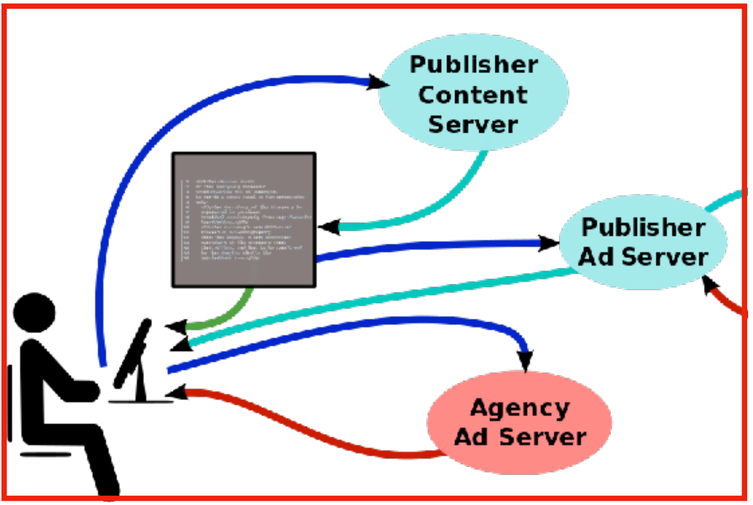
The Many Internet Players in Advertising



https://upload.wikimedia.org/wikipedia/commons/thumb/d/da/Adservingfull.svg/2880px-Adservingfull.svg.png

Online Advertising

Publishers

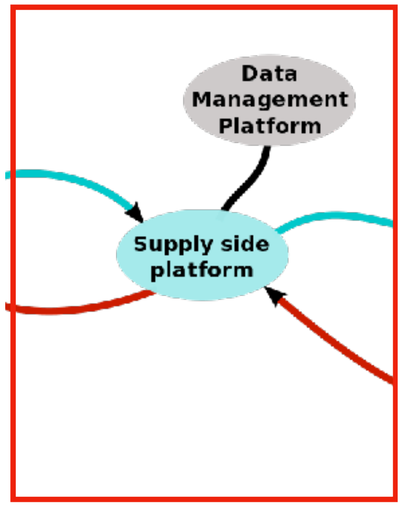


- Publishers (e.g., nytimes.com, cnn.com, other websites) often have advertising space that they are hoping to make revenue off of
- In some cases, publishers have explicit agreements with companies and can sell their space that way

https://upload.wikimedia.org/wikipedia/commons/thumb/d/da/Adservingfull.svg/2880px-Adservingfull.svg.png

Online Advertising

Supply Side Platforms

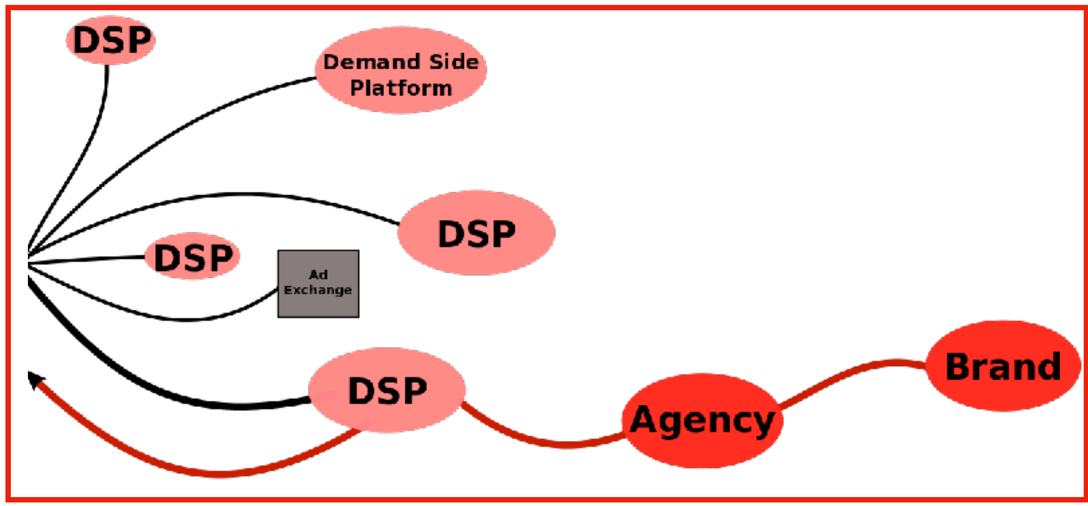


- If a publisher wants to place the ad spot on the open advertising market, they typically go through an intermediary called a Supply Side Platform (SSP)
  - Examples: Pubmatic, Rubicon Project, Verizon Media, etc.
- This aggregates information about the client (through a DMP) and participates in ad exchange

https://upload.wikimedia.org/wikipedia/commons/thumb/d/da/Adservingfull.svg/2880px-Adservingfull.svg.png

Online Advertising

Demand Side Platforms



- On the other end of the pipeline, you have advertisers
- There are analogous entities called demand side platforms, which participate in Real-Time Bidding, which is a real-time auction for ad space (examples: Google DoubleClick, QuantCast, Criteo, Adform)
  - Typically happens in < 100ms

https://upload.wikimedia.org/wikipedia/commons/thumb/d/da/Adservingfull.svg/2880px-Adservingfull.svg.png

Online Advertising

Ad Exchanges

- Advertising exchanges receive spots from supply side, and facilitate real time bidding from the demand side based on properties of the ad spot
    - Examples: Google DoubleClick, Facebook Exchange, PubMatic, Microsoft Advertising

https://upload.wikimedia.org/wikipedia/commons/thumb/d/da/Adservingfull.svg/2880px-Adservingfull.svg.png

Online Advertising

Bid Requests


```
        "site": {
            "id": "1234",
            "name": "Example Site",
            "domain": "examplesitedomain.com",
            "mobile": 1,
            "amp": 0,
            "pub": {
                "id": "9876",
                "name": "Example Publisher, Inc.",
                "domain": "examplepubdomain.com"
            }
        },
        "user": {
            "id": "a0af45c77890045deec100acb8443baff57c",
            "consent": "ihdknkhkq8y",
            "buyeruid": "fcd4282456238256034abcdef220d9aa5892",
            "yob": 1990,
            "gender": "F",
            "ext": {
                "consented_providers_settings": {
                    "consented_providers": [
                        1,
                        52,
                        45,
                        23
                    ]
                }
            }
        },
        "device": {
            "type": 4,
            "ifa": "8846d6fa10008bceaaf322908dfcb221",
            "ip": "1.2.3.4",
            "ua": "Mozilla/5.0 (Macintosh; U; Intel Mac OS X 10.6; en-US; rv:1.9.2.16) Gecko/20110319
Firefox/3.6.16",
            "make": "Apple",
            "model": "iPhone",
            "hwv": "6s",
            "os": 13,
            "osv": "11.4.1",
            "mccmnc": "310-005",
            "geo": {
```


https://protocol.bidswitch.com/rtb/request-examples.html

Online Advertising

Bid Response


```
{
    "id":"d7d1e107-987h",
    "cur":"usd",
    "ext":{
        "protocol":"6.0"
    },
    "seatbid":[
        {
            "seat":"4",
            "bid":[
                {
                    "id":"qwerty-098765",
                    "item":"asdf-7890",
                    "price":1.45,
                    "cid":"app-mraid-campaign-3442",
                    "burl":"https://adserver.com/winnotice?impid=102&winprice=${AUCTION_PRICE}",
                    "macro":[
                        {
                            "key":"TIMESTAMP",
                            "value":"1127987134"
                        }
                    ],
                    "ext":{
                        "agency_id":"agency_123",
                        "advertiser_name":"example advertiser"
                    },
                    "media":{
                        "ad":{
                            "id":"creative_id_1234",
                            "adomain":[
                                "example.com",
                                "example.io"
                            ],
                            "cat":[
                                "cat_1",
                                "cat_2"
                            ],
                            "lang":"en",
                            "attr":[
                                3,
                                7
```


https://protocol.bidswitch.com/rtb/response-examples.html

Online Advertising

Bidding for Ad Spots

- Real-time bidding is an auction process that is kicked off when a publisher tells an advertising network that they have an open ad-spot with certain properties
- Two most widely used methods of auctioning
  - Waterfall bidding
  - Header bidding

Online Advertising

Waterfall Bidding

- Publishers would pre-define a hierarchy of advertising networks that they wanted to ask in order (e.g., in a waterfall) about any given advertising spot
- Publishers would then set a floor bid rate that they needed for the ad spot
    - The first network to fulfill the floor would win the spot, but floor price goes down with lower priority
- Problems:
    - Slow (serial computation)
    - Anti-competitive!
        - Google had both an SSP and a DSP, which often meant they got first pick at ad spots

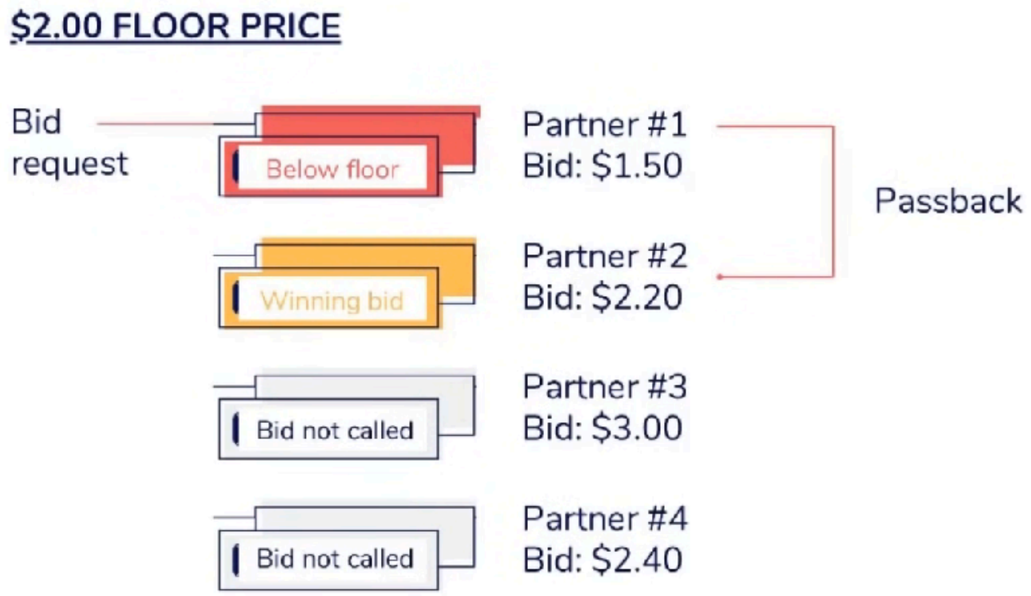


Image stolen from: https://www.youtube.com/watch?v=FcQnESUjOIY

Online Advertising

Header Bidding

- Every DSP is offered the auction at the *same time*, and DSPs are incentivized to provide their true value for the advertising spot (theoretically)
  - This typically happens in 100 – 200ms
- Two options:
  - Client-side header bidding (happens in JavaScript), potentially makes the page slower, but have finer grained access to cookies
  - Server-side header bidding (happens in the SSP), can be faster, but requires cookie syncing, could make things slower

When the business model *is* the privacy violation

APRIL 12, 2018 BY ARVIND NARAYANAN

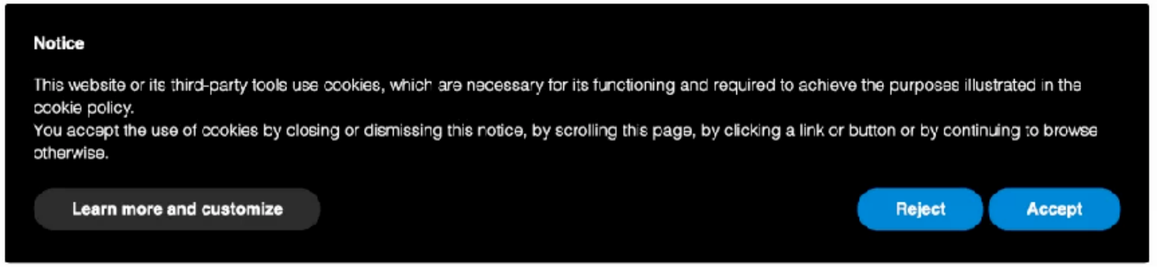

BRACE YOURSELVES
REGULATION IS COMING
memegenerator.net

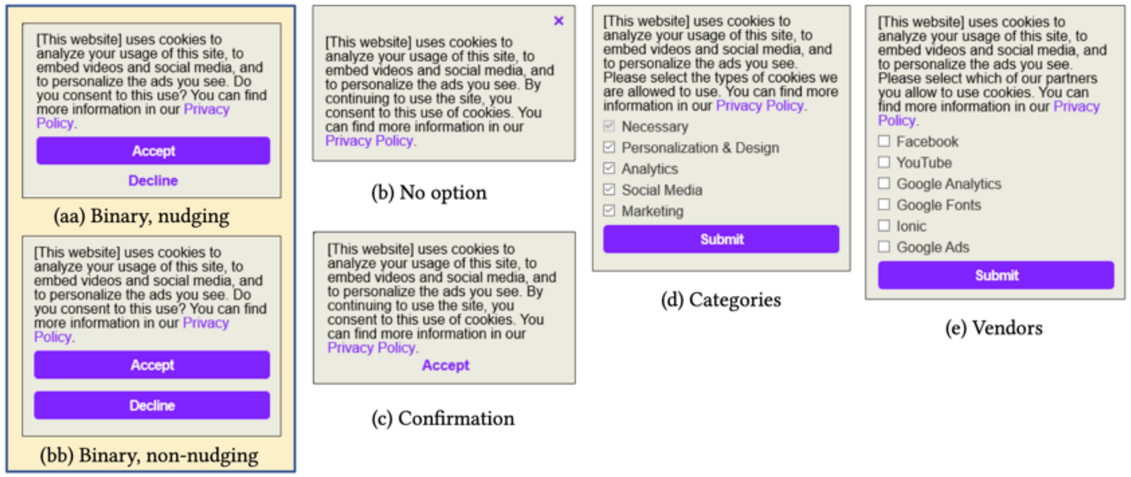
Regulation
GDPR, CCPA

- We've seen a big regulation push in the last five years around issues of online privacy and tracking
  - General Data Protection Regulation (GDPR), is an EU law on data protection and privacy for the European Economic Area
  - California Consumer Privacy Act (CCPA) is a state statue which aims to enhance consumer protections for Californians
- Both of these laws mandate all kinds of rules for the storing of personally identifiable data (e.g., IP addresses, cookies!), how long these things can be stored about users on the server side, etc.

Regulation
Cookie Banners

- If you use cookies, you must:
  - Inform users that your site/app uses cookies
  - Explain how cookies work and what the site uses them for
  - Obtain informed consent **prior** to storing those cookies on the user's device
- Need to provide users a **clear** and **easy** way to opt-out of cookie-tracking on a website
  - Steep fines (4% of annual revenue) if you do not comply
- Unfortunately, cookie-banners are being designed in terrible ways… and consent is broken





https://arxiv.org/pdf/1909.02638.pdf