2024 44th IEEE International Conference on Distributed Computing Systems

Dissecting the Applicability of HTTP/3 in Content Delivery Networks

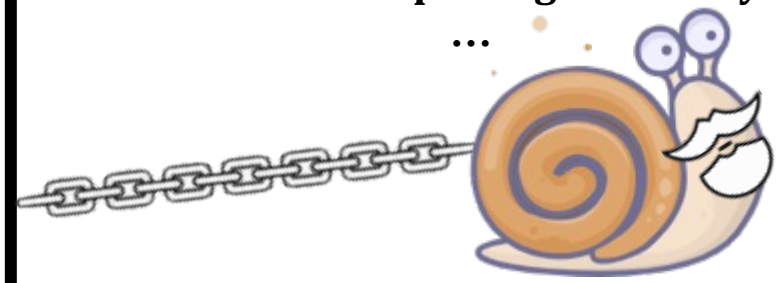
Mengying Zhou, Yang Chen, Shihan Lin, Xin Wang, Bingyang Liu, Aaron Yi Ding
Fudan University, HUAWEI, TU Delft

Nowadays HTTPs has fallen behind

People wants ...

Faster response Stable connection

- **Head-of-Line (HoL) Blocking**
- **Connection Establishment Latency**
- **Stream Multiplexing Efficiency**
- **...**



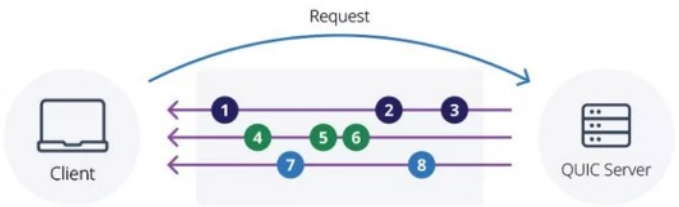
H1.1 or H2

However...

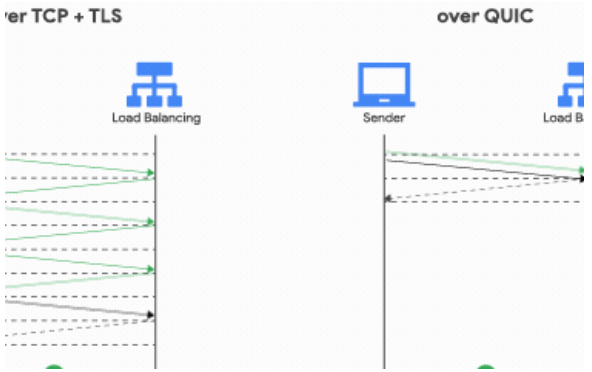
HTTP/1.x or HTTP/2 hardly satisfy these demands

HTTP/3 comes!

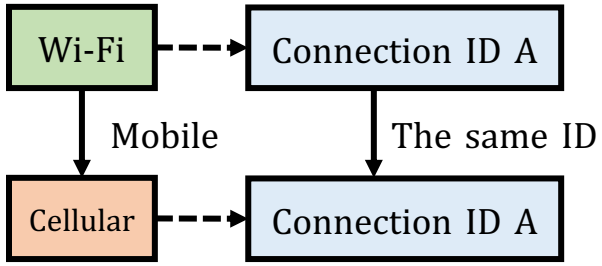
*HTTP/3 was designed to make HTTP traffic more **secure, efficient, and faster**.*



1. Efficient: stream multiplexing



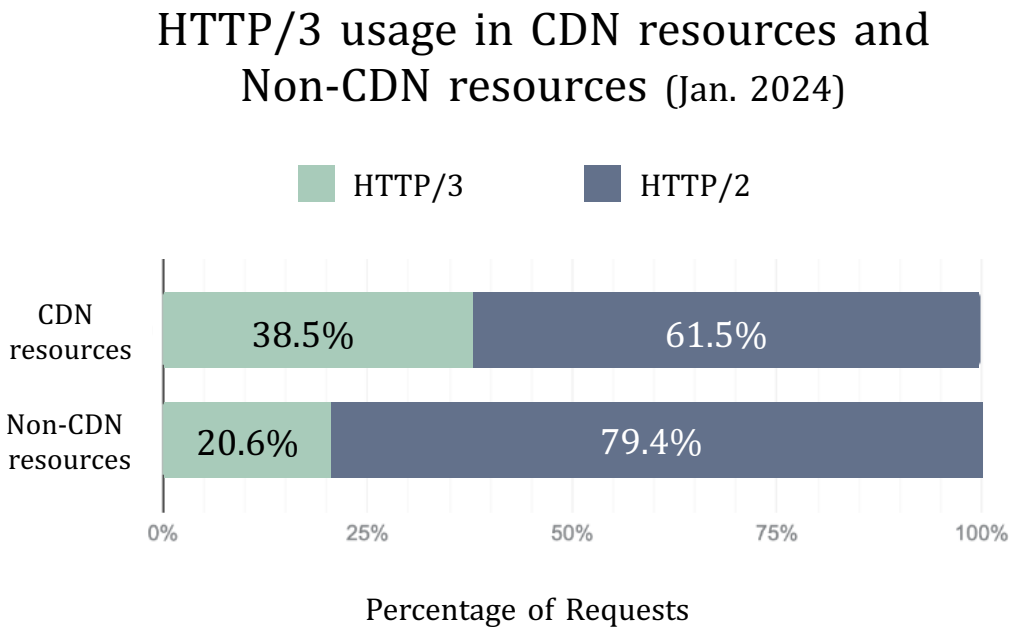
2. Fast: quick connection establishment



3. Flexible: migration without awareness

CDN: a main driver for H3

Provider	Release Year	Performance Report	H3 supported (%)
Cloudflare	2019 [1]	H3 performs 12.4% better in TTFB, but 1-4% worse in PLT than H2.	44.8%
Google Cloud CDN	2021 [2]	Reduce reach latency by 2%, video rebuffer times by 9%, and improves mobile device throughput by 7%.	95.7%
Fastly	2021 [3]	QUIC can represent an 8% in- crease in throughput.	8.1%
QUIC.Cloud	2021 [4]	H3 turns TTFB from 231ms to 24ms.	/
Amazon CloudFront	2022 [5]	N/A	7.7%
Meta	2022 [6]	H3 reduces tail latency by 20% and MTBR by 22%.	/
Akamai	2023 [7]	6.5% enhancement in users with TAT under 25ms; 12.7% improvement for requests ex-ceeding 1 Mbps.	/

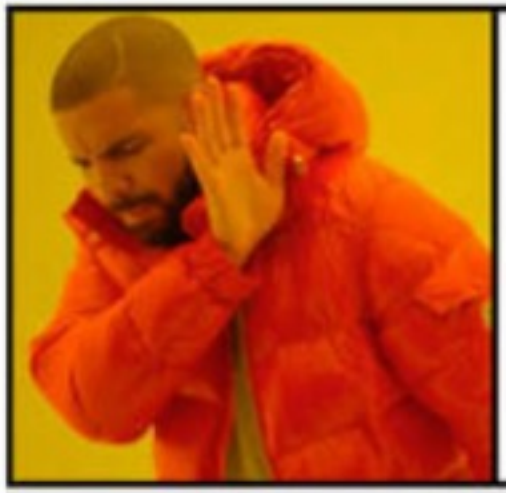


- H3 Adoption in Mainstream CDNs

- Higher usage of H3 in CDN resources

[1] A. Ghedini and R. Lalkaka, "HTTP/3: the past, the present, and the future," Available: https://blog.cloudflare.com/http3-the-past-present-an d-future, 2019
[2] "HTTP/3 gets your content there QUIC, with Cloud CDN and Load Balancing," Available: https://cloud.google.com/blog/products/network ing/cloud-cdn-and-load-balancing-support-http3, 2021
[3] "Making loveholidays 18% faster with HTTP/3," Available: https://tech.loveholidays.com/making-loveholidays-18-faster-with-h ttp-3-1860879528a7, 2021
[4] "QUIC.cloud CDN is Production Ready!" Available: https://www.quic .cloud/quic-cloud-cdn-production-ready, 2021
[5] C. Yun, "HTTP/3 Support for Amazon CloudFront," Available: https:// aws.amazon.com/blogs/aws/new-http-3-support-for-amazon-cloudfront, 2022
[6] T. Ingale, "Watch Meta's engineers discuss QUIC and TCP innovations for our network," Available: https://engineering.fb.com/2022/07/06/ne tworking-traffic/watch-metas-engineers-discuss -quic-and-tcp-innovati ons-for-our-network, 2022
[7] "HTTP/3 is added by default to a new Ion property," Available: https://techdocs.akamai.com/ion/changelog/may-15-2023-supportfor-http3, 2023

Why is H3 doing so well in CDN?



Previous research:

- Analyze CDN and H3 **separately**

Ours:

- Study CDN and H3's synergy **holistically**

Why is H3 doing so well in CDN?

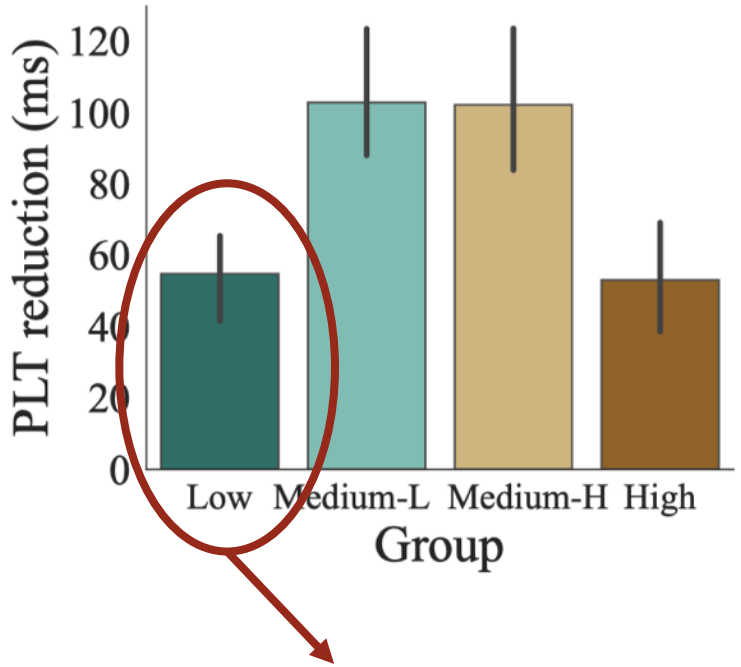
What makes these two compatible?

RQ1: What are the **inherent characteristics** of CDN services on webpages?

RQ2: What is the **synergistic collaboration** between H3’s features and these characteristics of CDN services?

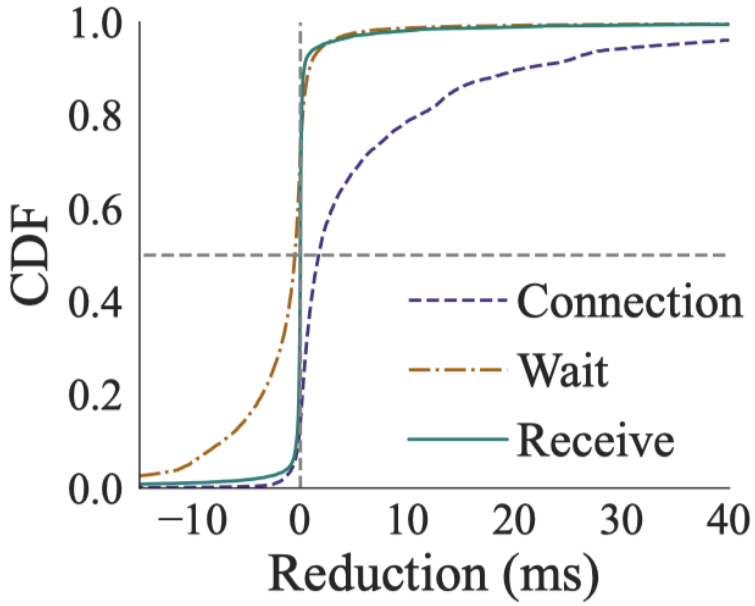
H3 adoption bring improvement

- PLT reduction for websites with different H3 adoption levels



Small-scale adoption

- PLT reduction in three request stages
  - connection, waiting, and receiving

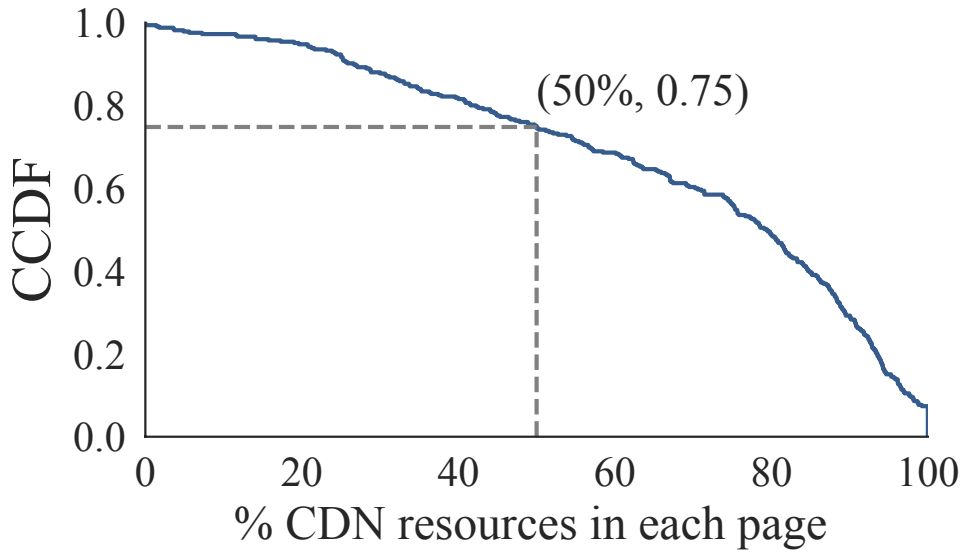


Fast connection contributes the most

Metric definition:

$X_{reduction} = X_{H2} - X_{H3}$ *X* includes *PLT, Connection time, Wait time, and Receive time*

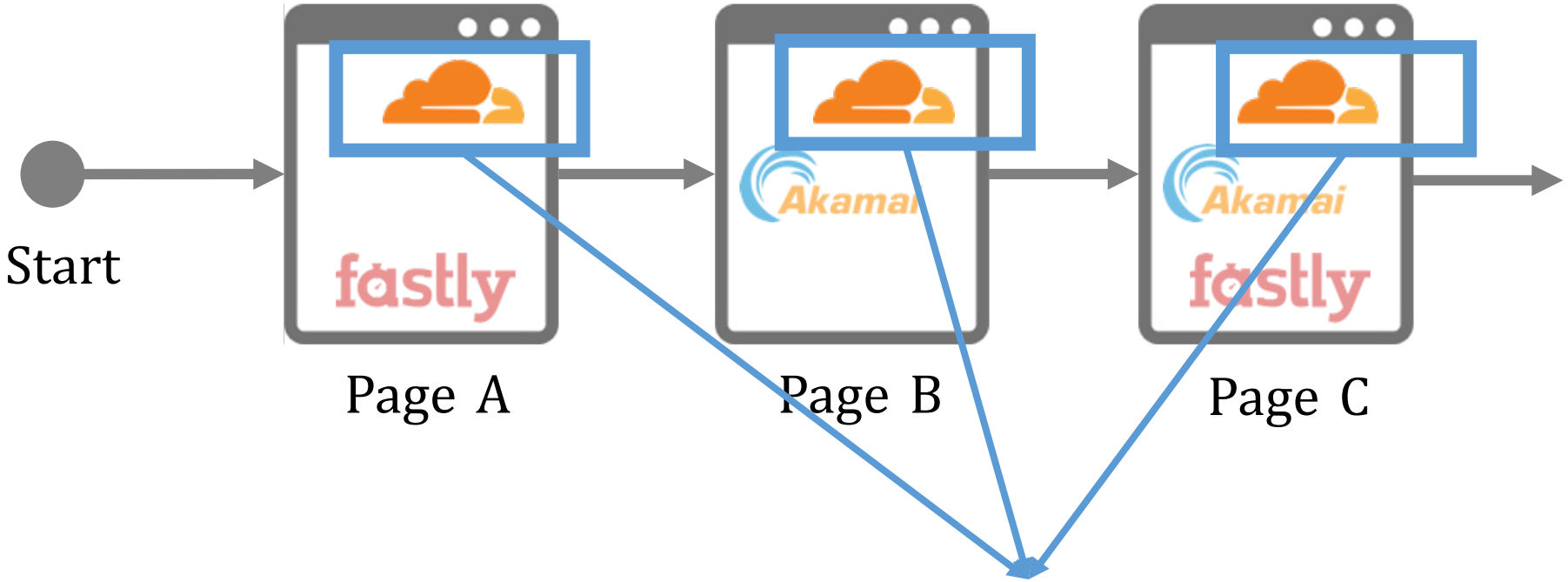
Why significant: dominance of CDN



- 75% of pages' CDN percentage are over 50% , amplifying H3's fast connection benefits.

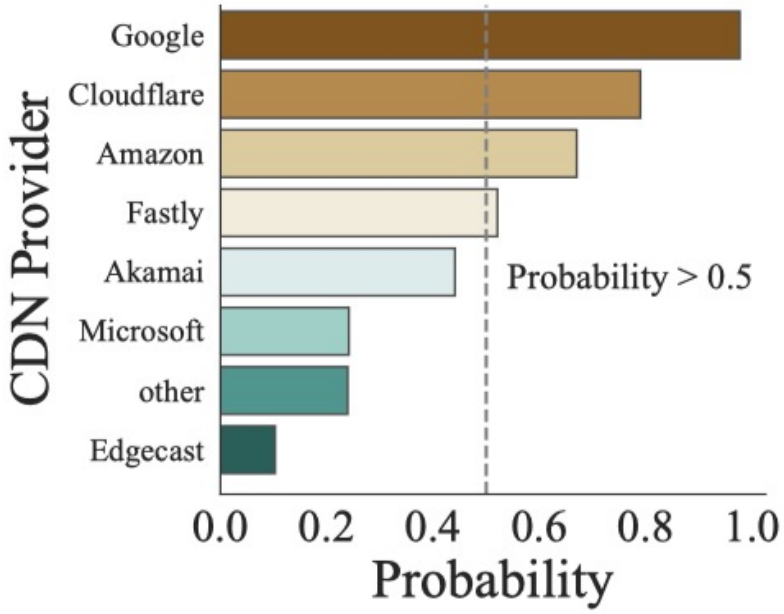
Takeaway 1:
- Fast connection in H3 contributes to accelerating page loading
- Dominant proportion of CDN resources amplifies such acceleration

A phenomenon in consecutive web browsing



Cross page shared-provider phenomenon

Shared-provider phenomenon

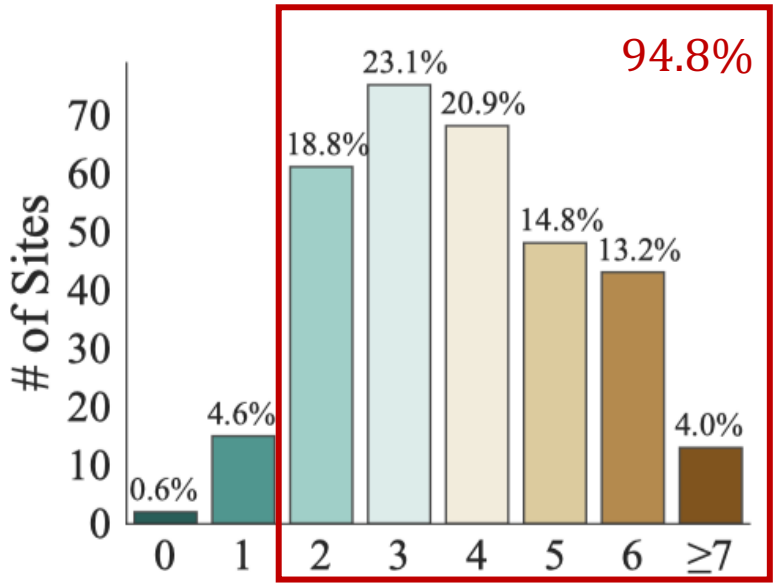


Giant CDN providers appear on almost every page

Shared-provider phenomenon

Giant CDN providers appear on almost every webpage

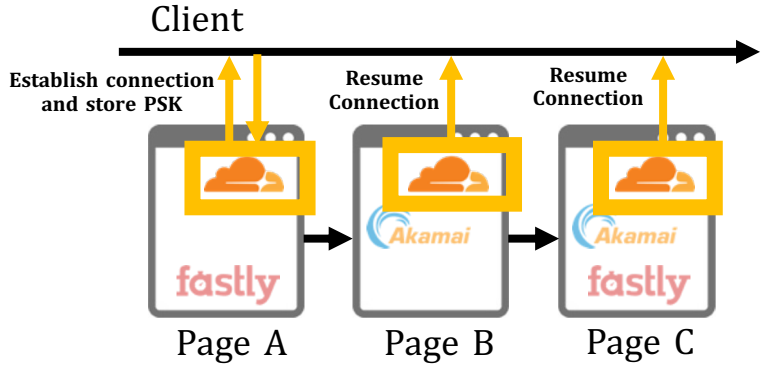
Result in



# of CDN Providers on Each Website

Giant providers are shared across different webpages

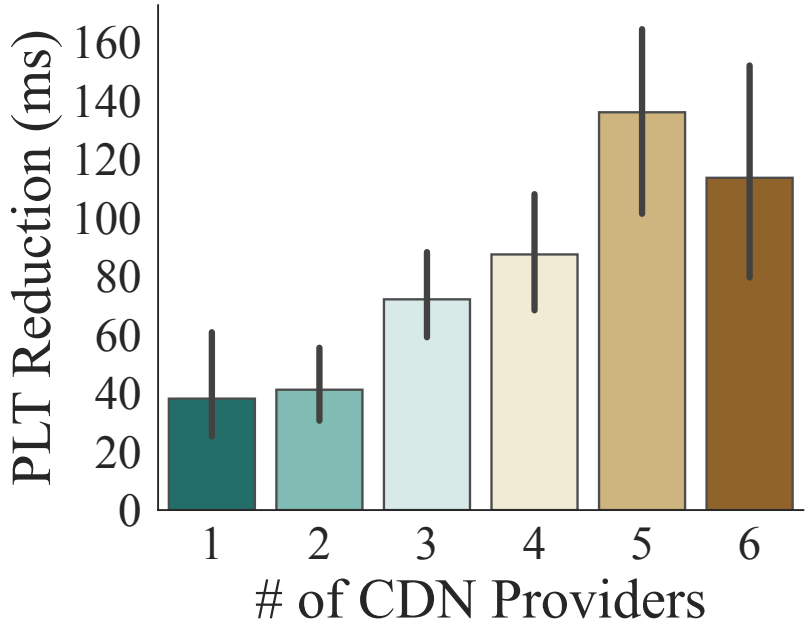
Can H3 connections be resumed* across pages by the same CDN provider?



* H3 connection can be resumed cross pages owing to pre-shared keys (PSK).

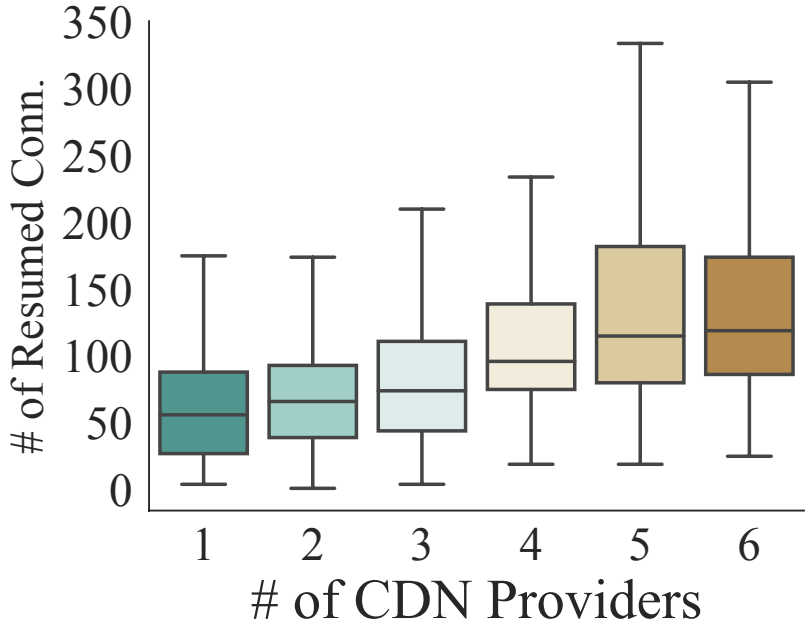
Shared-provider phenomenon reduces PLT with resumed connections*

- PLT reduction for websites with different numbers of used providers



- More shared-providers, more PLT reduction.

- Number of resumed connections for different numbers of providers



- More shared-providers, more resumed connections.

* Under consecutive visits scenario

Case study: Two shared-level groups

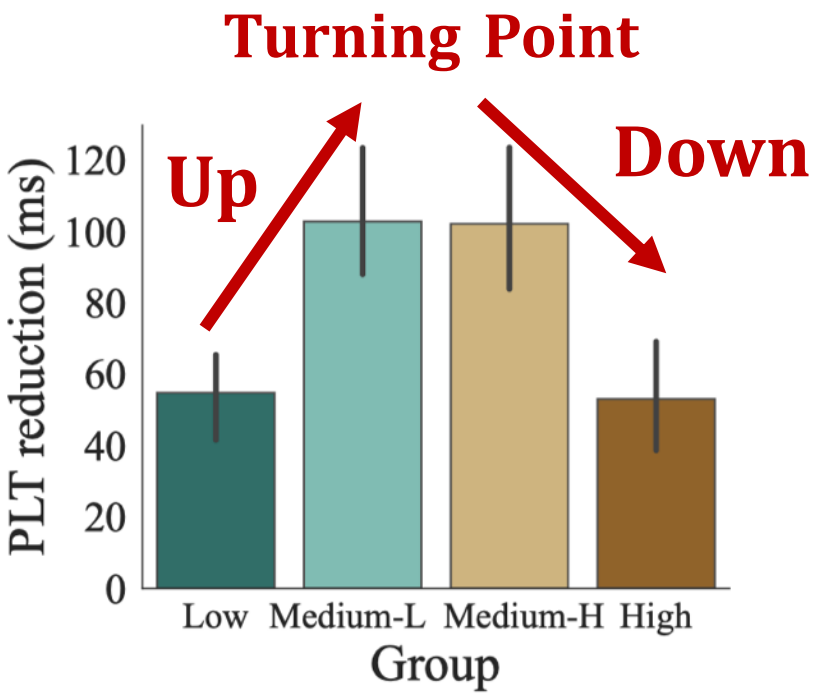
TABLE III: The PLT reduction comparison of two webpage groups with different sharing degrees

Metric	High sharing group C_H	Low sharing group C_L
Avg num. of shared providers	4.16	2.58
Avg num. of resumed connection	101.64	73.74
PLT reduction (ms)	109.3	54.35

- The higher the degree of sharing among these browsed pages, the more significant the optimization becomes.

Takeaway 2:
- There is a phenomenon of giant CDN providers being shared across different pages.
- This phenomenon accelerates page loading by triggering connection resumption of H3.

Full transition of CDN services to H3?

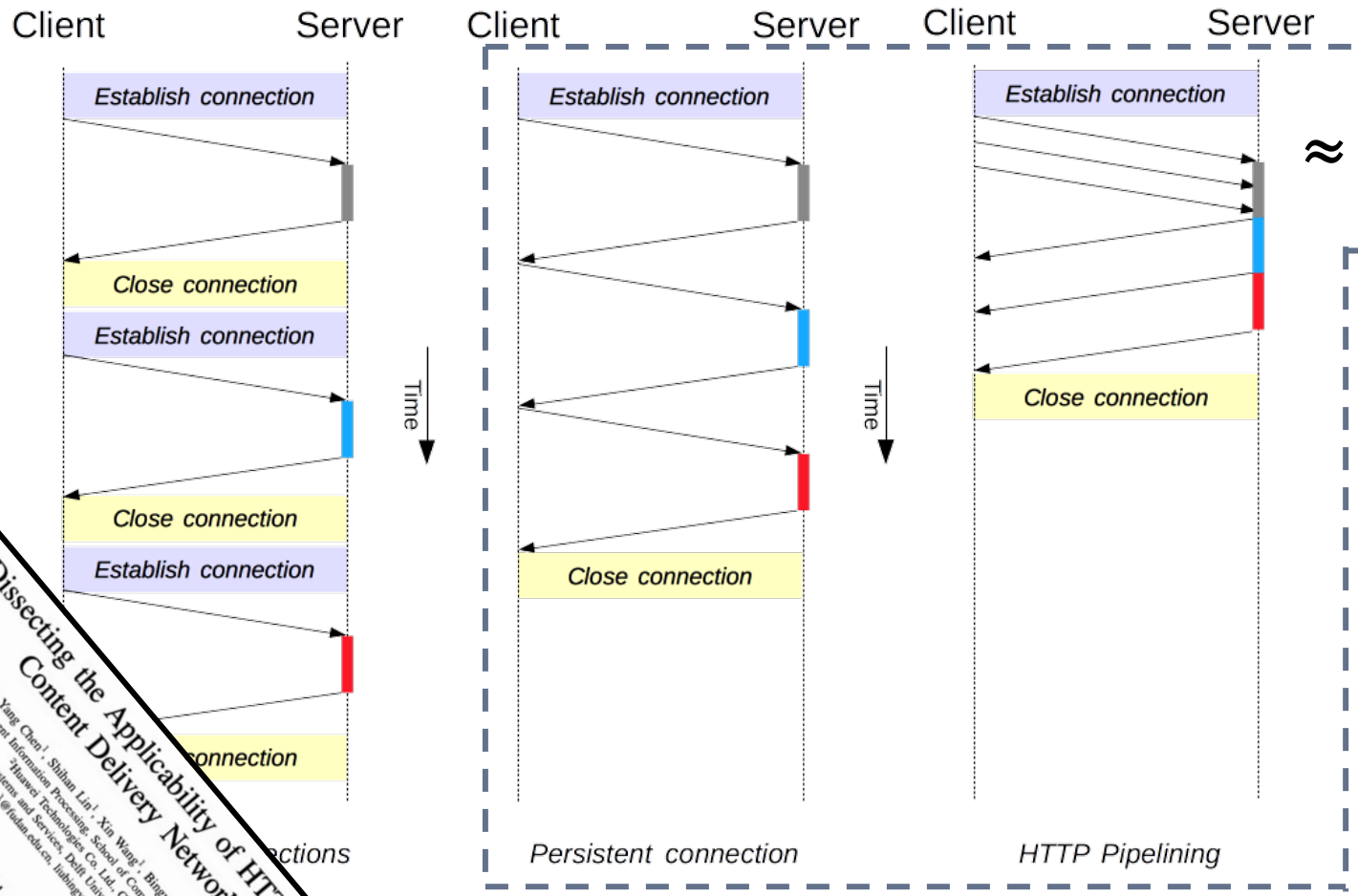


Our lesson:

- Watch out optimization turning points

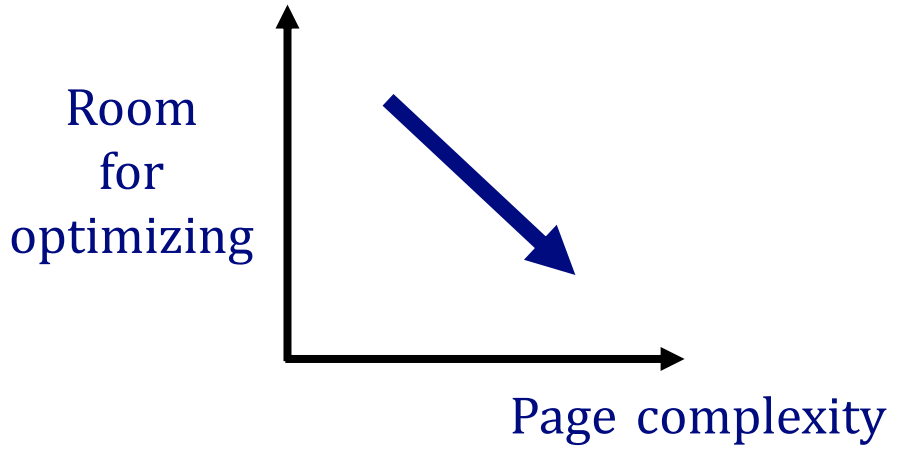
Reused HTTP connections diminish H3 benefits

Let us recall: connection time is the **primary factor** contributing to PLT reduction



Dissecting the Applicability of HTTP/3 in Content Delivery Networks

≈ Similar to H3's fast connection



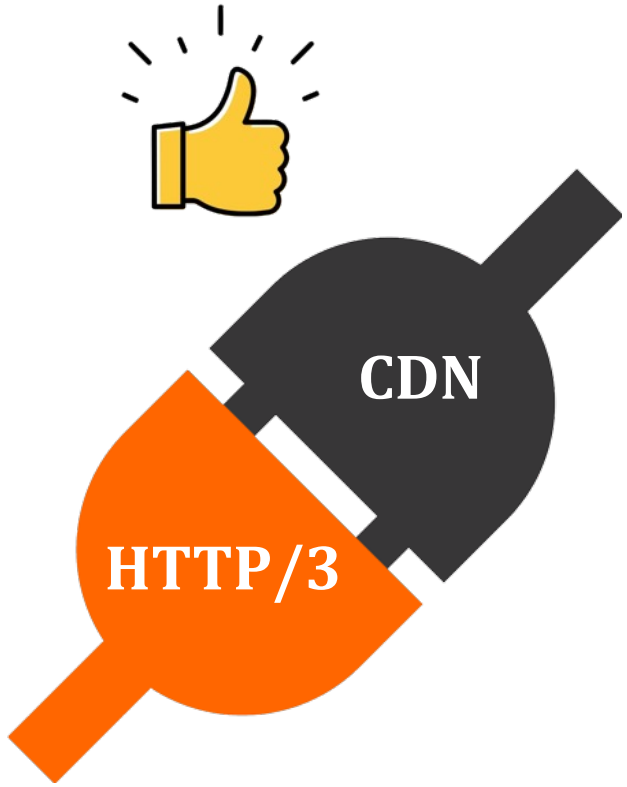
Conclusions

H3 in CDN: great applicability and compatibility

1. Dominant proportion of CDN resources amplifies benefit of H3's fast connection

2. Shared-provider phenomenon accelerates page loading by triggering connection resumption of H3.

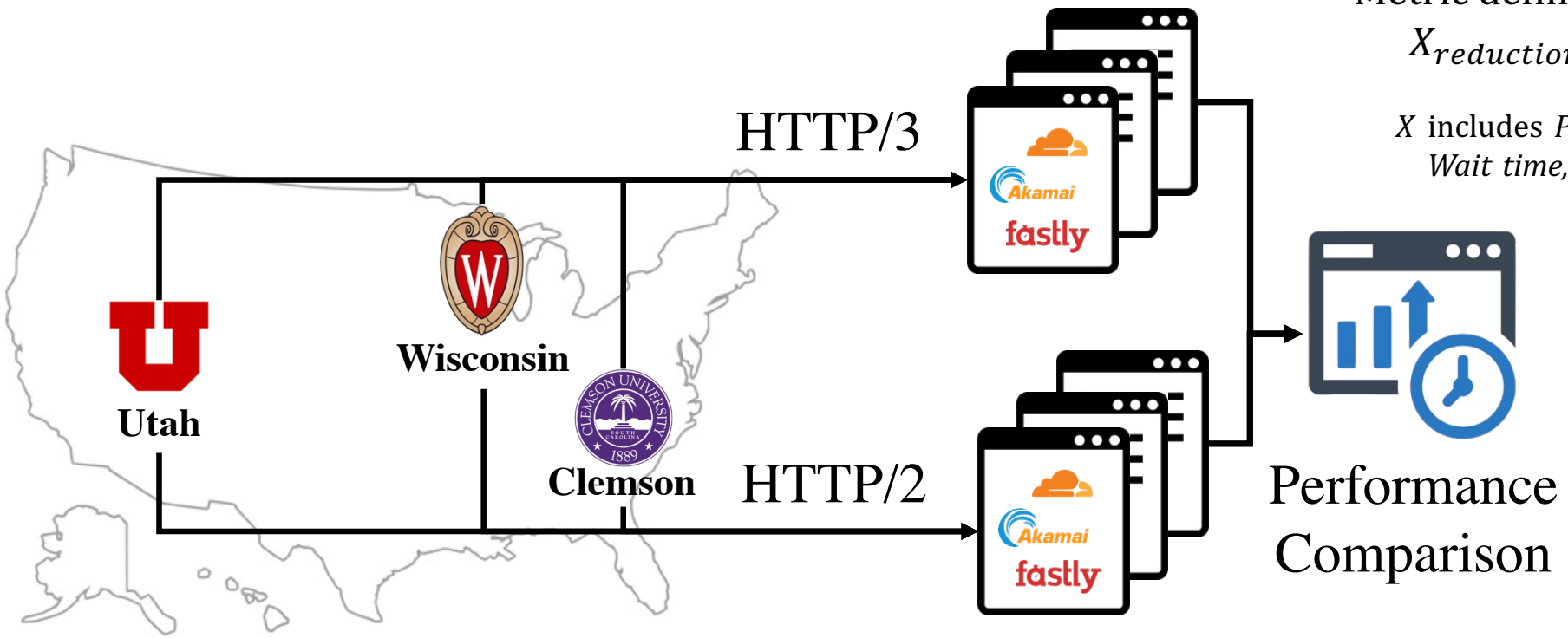
3. Watch out the optimization turning points, rather than adopting H3 blindly.



Thanks for your attention!

Homepage: https://mengyingzhou.github.io

Measurement setup



Metric definition:

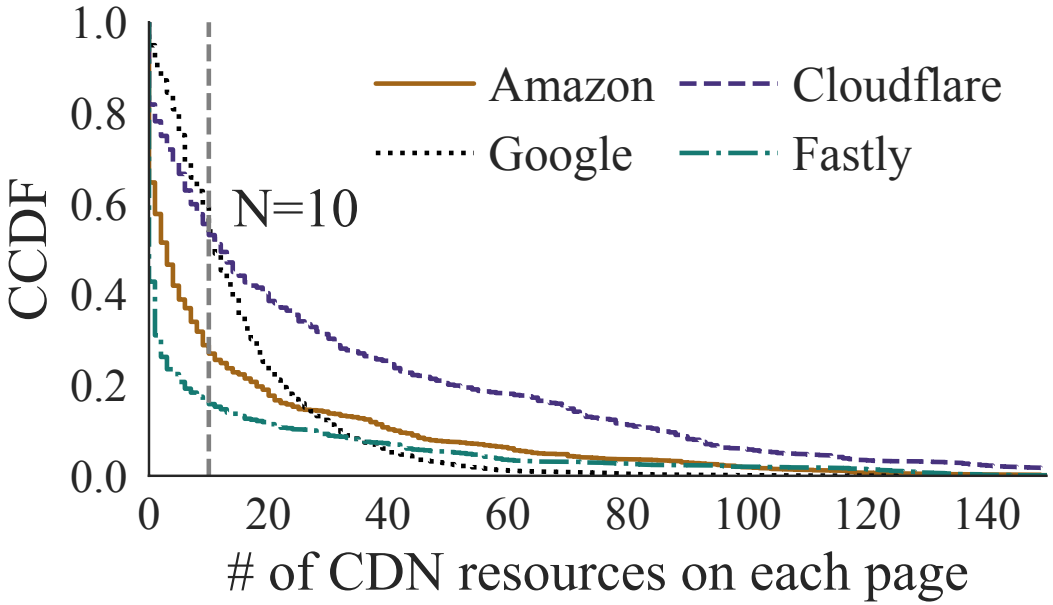
$$X_{reduction} = X_{H2} - X_{H3}$$

X includes *PLT, Connection time, Wait time, and Receive time*

Geo of Probes

Target Webpages (Alexa Top 500)

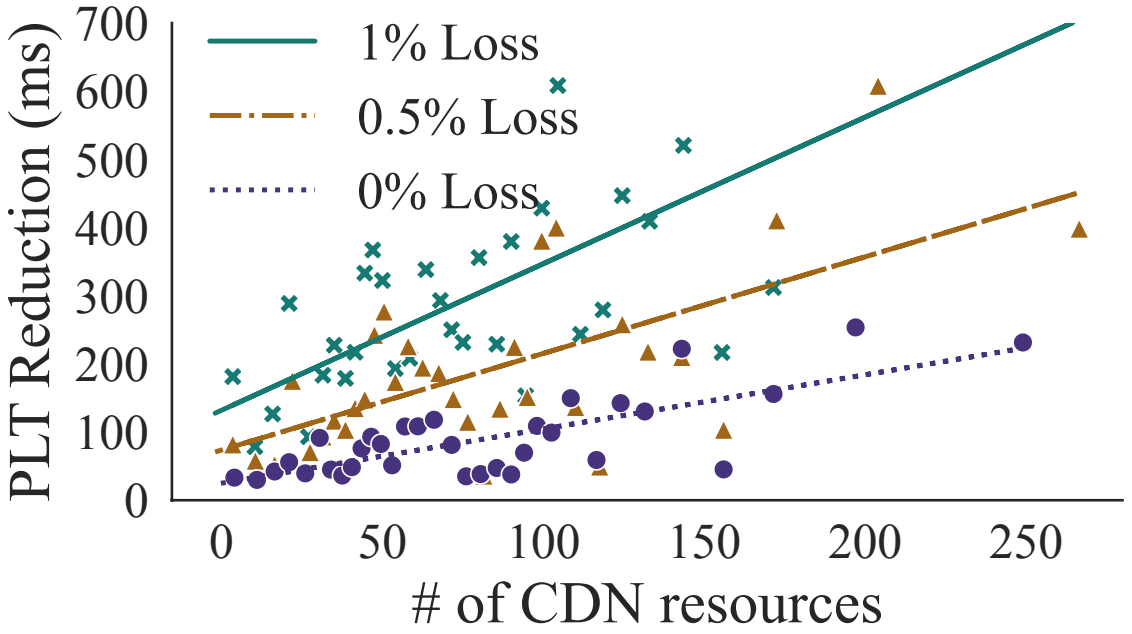
HoL problem with multiple CDN resources



The number of CDN entries per webpage for Amazon, Cloudflare, Google, and Fastly

TCP-based CDN providers are prone to HoL problem

Stream multiplexing eliminates HoL problem



Takeaway 3:

- Multiple CDN resources increase the risk of congestion.

- H3's stream multiplexing mitigates this problem.