

Recitation 12: Akamai

MIT - 6.033

Spring 2022

Henry Corrigan-Gibbs

Plan

- History of Akamai
- Recitation Qs (mixed in!)
- CDNs
 - * What they are
 - * How they work
- Discussion

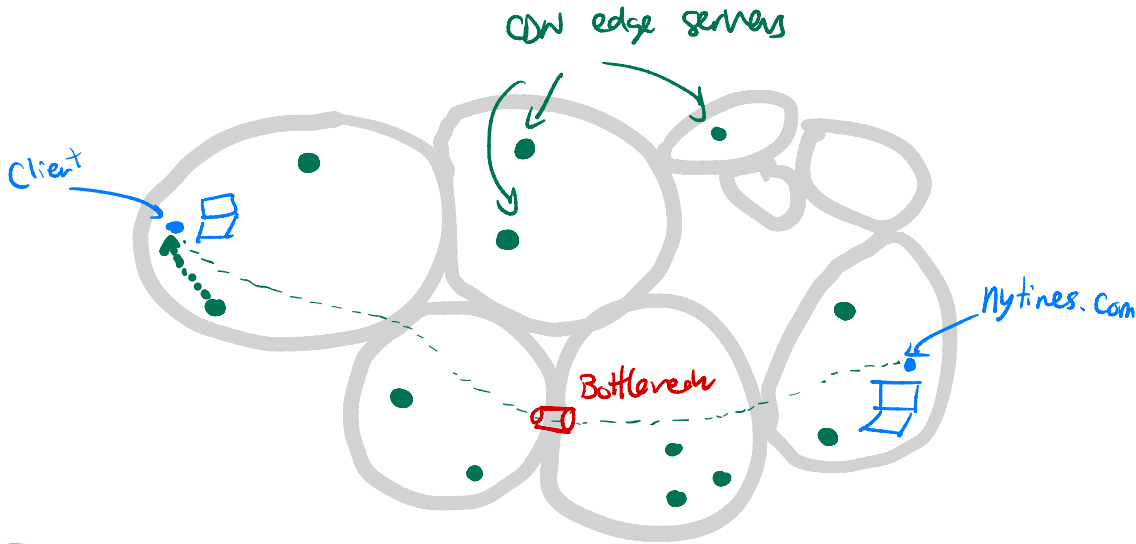
Logistics

- Fan in classroom (x3)
- Exam 3/31 - Open book!
↳ See website.
- Come prepared with Qs next Thursday (11am)

History of Akamai

- Tom Leighton, Danny Lewin, at MIT
 - ↳ Background, key results
- Lost Sloan biz competition (1997)
- Prototype 1998, Public in 1999

What is a CDN?



2. Which aspects of Internet's design is Akamai designed to overcome?

- Latency may be large - physical distance
- Throughput may be poor
- + Low latency
- + Fewer opportunities for bottlenecks

CDNs have many thousands of machines in many hundreds or thousands of physical locations around the world.

↳ Getting around the world is $\sim 40,000$ km ≈ 20 ms at speed of light. Too slow, even with no other slowdown

How does content get to edge servers?

Examples

Akamai (35% of market?)
CloudFlare (5%)
Fastly (13%)
Amazon CloudFront (9%)
Verizon EdgeCast (12%)

2016 #s

Throughput: 200 Tbps (2021), up from 1.5 Tbps

1. Who are Akamai's end users? Customers?

4. Why is it necessary for Akamai to overcome those aspects?

\$\$\$?

↳ Others?

Why use a CDN? (Ask!)

1. Latency / throughput

- Put data closer to client / edge
- Control full ^(IP) path — maximize throughput
- Latency / throughput relationship (BDP)

2. The Reddit Problem

- Traffic spikes (benign)
- DDos (malicious)
- Don't want to provision for worst case bad

3. Separation of concerns

- Lots of work to maintain big network of servers w/ high availability
- e.g. NY Times — newspaper, not a tech co

→ Does it really protect against outages?

Demo

- Latency measurement
 - ↳ Ping time to `cs.stanford.edu` vs. `www.Stanford.edu`
- Log into CloudFlare dashboard ?
 - ↳ Don't do this
- `CDNperf.com`
 - ↳ Cost \$
- Look at `nytimes.com`, notice that assets come from CDN often.
 - ↳ Inspect HTTP headers

How a page load works (simplified)

1. DNS lookup - RTT to DNS servers (or more)
nytimes.com → 1.2.3.4
2. TCP connection setup (RTT)
3. TLS connection (RTT)
4. Data transfer (RTT)

For each resource on page potentially.

With CDN...

1. Site DNS points to CDN
 - ↳ Use IP anycast to reach closest one (explain)
 - ↳ DNS points to a CDN server IP
2. Client makes request to CDN server IP (anycast?)
3. CDN server returns cached content (if it has it) or makes upstream req to origin server
 - ↳ Need to be careful about what can be cached
4. Origin server returns answer
5. CDN edge caches & returns to client

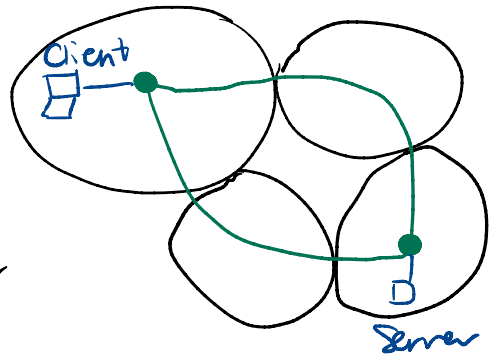
Challenge: What to cache & where?

Other features

Overlay network

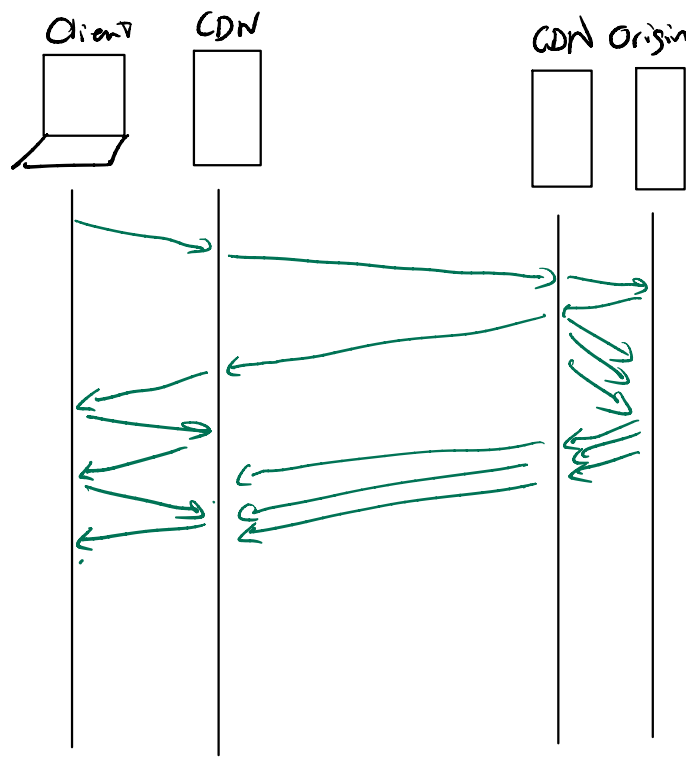
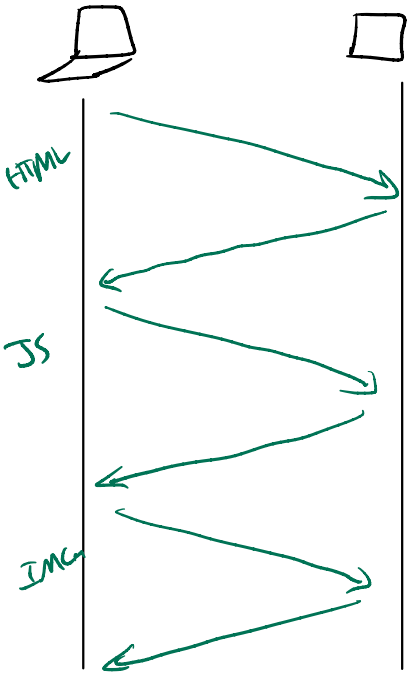
- ↳ Control both ends
- ↳ Use non-style ideas
- ↳ Add extra info in
- ↳ Resistent conns
 - ↳ Handle packet loss

Special-purpose tool is better



Prefetching

Without prefetch



Potential pitfalls (Discuss in groups)

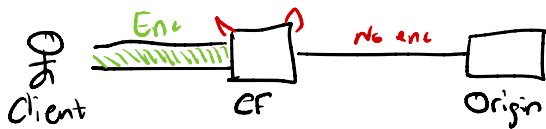
- Single point of failure

- ↳ Examples of Akamai going down?
- ↳ Config problems, etc.

July 2021, DNS bug: triggered by config change. took out FedEx, Airbnb, fidelity, VPS, Amazon, ...

- Not compatible (?) w/ E2E encryption

- ↳ Prefetching & compression rely on seeing traffic



- Maybe no enc on backed but green packet appears
- CDN can MITM traffic!

- Fast-changing content

- ↳ At odds w/ caching?

- ↳ With one CDN long ago, had to change file name to update image (e.g. NYT example)

To Discuss

You are a CDN. You get to place 1,000 servers. Where do you put them? Why?