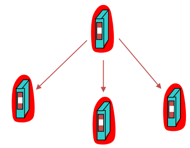


Distributed Systems for Content Delivery

Mike Freedman

Lecture 21

COS 418: Distributed Systems

1

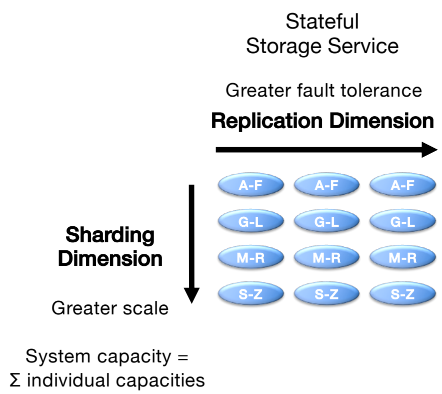
Problem Space

- Many clients accessing web content
- Approach #1: Scale-out web architectures
  - Use many independent instances of stateless web servers
  - Scale-out storage backends via sharding
- Approach #2: Replicate and cache data closer to users
  - Much web content is immutable and/or can be slightly stale

2

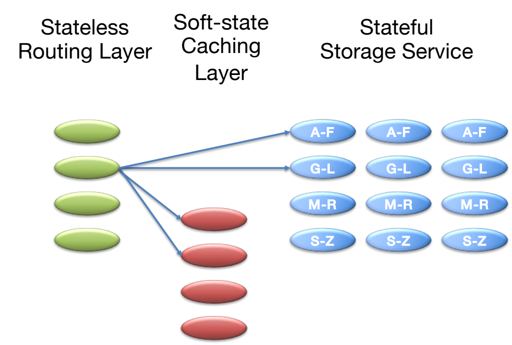
2

Modern Web Architectures



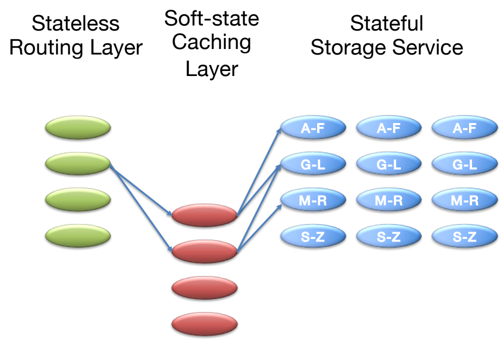
3

Modern Web Architectures



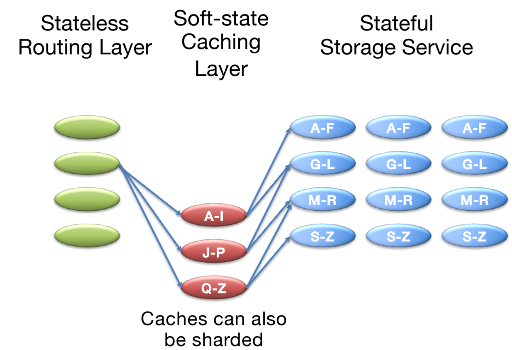
4

Modern Web Architectures



5

Modern Web Architectures



6

Types of State

- Soft state – State (information/data) which is used for efficiency, but is not essential for proper operation
  - Soft state can often be regenerated or replaced if needed
  - E.g., data caching is example of soft state used for performance improvement: If lost, cached data can be refetched from slower, more durable storage
- Hard State – State which is necessary for correctness
  - To date, most of our discussions in class have focused on hard state

Term introduced by David D. Clark (one of "designers" of Internet): *"The design philosophy of the DARPA internet protocols."* SIGCOMM, 1988. http://ccr.sigcomm.org/archive/1995/jan95/ccr-9501-clark.pdf

7

7

Sharded vs. Non-Sharded Caching

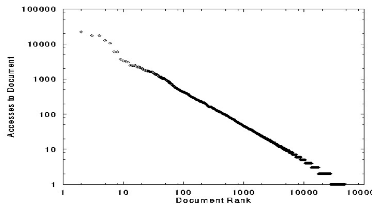
- Pros for sharding
  - Greater cache capacity (Σ individual capacities)
  - Adding servers increases both cache capacity and query throughput (although non-sharded can also scale query throughput)
- Cons for sharding
  - Clients need to maintain semi-accurate cache mappings, rather than just random / round-robin selection
  - Elasticity (adding/removing nodes) more complex, either requiring active moving content or cache misses during passive rebalancing

8

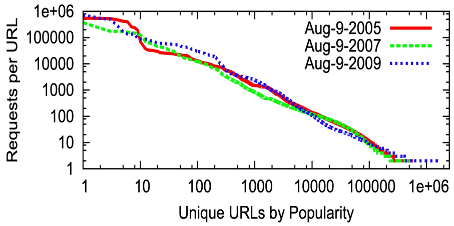
8

How much to cache?

Many Internet workloads have Power law (Zipf) distribution



Characteristics of WWW Client-based Traces. Cunha, Bestavros, Crovella, BU-CS, 1995

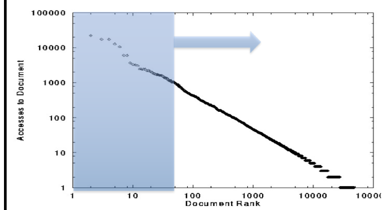


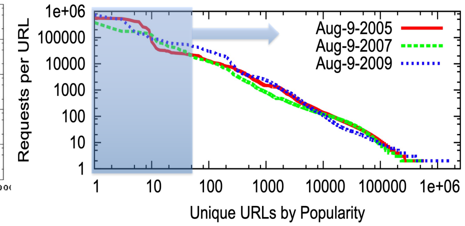
Experiences with CoralCDN: A Five-Year Operational View. Freedman. NSDI 2010

9

How much to cache?

Many Internet workloads have Power law (Zipf) distribution

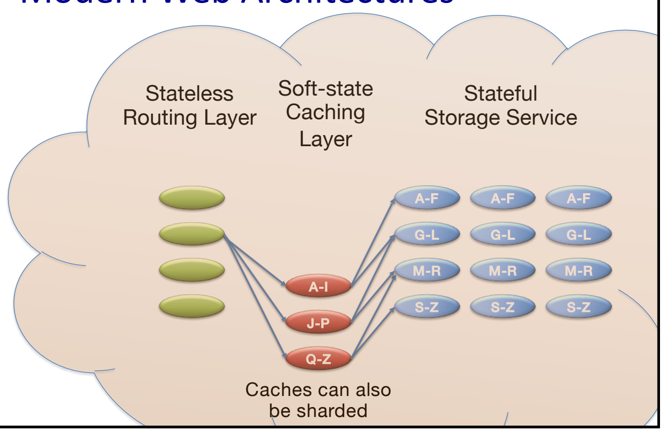




Significant benefits at beginning, but then reduced benefit for cache hit rate as cache size grows

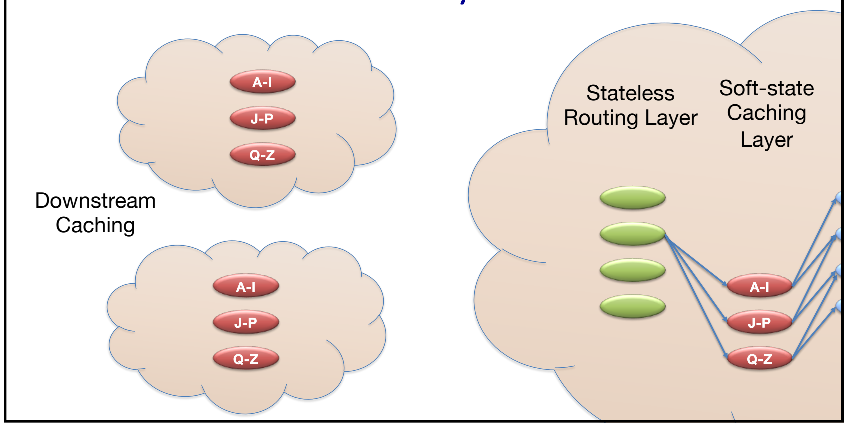
10

Modern Web Architectures



11

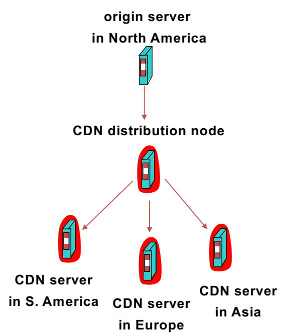
Content Delivery Networks



12

Content Distribution Network

- Proactive content replication
  - Content provider (e.g., CNN) contracts with a CDN
- CDN replicates the content
  - On many servers spread throughout the Internet
- Updating the replicas
  - Reactive by TTL or updates pushed to replicas when the content changes



13

Caching is Complicated

- Significant fraction (>50%?) of distinct HTTP objects may be uncacheable
  - Dynamic data: Stock prices, scores, web cams
  - CGI scripts: results based on passed parameters
  - Cookies: results may be based on passed data
  - Advertising / analytics: want to measure # hits (hint: use random strings)
- Yet significant fraction of HTTP bytes are cacheable
  - Images, video, CSS pages, etc.
- Goal: Maximize cachability, while limiting staleness of cached objects

14

U.S. INTERNATIONAL CANADA ESPAÑOL 中文 PLAY THE CROSSWORD Account

Saturday, April 9, 2022
Today's Paper

The New York Times

78°F 83° 72°
Dow +0.4%

World U.S. Politics N.Y. Business Opinion Tech Science Health Sports Arts Books Style Food Travel Magazine T Magazine Real Estate Video

LIVE Coronavirus Pandemic 30s ago Russia-Ukraine War 23m ago

Dozens Dead After Strike on Train Station in Ukraine

LIVE
At Least 50 Killed and Nearly 100 Wounded

- Ukrainian officials blamed Russia for a missile attack that targeted civilians who were trying to flee ahead of an expected stepped-up offensive in Ukraine's east.
- Western nations, which have been tightening sanctions on Russia, saw the attack as new justification to intensify their efforts. Follow updates.

Russian Blunders in Chernobyl: 'They Came and Did Whatever They Wanted'
After Russian generals took over the Chernobyl Exclusion Zone, one of the most toxic places on earth, scenes of disregard for safety were everywhere.

Kharkiv, Ukraine
Tyler Hicks/The New York Times

Land Mines on a Timer, Scattered Over a Ukrainian Town
Civilians near Kharkiv in northeastern Ukraine are discovering a new threat: munitions that eject up to two dozen small mines that explode at intervals.

"For the children": A chilling message was found at the Ukrainian train station destroyed in a missile strike.

2 Men Acquitted of Plotting to Kidnap Gov. Gretchen Whitmer of Michigan
Jurors acquitted two defendants and couldn't agree on a verdict for the other two in a setback to U.S. prosecutions of domestic terror.

Opinion
CHARLES M. BLOW
Cory Booker, 'You Had Our Back'

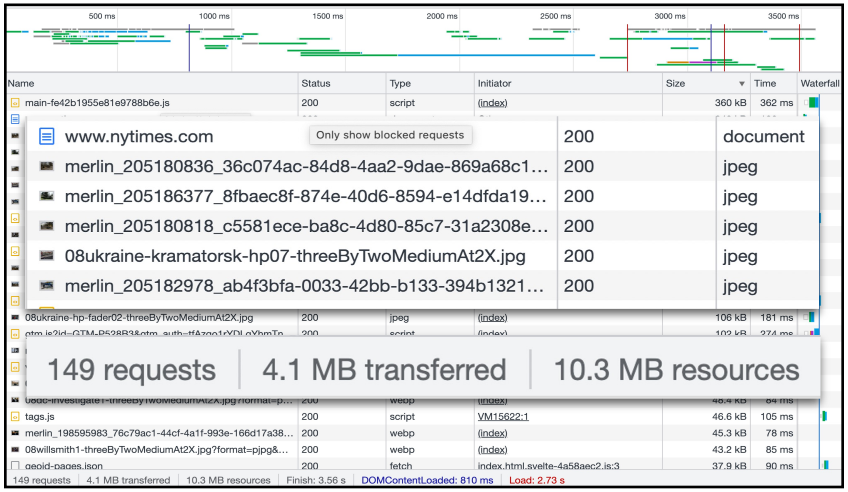
DAVID BROOKS

15

Name	Status	Type	Initiator	Size	Time	Waterfall
main-fe42b1955e81e9788b6e.js	200	script	(index)	360 kB	362 ms	
www.nytimes.com	200	document	Other	343 kB	106 ms	
merlin_205180836_36c074ac-84d8-4aa2-9dae-869a68c1…	200	jpeg	(index)	301 kB	321 ms	
merlin_205186377_8fbaec8f-874e-40d6-8594-e14dfda19…	200	jpeg	(index)	259 kB	244 ms	
merlin_205180818_c5581ece-ba8c-4d80-85c7-31a2308e…	200	jpeg	(index)	230 kB	282 ms	
08ukraine-kramatorsk-hp07-threeByTwoMediumAt2X.jpg	200	jpeg	(index)	207 kB	223 ms	
merlin_205182978_ab4f3bfa-0033-42bb-b133-394b1321…	200	jpeg	(index)	190 kB	210 ms	
grumi.js	200	script	grumi-ip.js:1	187 kB	367 ms	
merlin_205185192_5d3658e2-0849-4392-aa7d-f9758a6d…	200	jpeg	(index)	167 kB	247 ms	
home-f8887cada20e9963da7f.js	200	script	(index)	153 kB	339 ms	
merlin_205205850_c0ce44bd-b4d0-4cb9-8050-2f137d8e…	200	jpeg	(index)	144 kB	173 ms	
08ukraine-hp-fader01-threeByTwoMediumAt2X.jpg	200	jpeg	(index)	144 kB	176 ms	
pubads_impl_2022040501.js	200	script	grumi-ip.js:1	128 kB	186 ms	
08ukraine-hp-fader02-threeByTwoMediumAt2X.jpg	200	jpeg	(index)	106 kB	181 ms	
gtm.js?id=GTM-P528B3&gtm_auth=tfAzqo1rYDLgYhmTn…	200	script	(index)	102 kB	274 ms	
merlin_205192041_5f20a742-ad9b-467b-985e-919856a5…	200	jpeg	(index)	80.3 kB	226 ms	
vendor-f6dbc528114fb1fda428.js	200	script	(index)	74.9 kB	305 ms	
08king1-threeByTwoMediumAt2X.jpg?format=pjpg&qualit…	200	webp	Other	51.5 kB	66 ms	
08dc-investigate1-threeByTwoMediumAt2X.jpg?format=p…	200	webp	(index)	48.4 kB	84 ms	
tags.js	200	script	VM15622:1	46.6 kB	105 ms	
merlin_198595983_76c79ac1-44cf-4a1f-993e-166d17a38…	200	webp	(index)	45.3 kB	78 ms	
08willsmith1-threeByTwoMediumAt2X.jpg?format=pjpg&…	200	webp	(index)	43.2 kB	85 ms	
geoid-pages.json	200	fetch	index.html.svelte-4a58aec2.js:3	37.9 kB	90 ms	

149 requests | 4.1 MB transferred | 10.3 MB resources | Finish: 3.56 s | DOMContentLoaded: 810 ms | Load: 2.73 s

16



17

Caching is powerful: Modern HTTP Video-on-Demand

- Download "content manifest" from origin server
- List of video segments belonging to video
  - Each segment 1-2 seconds in length
  - Client can know time offset associated with each
  - Standard naming for video resolutions/formats: eg, 320dpi, 720dpi, 1040dpi
- Client downloads video segment (at certain resolution) using standard HTTP request.
  - HTTP request can be satisfied by cache: it's a static object
- Client observes download time vs. segment duration, increases/decreases resolution if appropriate

18

18

CDN Case Study: How Akamai Works

Akamai Network

- Servers: ~365,000
- Networks: 1,350
- Countries: 135

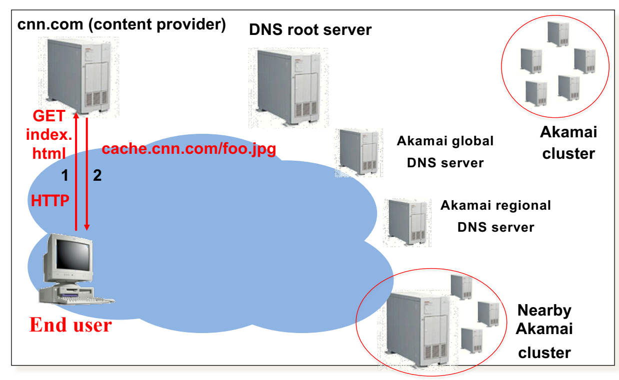
https://www.akamai.com/company/facts-figures

19

19

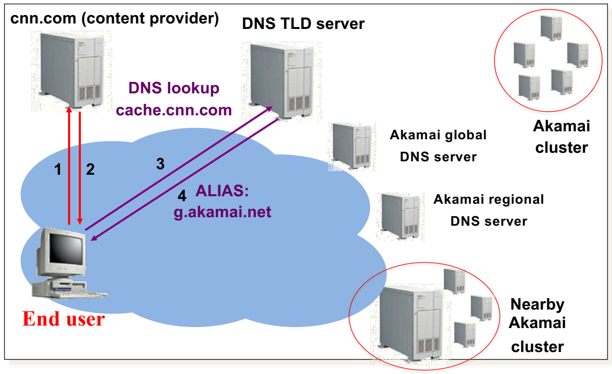
20

How Akamai Uses DNS



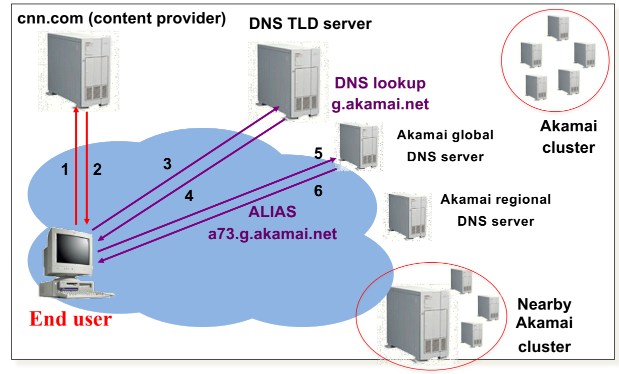
20

How Akamai Uses DNS



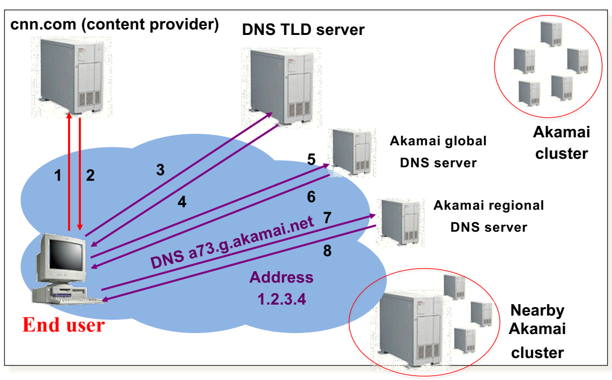
21

How Akamai Uses DNS



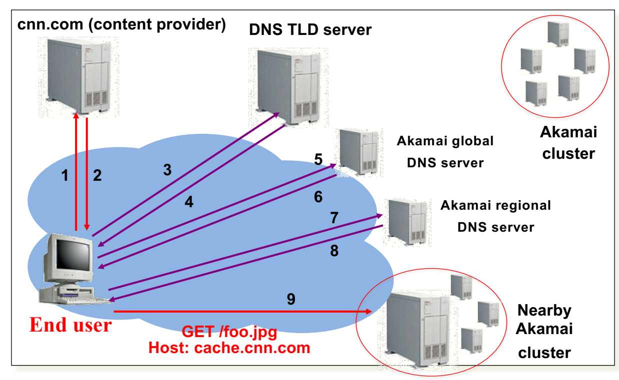
22

How Akamai Uses DNS



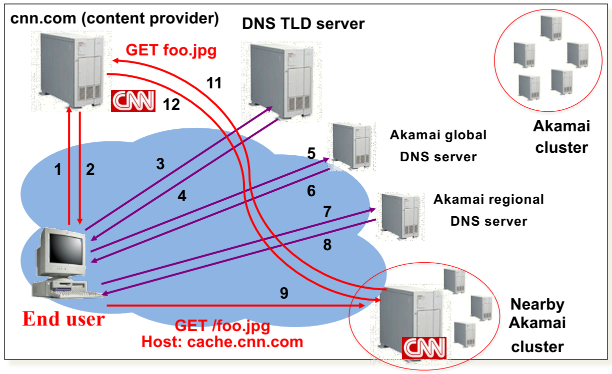
23

How Akamai Uses DNS



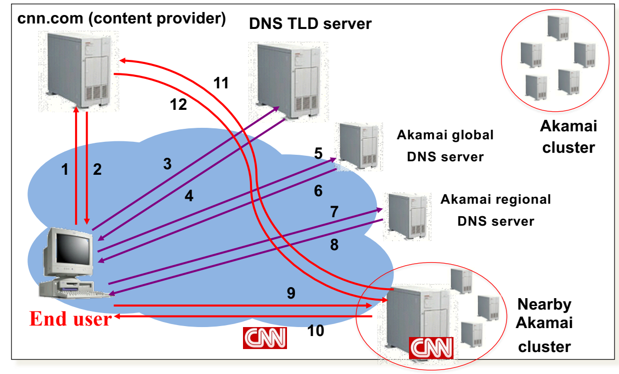
24

How Akamai Uses DNS



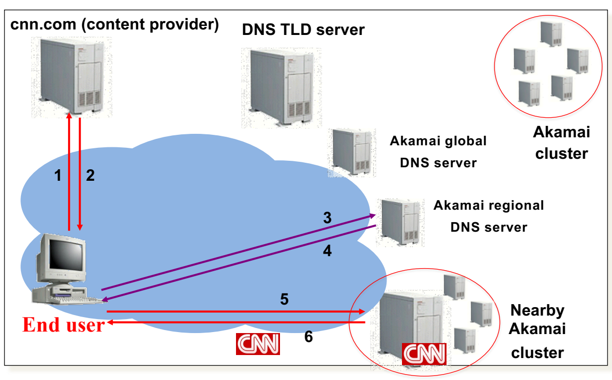
25

How Akamai Uses DNS



26

How Akamai Works: Cache Hit



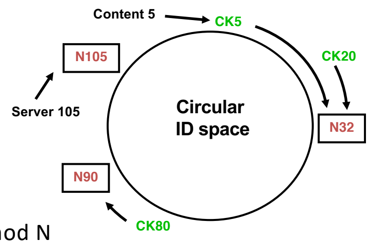
27

Routing Client Requests within Map

- Mapping system collects data about each “group” of IP addresses, based on network latency, loss, connectivity
- Map each IP group to a preferred server cluster
  - Updated roughly every minute
    - **Short, 60-sec DNS TTLs** in Akamai regional DNS accomplish this
- Map client request to a server in the cluster
  - Load balancer selects a specific server
  - E.g., to **maximize** the cache hit rate

28

Selecting server inside cluster



- Consistent hashing
  - content_key = hash(URL) mod N
  - node_key = hash(server ID) mod N
  - Content belongs to server's node_key is "closest" to URL's content_key

29

29

"Consistency"?

(and/or limiting the staleness of cached objects)

30

30

How long should the client cache for?

- Clients (and proxies) cache documents
  - When should origin be checked for changes?
  - Every time? Every session? Date?
- HTTP includes caching information in headers
  - HTTP 0.9/1.0 used: "Expires: <date>"; "Pragma: no-cache"
  - HTTP/1.1 has "Cache-Control"
    - "No-Cache", "Max-age: <seconds>"
    - "ETag: <opaque value>

31

31

Why the changes between 1.0 and 1.1?

- Timestamps
  - Server hints when an object "Expires" (Expires: xxx)
  - Server provides last modified date, client can check if still valid
- Problems
  - Client and server might not have synchronized clocks
  - Server replicas might not have synchronized clocks
  - Max-age solves this: relative seconds, not absolute time

32

32

What if cache expires?

- Store past expiry time (if room in cache)
- Upon request, first revalidate with server

GET / HTTP/1.1
Accept-Language: en-us
If-Modified-Since: Mon, 29 Jan 2001 17:54:18 GMT
Host: www.example.com
Connection: Keep-Alive

HTTP/1.1 304 Not Modified
Date: Tue, 27 Mar 2001 03:50:51 GMT
Connection: Keep-Alive

33

33

Conditional GETs

- Revalidate cache content if still valid
- Redownload new version if modified

GET / HTTP/1.1
Accept-Language: en-us
If-Modified-Since: Mon, 29 Jan 2001 17:54:18 GMT
Host: www.example.com
Connection: Keep-Alive

HTTP/1.1 304 Not Modified
Date: Tue, 27 Mar 2001 03:50:51 GMT
Connection: Keep-Alive

34

34

Another clock sync problem!

- What if server replicas don't have aligned modification times?

HTTP/1.1 200
Date: Tue, 27 Mar 2001 03:50:51 GMT
ETag: 686897696a7c876b7e

GET / HTTP/1.1
Accept-Language: en-us
If-None-Match: "686897696a7c876b7e"
Host: www.example.com
Connection: Keep-Alive

35

35

Conclusion

- Content distribution is hard
  - Many, diverse, changing objects
  - Clients distributed all over the world
- Moving content towards client is key
  - Reduces latency, improves throughput, reliability
  - CDNs evolved into complex distributed systems
- Cache controls and revalidation are a key part of managing content freshness with decentralized caching

36

36