

The Web (part 3)

Lecture 8

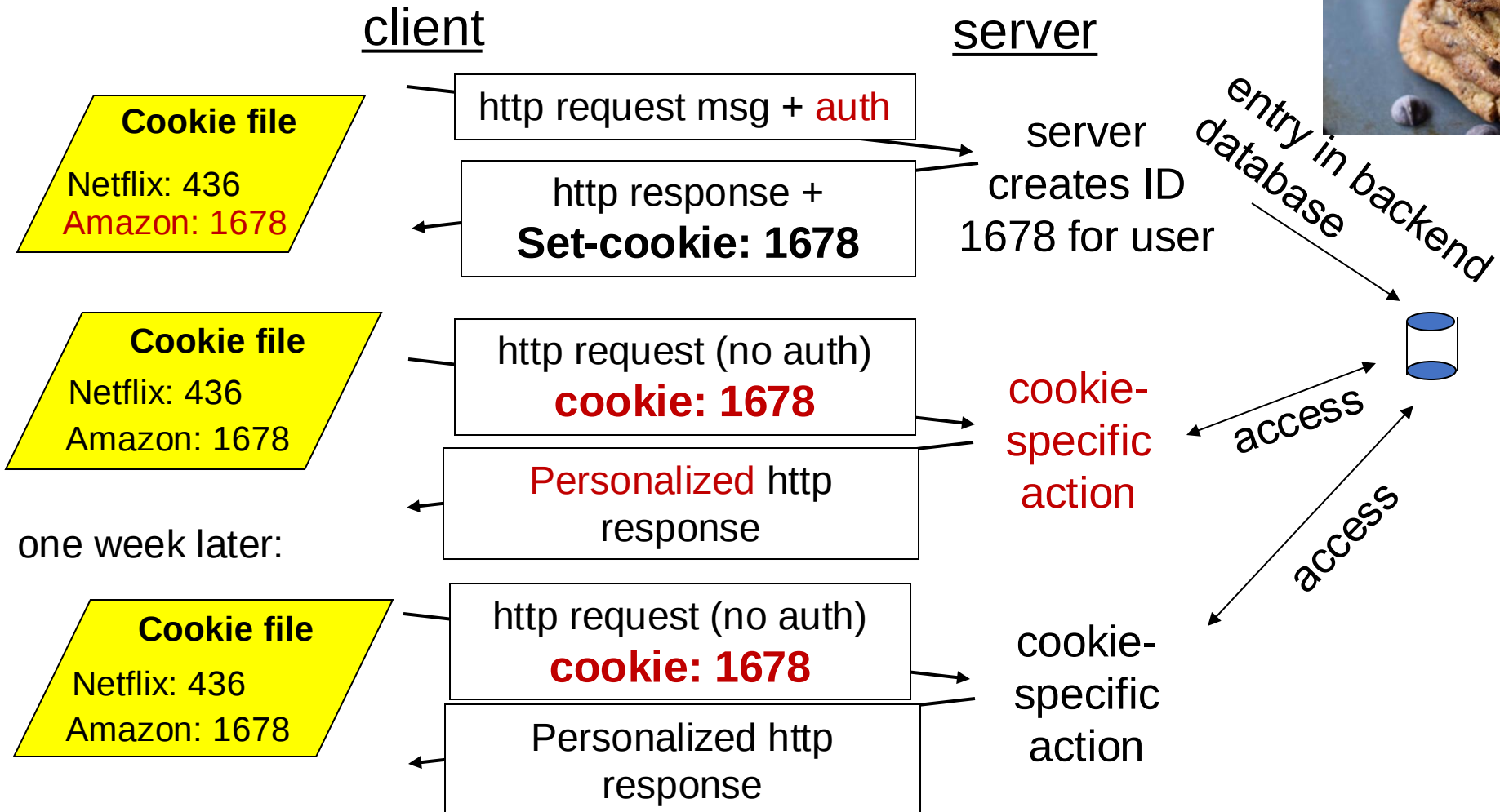
<http://www.cs.rutgers.edu/~sn624/352-F24>

Srinivas Narayana

Cookies: Keeping user memory



Cookie is typically opaque to client.



How cookies work

Collaboration between client and server to track user state.

Four components:

1. cookie header line of HTTP response message
2. cookie header line in HTTP request message
3. cookie file kept on user endpoint, managed by user's browser
4. back-end database maps cookie to user data at Web endpoint

Cookies come with an expiration date (yet another HTTP header)

Cookies have many uses

- The good: Awesome user-facing functionality
 - Shopping carts, auth, ... very challenging or impossible without it
- The bad: Unnecessary recording of your activities on the site
 - First-party cookies: performance statistics, user engagement, ...
- The ugly: Tracking your activities across the Internet
 - Third-party cookies (played by ad and tracking networks) to track your activities across the Internet
 - personally identifiable information (PII)
 - Ad networks target users with ads; may sell this info
 - Scammers can target you too

PSA: Cookies and Privacy

- Disable and delete unnecessary cookies by default
- Suggested privacy-conscious browsers, websites, tools:
 - DuckDuckGo (search)
 - Brave (browser)
 - AdBlock Plus (extension)
 - ToR (distract targeting)
 - ... assuming it doesn't break the functions of the site



<https://gdpr.eu/cookies/>

Web Caching

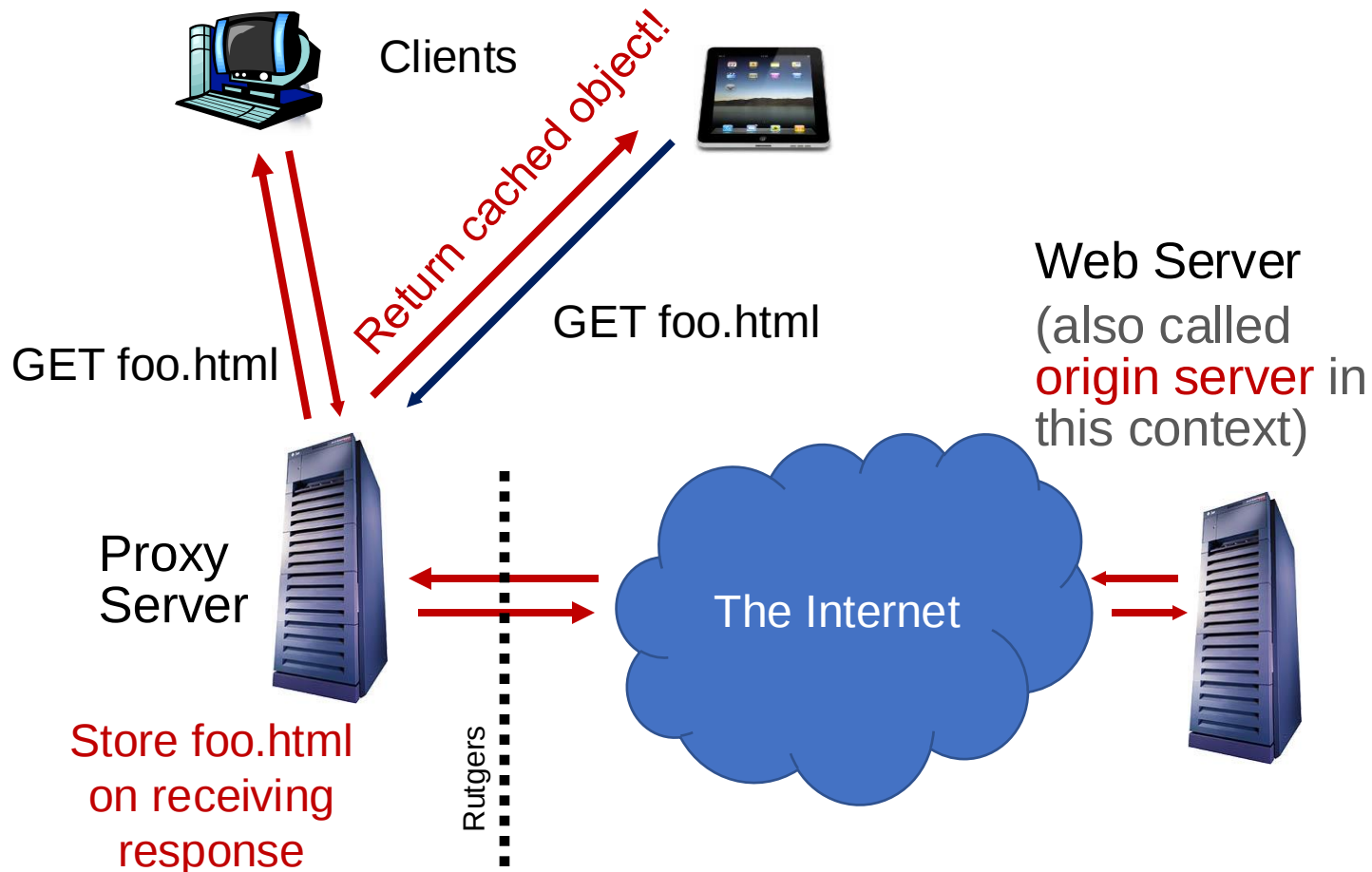
Web caches

Web caches: Machines that remember web responses for a network

Why cache web responses?

- Reduce response time for client requests
- Reduce traffic on an organization's access link

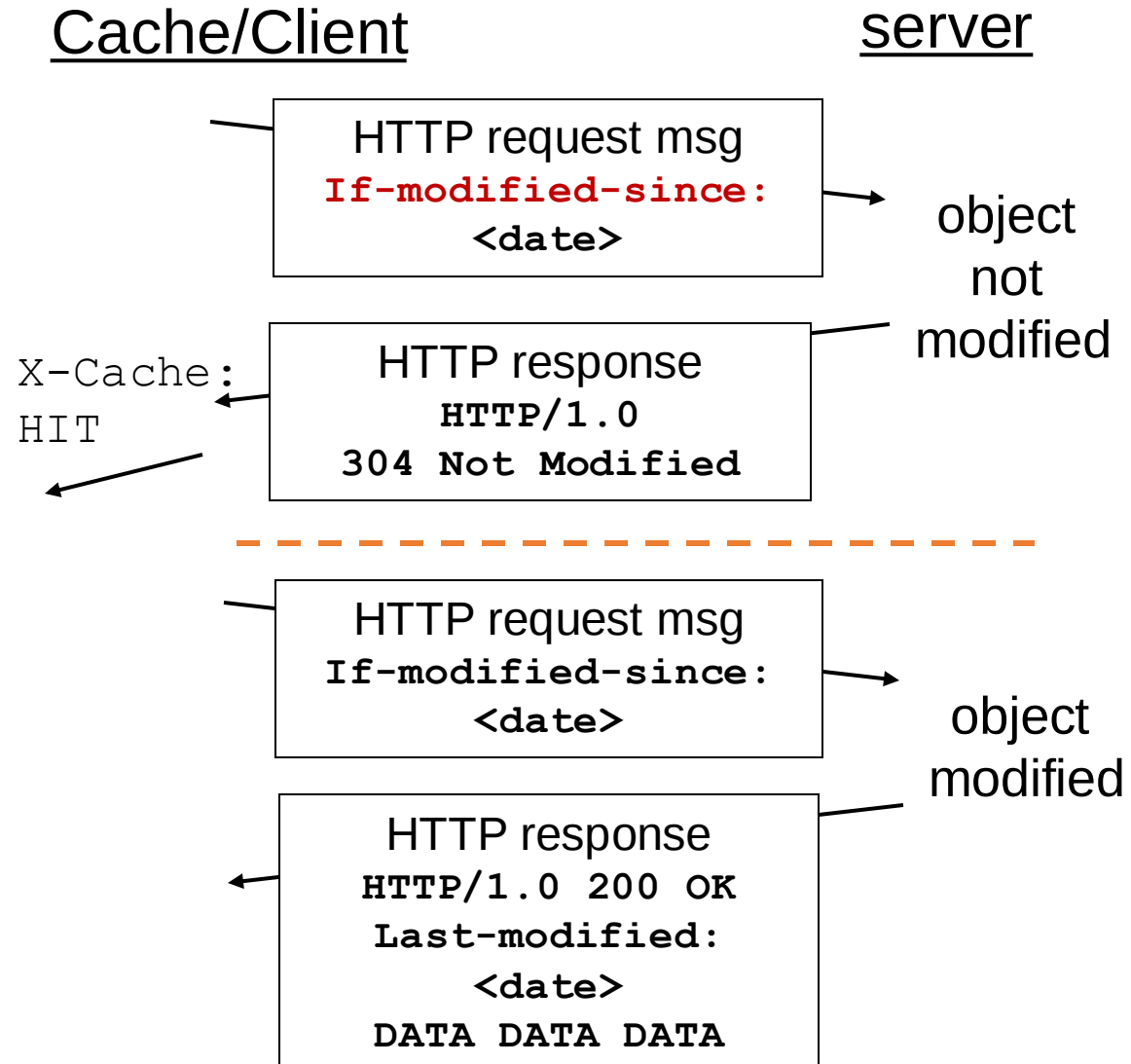
Web caching using a proxy server



- You can configure a HTTP proxy on your laptop's network settings.
- If you do, your browser sends all HTTP requests to the proxy (cache).
- Hit: cache returns object
- Miss: obtain object from originating web server (**origin server**) and return to client
 - Also cache the object locally

Caching in the HTTP protocol

- **Conditional GET**
guarantees cache content is up-to-date while still saves traffic and response time whenever possible
- Date in the cache's request is the last time the server provided in its response header **Last-Modified**



Content Distribution Networks (CDNs)

A global network of web caches

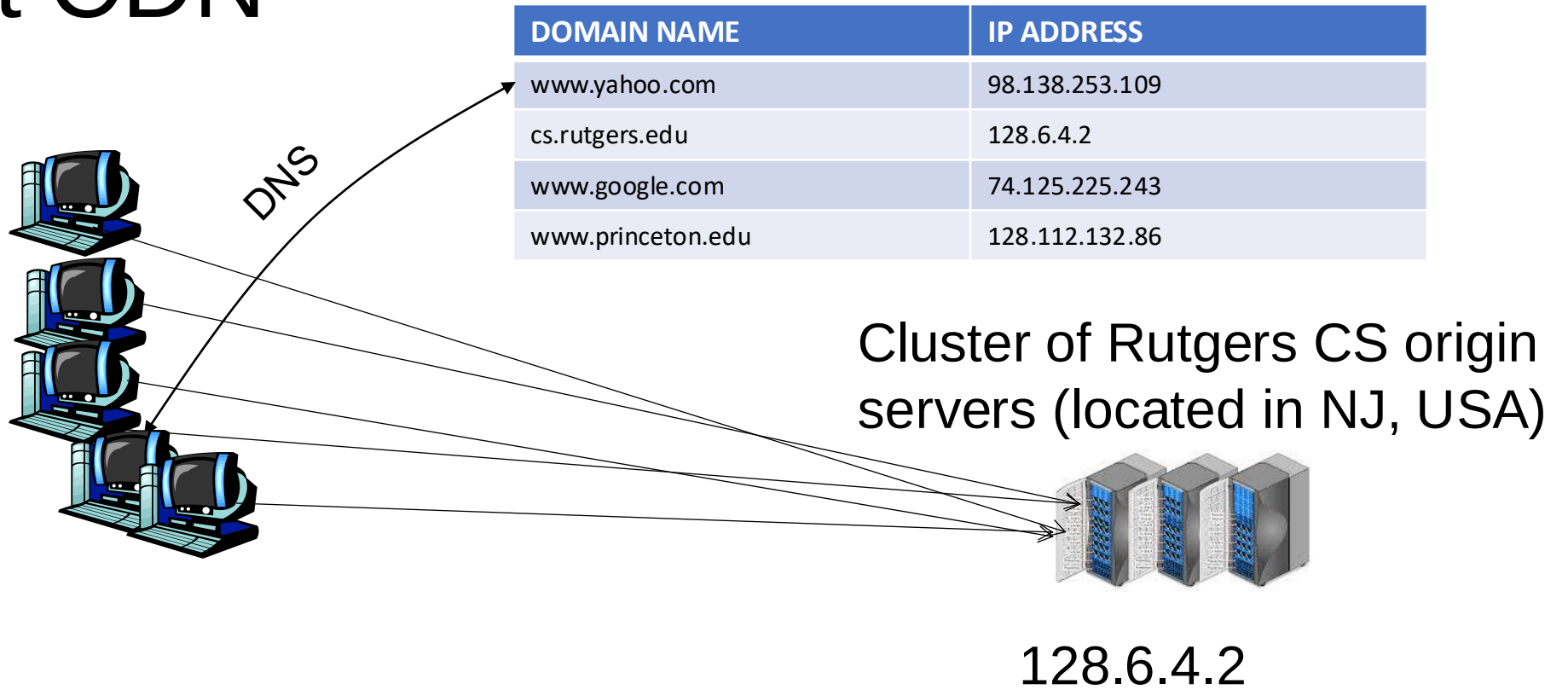
- Provisioned by ISPs and network operators
- Or content providers, like Netflix, Google, etc.

Uses (overlaps with uses of web caching in general)

- Reduce traffic on a network's Internet connection, e.g., Rutgers
- Improve response time for users: CDN nodes are closer to users than origin servers (servers holding original content)
- Reduce bandwidth requirements on the content provider
- Reduce cost to maintain origin servers

Without CDN

Clients
distributed
all over the
world



- Problems:
- Huge bandwidth requirements for Rutgers
- Large propagation delays to reach users

Where the CDN comes in

- Distribute content of the origin server over geographically distributed **CDN servers**
- But how will users get to these CDN servers?
- **Use DNS!**
 - DNS provides an additional layer of indirection
 - Instead of returning an IP address, return another DNS server (NS record)
 - Much like a response to any other iterative query
 - The second DNS server (run by the CDN) returns the IP address of the client
- The CDN runs its own DNS servers (**CDN name servers**)
 - Custom logic to send users to the “closest” CDN web server

With CDN

NS record delegates the choice of IP address to the CDN name server.

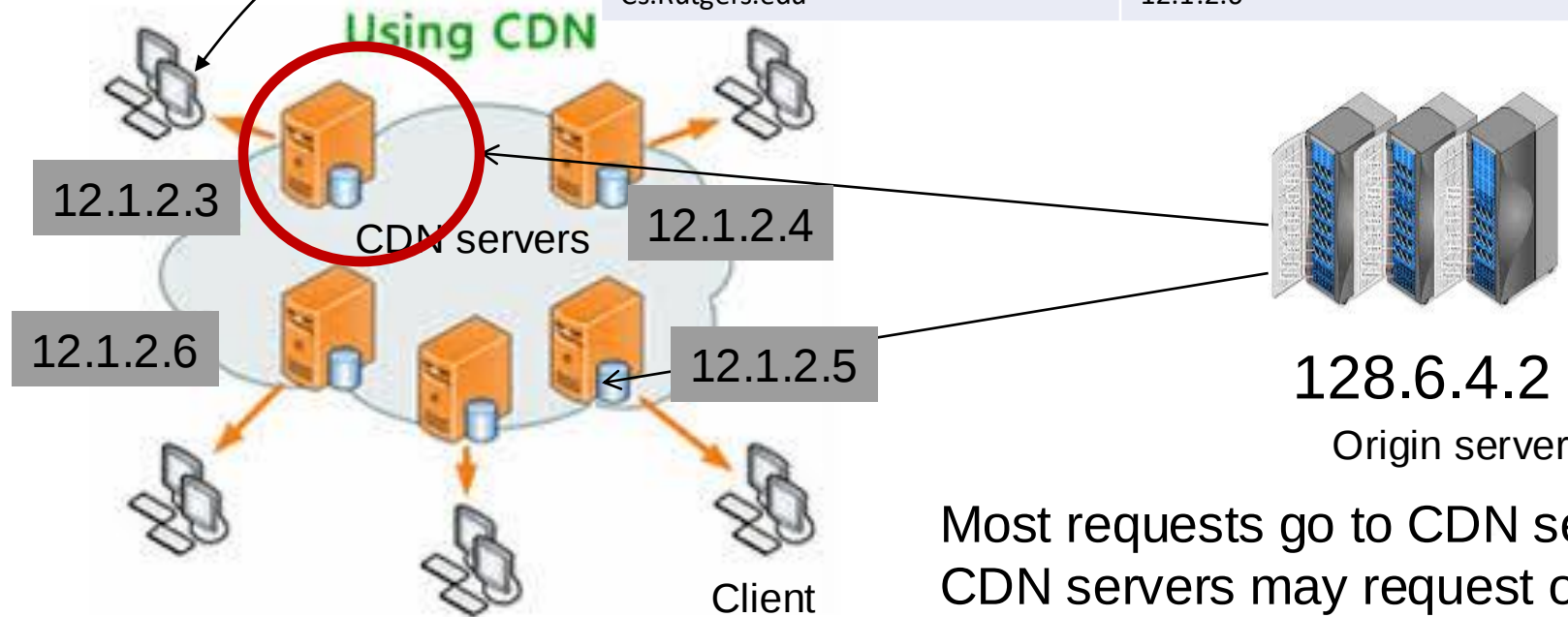
DOMAIN NAME	IP ADDRESS
www.yahoo.com	98.138.253.109
cs.rutgers.edu	124.8.9.8 (NS record pointing to CDN name server)
www.google.com	74.125.225.243

CDN Name Server (124.8.9.8)

DOMAIN NAME	IP ADDRESS
Cs.Rutgers.edu	12.1.2.3
Cs.Rutgers.edu	12.1.2.4
Cs.Rutgers.edu	12.1.2.5
Cs.Rutgers.edu	12.1.2.6

Custom logic to map ONE domain name to one of many IP addresses!

Popular CDNs:
CloudFlare
Akamai
Level3
...



Most requests go to CDN servers (caches).
CDN servers may request object from origin
Few client requests go directly to origin server

Seeing a CDN in action

- `dig +trace freshtohome.com`
- `dig web.mit.edu (or) dig +trace web.mit.edu`

Summary of HTTP

- Request/response protocol
- ASCII-based human-readable message structures
- Enhanced stateful functionality using cookies
- Improve performance using caching and CDN
- Persistence and pipelining to improve performance
- Simple, highly-customizable protocol
 - Just add headers
- The protocol that is the basis of the web we enjoy today

Multimedia over the Internet

Internet Multimedia

- Many applications on the Internet use audio or video
- Comparison with traditional web/HTTP:
 - Cannot tolerate loss, but a little delay may be ok
 - Data used after the transfer is complete
- Multimedia is more real-time
 - Performance *during* the data transfer matters

