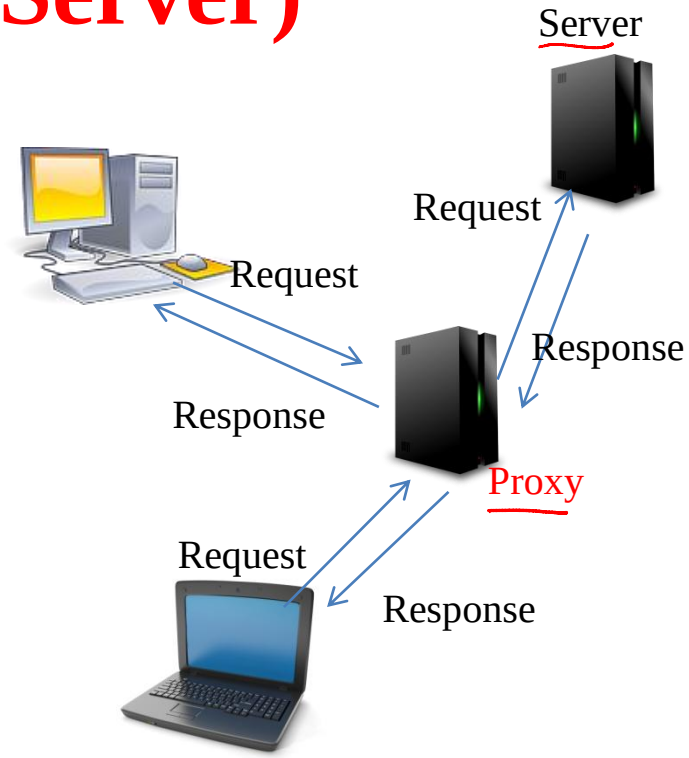
The World Wide Web (WWW): Web Caches

Kameswari Chebrolu

Web Caches (Proxy Server)

- Browsers access web via a cache (proxy server)
  - Can specify proxy address and port as part of browser's network settings
- If HTTP object request in cache, proxy returns the object (origin server not involved)
- Else, Proxy contacts original server, obtains object and returns it to client



Proxy Server

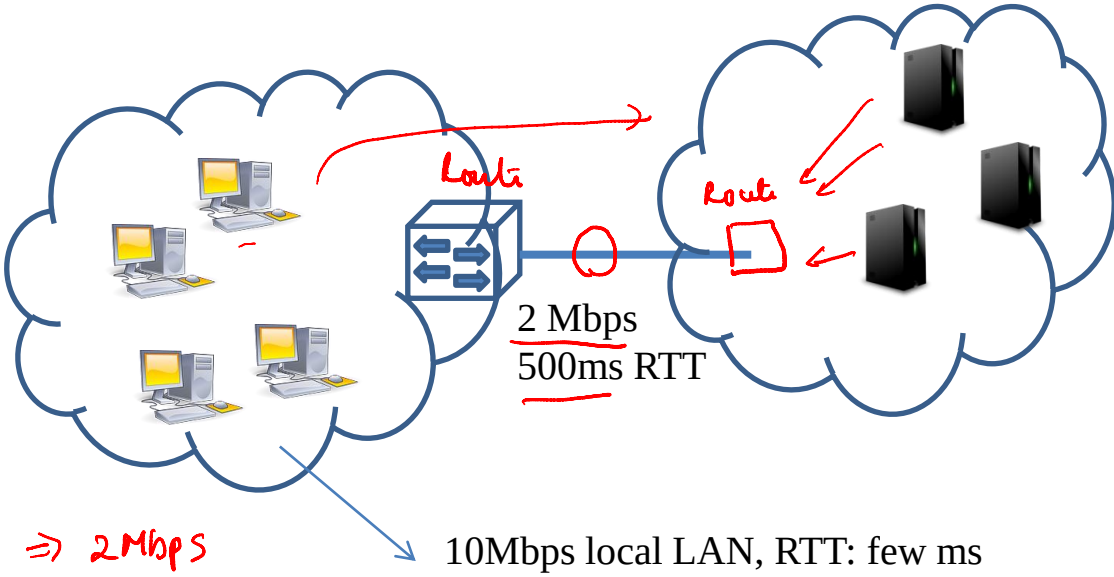
- Proxy server installed by ISP
  - Universities, Companies, local ISPs
- Proxy acts as server to clients and as client to external servers
- Internet dense with caches

Advantages

- Reduces response time of client request
- Saves bandwidth (prevents downloading of same content multiple times)
- Helps log usage, block undesired sites etc
- Enables “poor” content providers to effectively deliver content

Example

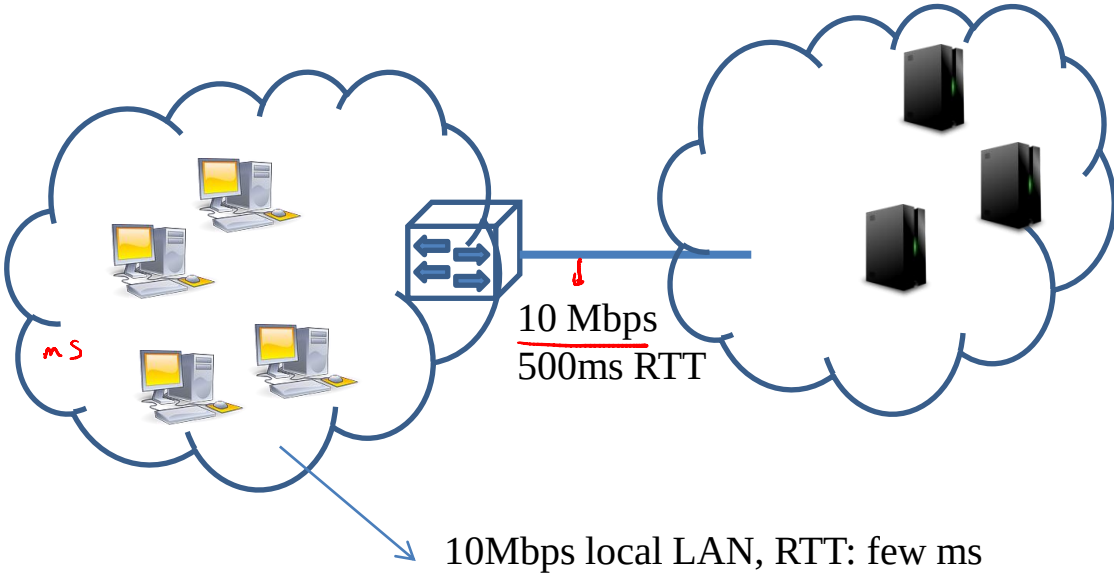
- Assumptions:
  - Average object size is 100Kbits
  - Request rate inside organization is 20 requests/sec



- Approximately, what is the ~~total~~ average response time?
  - ➢ Order of minutes (governed by queuing delay; considering TCP handshake time)

High Bandwidth Link

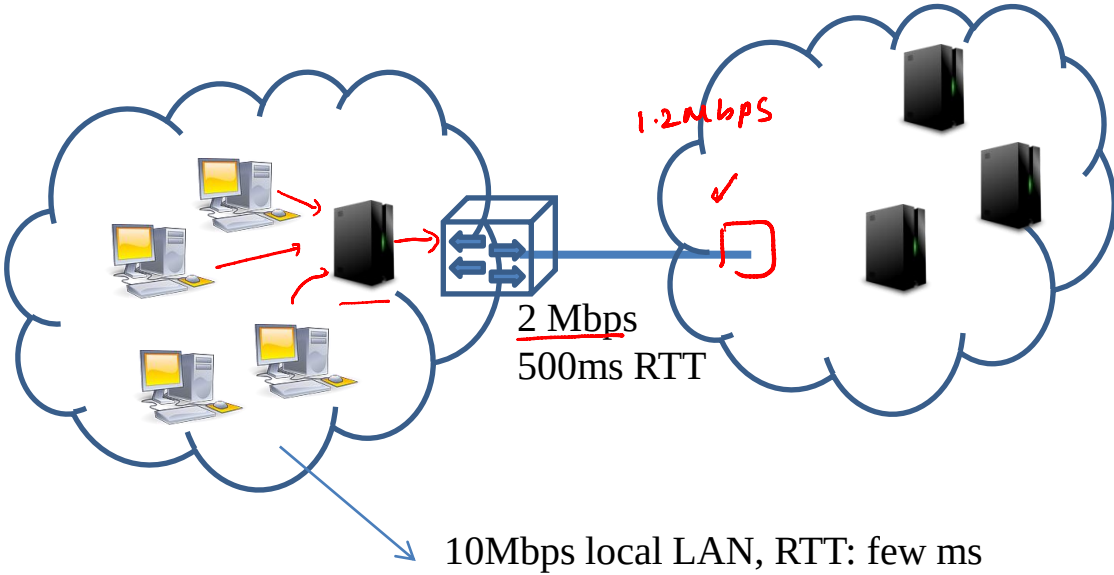
- Assumptions:
  - Average object size is 100Kbits
  - Request rate inside organization is 20 requests/sec



- Approximately, what is the total average response time?
  - ➢ Around 1 sec but can be a costly upgrade

Proxy Server

- Assumptions:
  - Average object size is 100Kbits
  - Request rate inside organization is 20 requests/sec
  - Hit rate of 0.4



- Approximately, what is the total average response time?
  - ➢ Roughly 1 sec [60% requests take ~ 1 sec, while 40% take a few ms]

Stale Cache

- How to verify if object in cache is up to date?
- HTTP 1.1 provides means to do this via cache-control headers
  - max-age=[seconds] : indicates freshness of the object (60 sec)
  - public : content is cacheable
  - private : content is not cacheable at proxy (can be cached at browser)
  - no-cache : need validation before releasing content
  - no-store : do not cache under any condition
  - E.g. Cache-Control: max-age=3600, public

Cache Validation

- Last Modified + Conditional GET
  - GET method has an If-Modified-Since header line
- Etag (hash based) + Conditional GET
  - GET method has an If-None-Match
- Server responds with “304 Not Modified” status code if cached copy is still good
  - Cache will use past cache-control headers or new ones if included in response

Example

GET /info.html HTTP/1.1

HOST: www.xyz.com

HTTP/1.1 200 OK

Date: Sat, 01 Nov 2013 10:23:21 GMT

Server: Apache

Last-Modified: Mon, 11 Aug 2013 09:54:12 GMT

Content-Type: text/html

(data)

Example

GET /info.html HTTP/1.1

HOST: www.xyz.com

If-modified-since: Mon, 11 Aug 2013 09.54.12 GMT

HTTP/1.1 304 Not Modified

Date: Sat, 09 Nov 2013 10:23:21 GMT

Server: Apache

(empty body)

Summary

- Proxy servers help enhance web performance and provide other useful features
- Stale caches are problematic but can be handled via cache-control headers