CSE561 – Traffic / Content Distribution

David Wetherall
djw@cs.washington.edu

Project presentations & reports

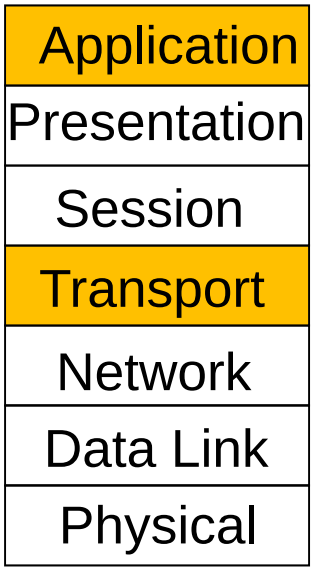
- Presentations in class, last day of classes
- 10? minutes per team
- Quick refresher of the problem, focus on what you have accomplished and what you have learned

- Accompanying writeup of no more than 6 pages (11pt)
- Can be turned in up to EOD the following day
- Captures your problem / approach / results / learnings and future directions.

Traffic and Content Distribution

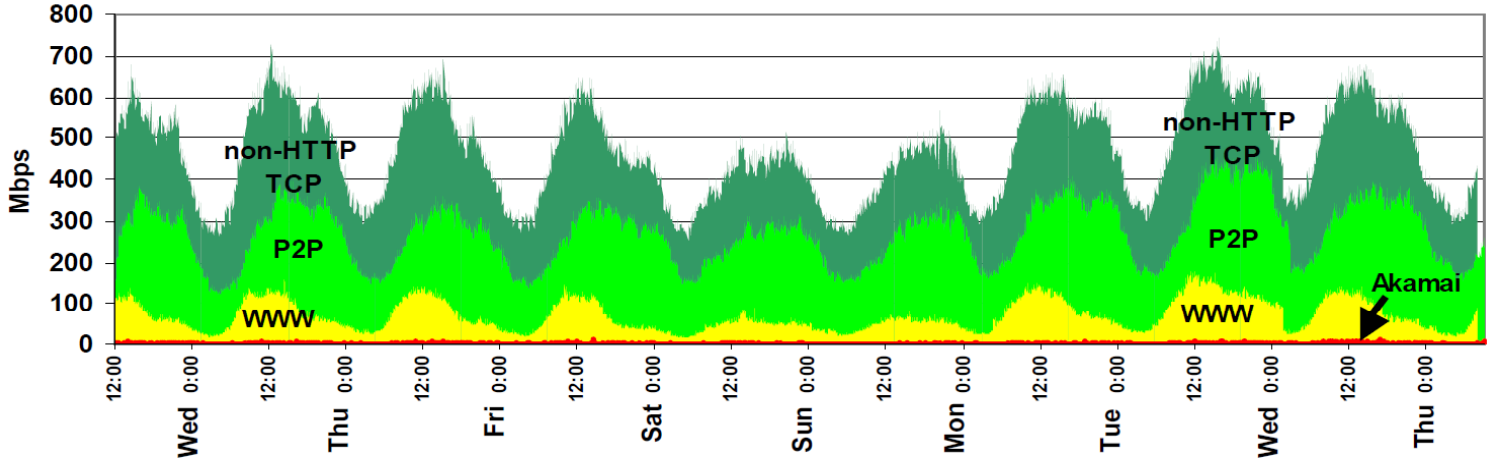
- Focus:
  - Things you should know about Internet workloads

- Traffic characteristics
- Caching
- CDNs
- Peer-to-peer



Rapid shifts in dominant applications

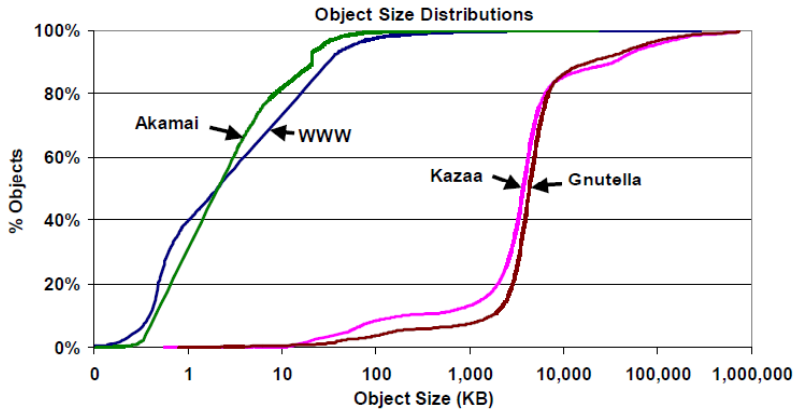
- The rise of P2P over Web in 2002 (Sariou et al.)

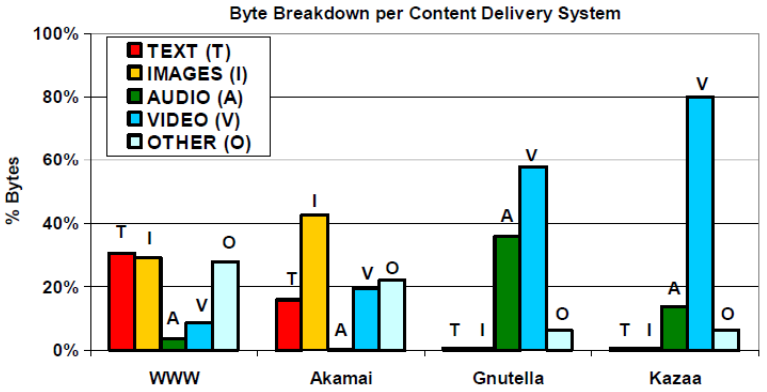


- Email, Web, P2P/BitTorrent, Skype, YouTube, Facebook

Leads to shifts in traffic features

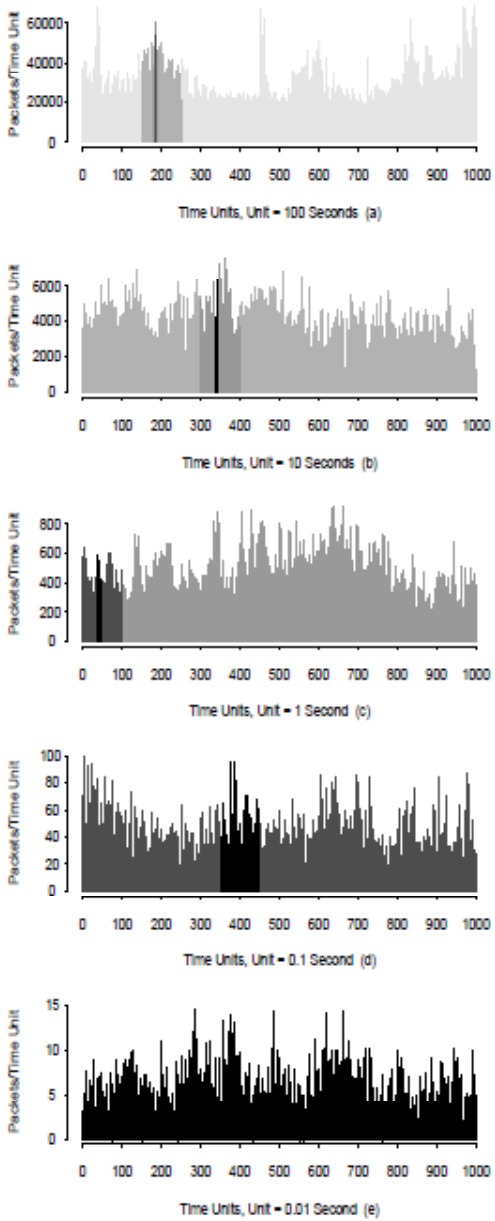
- With the rise of video, shift from many short connections to much longer/larger transfers





Self-similarity

- Network traffic is bursty over all timescales; Poisson is only a good model for human-driven
  - "On the self-similar nature of Ethernet traffic," Leland et al., 1993
- Aggregating Poisson traffic (exponential inter-arrival times) smoothes it
- But aggregating self-similar traffic just makes it burstier



More on objects …

- The size of transfers is heavy-tailed
  - Mostly small connections yet most bytes in a few large ones
- The popularity of objects has a power-law distribution
  - Zipf/Pareto for Web pages

Caching

- Implication of power-law traffic models is that caching is not very effective (by traditional measures)
  - More like “50%” than “95%”

- One rule of thumb:
  - Cache hit rate grows logarithmically with cache size

How to speed up content distribution

- Model is that many clients want the same objects
- 1. Remove server bottleneck
  - Replicate it
- 2. Place content close to clients
  - Reduces network load, speeds transfers (TCP effects etc.)

Content Distribution Networks

- Akamai as example

- Replicate content at locations near clients
- Replicas are really cached copies

- Magic is to connect client with nearby replica
  - Overrides DNS resolution for deployment
  - Client still uses URL, contacts server to get page
  - DNS maps Akamai server name to IP of nearby replica
  - Nearby might be RTT to client nameserver, or better

Peer-to-Peer

- BitTorrent as example
- Users serve dual role as replicas for each other
  - Issues of participation incentives
- Magic is to connect client with a set of nearby replicas
  - Application search process that favors better/faster partners
  - Emphasis on decentralization; no single authority or contact