Today

- Student Topics
- End-to-End Performance (15.4)
- Background (4 Internet Drafts)
- Technologies (4 papers)
- Break
- Performance (2 papers)
- For next week
End-to-end Performance

- Krishnamurthy and Wills, 2000. (WWW9)

- Performed repeated retrievals of objects on popular sites from multiple client locations.

- Studied effect of
  - Protocol options
  - Caching
  - Content from multiple servers
  - Range requests
- Tried to account for
  - Network delays
  - Server loads
  - Number of objects and bytes

By repeating tests and using large enough datasets.
• Did not consider
  • DNS lookup times
  • HTTP redirection
  • Dynamic content
  • Packet-level performance

• Ran 4 versions
  • HTTP/1.0 serial (single connection)
  • HTTP/1.0 burst (multiple parallel connections)
  • HTTP/1.1 serial (single persistent)
  • HTTP/1.1 burst (single pipelined, persistent)
Results

- HTTP/1.1 – pipelined, persistent connection had best performance, if connection is not closed prematurely, and there are requests for more than a few small objects.
- Caching is good; when validation is needed, persistent connections are essential.
- Multiple servers can improve performance if significant content is transferred.
Homework for next week

- Send me 2-3 page version of research proposal (or see me!!!).
- Read Search Engine and Spider papers.
- Work on semester research project.
- As always,
  - Post questions and comments to the class discussion group.
  - Bring something interesting to discuss, relating to Web performance.
Outline For Next Week

Search Engines
- Saratava et al., 2001. Rank-Preserving Two-Level Caching for Scalable Search Engines. SIGIR.

Spiders