On the Effect of Server Adaptation for Web Content Delivery

Yin Zhang (AT&T)

Joint work with

Balachander Krishnamurthy (AT&T)
Craig Wills (WPI)

Motivation

- Web sites have a strong incentive to reduce time-to-glass
- Challenge
  - client connectivity is heterogeneous
- Natural solution – server adaptation
  - client connectivity + content characteristics + client capability + server load + ...
  - → action to take

Internet

foobar.com

Study: What?

- Basic question – What exactly is the performance impact of server adaptation?
  - When and how much can server adaptation help?
  - Which action should the server take?
- Lots of previous work … but typically focusing on one individual action
- This study –
  - Provides a unified framework for assessing the impact of different server actions
  - Obtains useful insights through multi-site wide-area measurements

Factors Considered

- Client connectivity
  - Latency, bandwidth
- Content characteristics
  - Criteria: total bytes, container bytes, #objects
  - 3x3x3 = 27 buckets
    - derived from large proxy logs
    - further justified by examining popular Websites' pages
- Server actions
  - Altering the content
    - reducing number of images, reducing image size
  - Altering the location of the content
    - using a Content Distribution Network (CDN)
  - Altering manner of delivery
    - compression, bundling
  - Altering protocol options
    - using persistent connections
  - Combination of different actions

Experiment Methodology
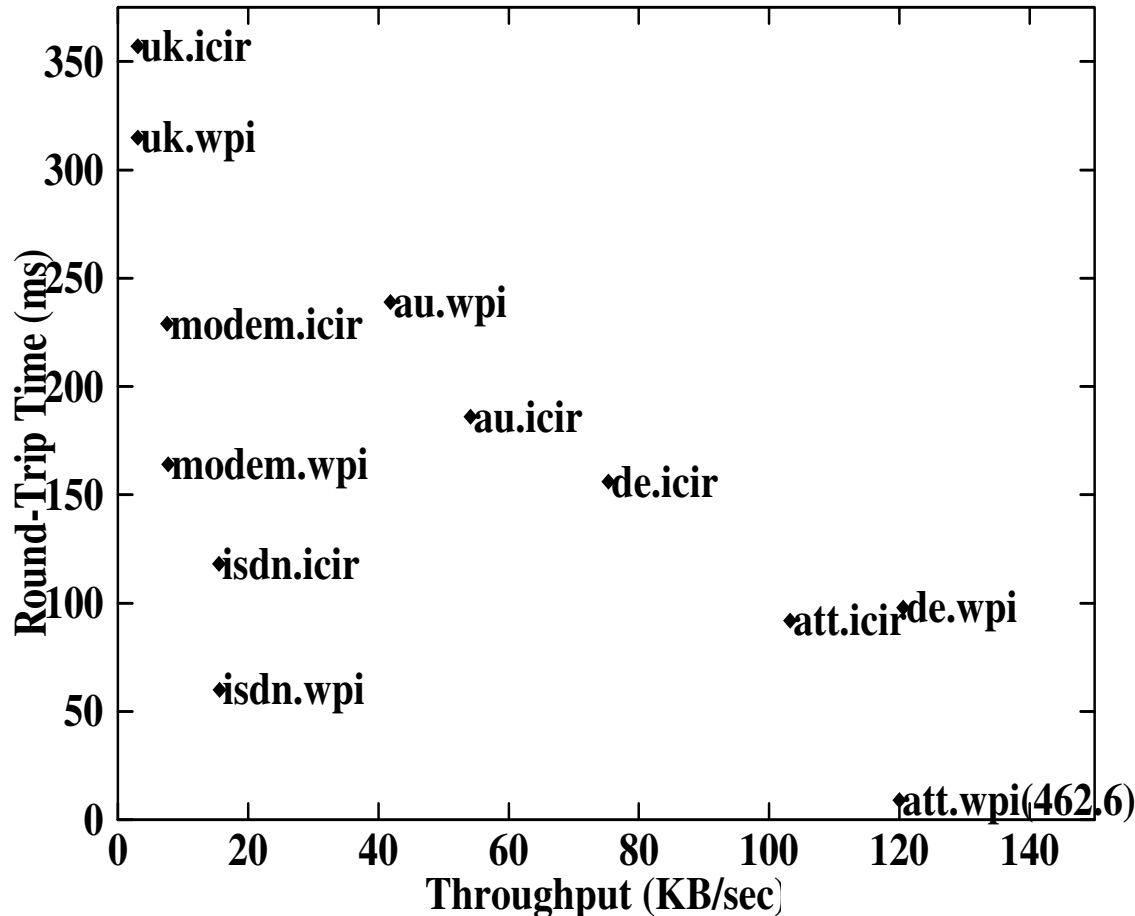
- A multi-site study
  - Server: Apache
    - West coast: icir
    - East coast: wpi
  - Client: httperf
    - US: att, modem, isdn
    - Intl: de, au, uk
- Canonical content served at each site
  - covering the space of buckets
- Experiments repeated at different times of day

Results

- Compression of HTML is not universally useful
    - It only works for bandwidth-constrained clients
- Persistent connections alone has limited benefit
    - Little improvement for all client/server combo
- Pipelining gives significant improvement
    - Exception: bandwidth-constrained clients
- Bundling gives significant improvement
    - Bundling alone is similar to pipelining
    - Compressed bundles help a lot under all conditions
    - CDN-served bundles – good idea for well-connected clients
- Reducing image size by half has little benefit
- Reducing the number of objects by half helps a lot under most conditions

Baseline: 4 parallel HTTP/1.0 connections

Contribution and Further Work

- Contribution
  - A unified framework for evaluating the impact of server adaptation
    - Can be applied by individual Web site
  - Insights we gained can be useful for improving client performance
- Further work
  - Evaluation of the feasibility of online client classification and server adaptation through real implementation
    - Our results are encouraging

Acknowledgments

- People who gave us accounts / logs