End-to-end
Information Exchange

Andrew Mall
CSC 450
10/15/01
Outline

- Server Volumes
- Proxy Filters
- Volumes and Filters: Practical Details
- Volume-Construction Algorithms
- Evaluation of Volume-Construction Algorithms
- Summary
Server Volumes

- Constructed using a heuristic:
- Accessed close together in time (temporal access)
- Share a syntactic prefix (in the same directory)
- Share resource attributes (similar content types, sizes, etc.)
- Modified together
- May have multiple volume lists
Proxy Filters

- Volume construction draws on information available at the server, but the proxy may not be interested in all of the resources in the volume.
- The proxy includes information about its caching policies with the HTTP request as a filter.
- Server applies this filter to the volume, returning hints that are relevant to what the proxy needs.
- Server doesn’t keep state for proxies, and proxies can use same filters across multiple servers.
## Practical Details

### Range of filter elements

<table>
<thead>
<tr>
<th>Filter element</th>
<th>Typical values</th>
<th>Implication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hints</td>
<td>On/Off</td>
<td>Should hints be sent?</td>
</tr>
<tr>
<td>RPV</td>
<td>{1, 2, 4}</td>
<td>Server should not send listed volumes</td>
</tr>
<tr>
<td>MaxPiggy</td>
<td>5, 10</td>
<td>Maximum # of hints in a message</td>
</tr>
<tr>
<td>MinAccess</td>
<td>100</td>
<td>Minimum # of accesses of a resource</td>
</tr>
<tr>
<td>MinProb</td>
<td>0.8</td>
<td>Minimum probability of next access</td>
</tr>
<tr>
<td>Level</td>
<td>0, 1, 2</td>
<td>Hints share prefixes to specified level</td>
</tr>
<tr>
<td>Size</td>
<td>1 MB, 10 KB</td>
<td>Volume elements size limitation</td>
</tr>
<tr>
<td>OKType</td>
<td>HTML</td>
<td>Limit hints to specified content types</td>
</tr>
<tr>
<td>NotOKType</td>
<td>CGI</td>
<td>Do not send hints about this type</td>
</tr>
<tr>
<td>Lastmod</td>
<td>1 min</td>
<td>Minimum time since last modification</td>
</tr>
<tr>
<td>HTTP method</td>
<td>PUT</td>
<td>Resources must allow listed method</td>
</tr>
<tr>
<td>HTTP header</td>
<td>ETag, Accept</td>
<td>Resource must match header criteria</td>
</tr>
</tbody>
</table>

Table 13.1, page 494
Practical Details

- Filters and hints could be included as a header on HTTP requests and responses, but this produces latency problems in satisfying requests.
- Instead, HTTP/1.1 supports the use of the Trailer header.

- GET /random.html HTTP/1.1
  Host: smartserver.com
  TE: trailers
  Piggy-filter: maxpiggy=5;
  content-type=gif

- HTTP/1.1 200 OK
  Trailer: P-volume
  Transfer-encoding: chunked
  < size of chunk >
  < data >
  0
  P-volume: vol=7;
  pe="/random-img1.gif,895527629,546";
  pe="/random-img3.gif,891527021,192";
  pe="/random-img21.gif,821993421,900"

CRLF
Volume-Construction Algorithms

- **Performance metrics:**
  - Hint size – average number of hints on a response
  - Recall – how many client request predicted by hints
  - Precision – fraction of hints that are successful
  - Update fraction – was request predicted and resource was requested from the proxy in the past

- **May be combined**
  - Maximize recall subject to an upper bound on average hint size
  - Minimize hint size subject to a lower bound on recall
  - Maximize recall subject to a lower bound on precision
  - Maximize precision subject to a lower bound on recall
Volume-Construction Algorithms

- Example algorithm based on temporal locality (accessed close together in time)
- Volume $v$ should include resource $s$ if a request for $r$ is typically followed by a request for $s$ by the same client
- Subject to time constraint
- $(\# \text{ of requests for } r \text{ followed by requests for } s) / (\text{total } \# \text{ of requests for } r) = \text{ratio}
- If $(\text{ratio} > \text{threshold})$ then $s$ is included in $v$
Evaluation of Algorithms

- Server logs indicate that 85% of requests were for 10% of resources.

- Experiments across a collection of logs indicate that the two-pass (thinning) algorithm resulted in a recall of 60-80%, precision of 80-88%, and hint size of 2-10.

- Accuracy not significantly affected by sampling logs.
Summary

- Server groups related resources into volumes and applies filters, customizing hints for the requesting proxies.
- Proxy filters and server hints can be used with current HTTP/1.1.
- Experiments have shown that successful algorithms rely on close temporal access and omit duplicate predictions.
- Wide variety of proxy filters?
- Constructing volumes grouping resources that are accessed together?