COS 316 Precept #6: Cache Eviction (Replacement)

git clone https://github.com/cos316/precepts
Navigate to precept7 directory
Overview of Web Caching

• Basic idea:
  • Bring objects “closer” to clients
• Three primary features:
  • Reduce network bandwidth
  • Reduce client-perceived delays
  • Reduce load on server
• Cache Replacement Strategy
  • When a cache becomes full, which object should be **evicted/replaced**?
Cache Eviction Algorithms

• High level
  • Client requests a new object
  • If object is in cache, return the object
  • If object is not in cache:
    • Get object from server/provider and return the object
    • Cache full:
      • Identify an object in cache to evict
      • Evict the object in the cache
      • Replace with new object (insert new object)
    • Cache not full:
      • Admit the new object to the cache
Cache Eviction Algorithms

• Least recently used (LRU): Evict the object from the cache whose last request is the oldest

• First-in, First-out (FIFO): Evict the object from the cache that has been in the cache the longest

• Many others...
### LRU Cache

<table>
<thead>
<tr>
<th>ID</th>
<th>Size</th>
<th>Request Time</th>
<th>Admit Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>2</td>
<td>13:00</td>
<td>11:00</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>13:45</td>
<td>13:45</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>15:01</td>
<td>12:01</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>11:53</td>
<td>11:33</td>
</tr>
<tr>
<td>7</td>
<td>17</td>
<td>13:30</td>
<td>13:30</td>
</tr>
<tr>
<td>8</td>
<td>10</td>
<td>16:00</td>
<td>16:00</td>
</tr>
</tbody>
</table>

Current time: 16:00

Current cache size: 45

Cache capacity: 50

*Note: The last request is shown as **16:00** since the current time is 16:00.*
The document appears to be about the LRU (Least Recently Used) cache algorithm. Here is the text representation:

**LRU**

<table>
<thead>
<tr>
<th>id</th>
<th>size</th>
<th>request</th>
<th>admit</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>2</td>
<td>13:00</td>
<td>11:00</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>13:45</td>
<td>13:45</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>15:01</td>
<td>12:01</td>
</tr>
<tr>
<td>11</td>
<td>8</td>
<td>11:30</td>
<td>11:30</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>11:53</td>
<td>11:33</td>
</tr>
<tr>
<td>7</td>
<td>17</td>
<td>13:30</td>
<td>13:30</td>
</tr>
<tr>
<td>8</td>
<td>10</td>
<td>16:00</td>
<td>16:00</td>
</tr>
</tbody>
</table>

Cache capacity = 50
Cache size = 45

Current time: 16:00
### FIFO Cache

<table>
<thead>
<tr>
<th>ID</th>
<th>Size</th>
<th>Request Time</th>
<th>Admit Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>15:01</td>
<td>12:01</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>13:45</td>
<td>13:45</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>13:00</td>
<td>11:00</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>11:53</td>
<td>11:33</td>
</tr>
<tr>
<td>11</td>
<td>8</td>
<td>11:30</td>
<td>11:30</td>
</tr>
<tr>
<td>7</td>
<td>17</td>
<td>11:30</td>
<td>11:30</td>
</tr>
</tbody>
</table>

Current time: 16:00

Cache capacity = 55
Cache size = 45
Experiments

> cd <Precepts repo> #https://github.com/cos316/Precepts

> Make sure you have the main branch checked out

> git pull

> cd precept7/webcachesim-master

> make
Trace File Form

• Request traces must be given in a space-separated format with three columns
• time - long long int
• id - long long int, used to uniquely identify objects
• size should be a long long int, object's size in bytes

• Example
  
<table>
<thead>
<tr>
<th>time</th>
<th>id</th>
<th>size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>120</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>64</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>120</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>14</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>120</td>
</tr>
</tbody>
</table>

• See test.tr
Using the Simulator*

> ./webcachesim test.tr LRU 1000

LRU:1000 bytes, 10492 reqs, 8495 hits, 81 hits/reqs(%) 

> ./webcachesim test.tr FIFO 1000

FIFO:1000 bytes, 10492 reqs, 8206 hits, 78 hits/reqs(%) 

* Derived from https://github.com/dasebe/webcachesim
Experiments

- Trace data from a production CDN
  - cd1-10M.tr *
  - 10 million requests /
    Object sizes from 10 byte to .7GB
- LIFO and FIFO
- Vary cache sizes
- 16000000
- 32000000
- 64000000
- 128000000
- 256000000
- 512000000
- 1024000000
- 2048000000
- 4096000000

- Create a Google Sheet
- Three columns
- SIZE LRU FIFO
- Copy results accordingly
- Select three columns to create line chart

* Practical Bounds on Optimal Caching with Variable Object Sizes Daniel S. Berger, Nathan Beckmann, Mor Harchol-Balter. ACM SIGMETRICS, June 2018
Experiments

- LRU and FIFO
- Vary cache sizes
  - 80
  - 160
  - 320
  - 640
  - 1280
  - 2560
  - 5120

- Create a Google Sheet
- Three columns
- SIZE LRU FIFO
- Copy results accordingly
- Select three columns to create line chart