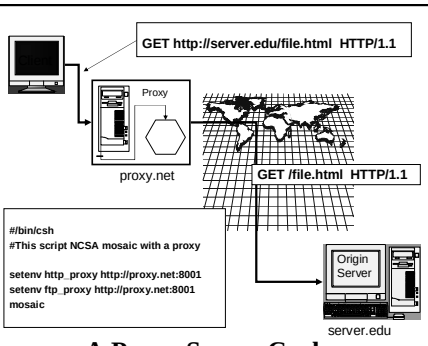
Internet Caching Architecture

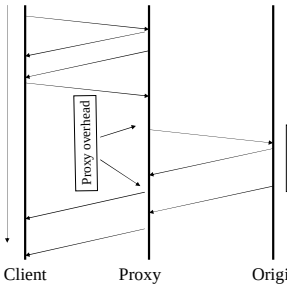
34



A Proxy Server Cache

INTERNET ENGINEERING

Cache Performance

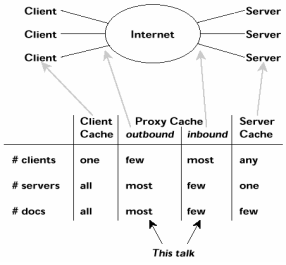


Hit Ratio

$T_{avg}=h.T_{cache}+(1-h).T_{origin}$

INTERNET ENGINEERING

Cache Location



	Client Cache	Proxy Cache *outbound*	Proxy Cache *inbound*	Server Cache
# clients	one	few	most	any
# servers	all	most	few	one
# docs	all	most	few	few

This talk

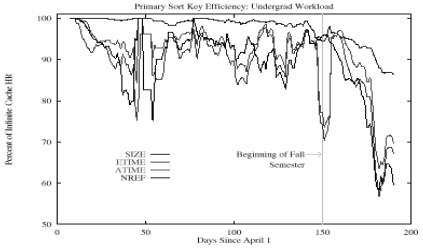
INTERNET ENGINEERING

Cache Replacement Policies

- It is a delicate balance between the cache and performance. Following are some attributes which can be used for cache document replacement:
  - SIZE document size (bytes)
  - ETIME time doc entered cache
  - NREF number references to doc
  - ATIME time doc last accessed

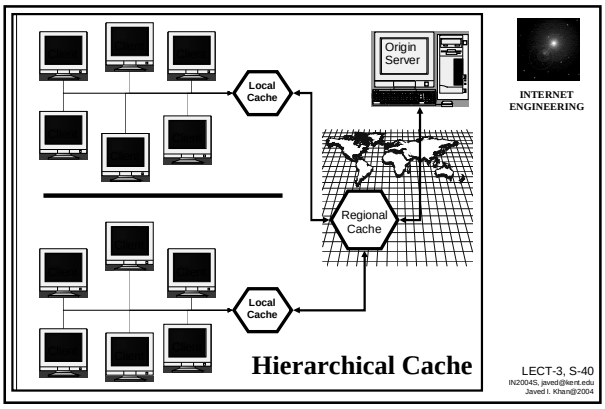
INTERNET ENGINEERING

A Typical Performance Result: Primary Key Comparison (Cache Size = 10% of max needed)

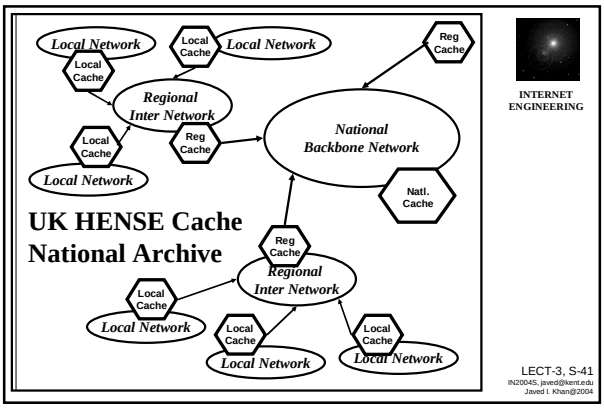


Courtesy: Removal policies in network caches for World-Wide Web documents Marc Abrams, Charles R. Standridge, Ghaleb Abdulla, Edward A., Fox and Stephen Williams, ACM SIGCOMM '96. Aug. 26-30, 1996, Stanford, CAPages 293-305

INTERNET ENGINEERING



INTERNET ENGINEERING



INTERNET ENGINEERING

Caching Protocols

Cache Related Protocols

- ICP: Internet Cache Protocol
  - Proposed in 1995 as a part of Harvest system, later Squid Cache.
- CARP: Cache Array Resolution Protocol
  - Proposed in 1996 as “Super Proxy Script” by SHARP and later adopted by Microsoft.
- CDP: Cache Digest Protocol
  - Also part of Squid. Evolved around 1998.
- WCCP: Web Cache Coordination Protocol
  - Evolved around 1999 and used by CISCO router/switch interceptors.

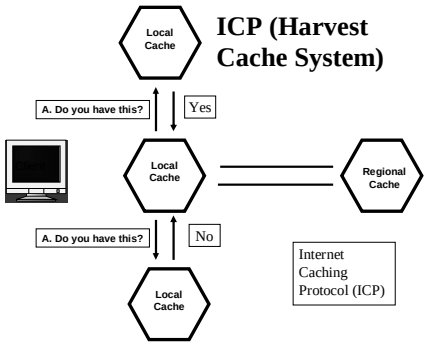
INTERNET ENGINEERING

ICP: Internet Cache Protocol

- ICP [Chankhunthod et. al. 1995] is used by a set of caches to query each other to find an object.
- Caches can be organized in multilevel hierarchy.
- Sends parallel query to siblings. Fastest to reply gets the request.
- If no sibling responds with a specified time query is sent to parent. Parent repeats the process.
- When object is found all caches in the path may store the response for their own use.

INTERNET ENGINEERING

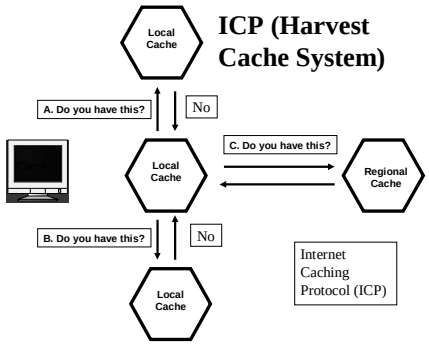
ICP (Harvest Cache System)



- Parallel Neighbor and Parent check.
- Proactive caching.
- UDP messaging

INTERNET ENGINEERING

ICP (Harvest Cache System)



Internet Caching Protocol (ICP)

- Parallel Neighbor and Parent check.
- Proactive caching.
- UDP messaging

INTERNET ENGINEERING

ICP: Internet Cache Protocol (2)

- ICP query uses UDP Messages. More efficient than HTTP. But may face firewall problems.
- Each system has to be manually configured and thus is cumbersome to manage for a large system.
- If object is not found in one hop, then typically delay increases. It takes some time before the response percolates to the original client.

INTERNET ENGINEERING

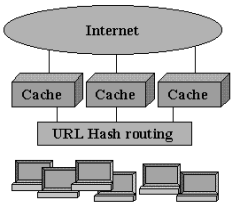
Cache Digest

- Squid keeps it's "list" of objects in an in-memory hash. The hash table (which is based on MD5) helps Squid find out if an object is in the cache without using huge amounts of memory or reading files on disk. Periodically Squid takes this table of objects and summarizes it into a small bitmap. If a bit in the map is on, it means that the object is in the store, if it's off, the object is not.
- This bitmap/summary is available to other caches, which connect on the HTTP port and request a special URL. If the client cache wants to know if the server has an object, it simply performs the same mathematical function that generated the values in the bitmap. If the server has the object, the appropriate bit in the bitmap will be defined.
- There are various advantages to this idea: inter-cache communication can use significant amounts of bandwidth. Each request to one cache sparks off a series of requests to all neighboring caches. Each of these queries also causes some server load: the networking stack has to deal with these extra packets, for one thing. With cache-digests, however, load is reduced. The cache digest is generated only once every 10 minutes (the exact value is tunable).
- The transfer of the digest thus happens fairly seldom, even if the bitmap is rather large (a few 100kbytes is common.) If you were to run 10 caches on the same physical network, however, with each ICP request being a few hundred bytes, the numbers add up. This network load reduction can give your cache time to breathe too, since the kernel will not have to deal with as many small packets.
- ICP packets are incredibly simple: they essentially contain only the requested URL. Today, however, a lot of data is transferred in the headers of a request. The contents of a static URL may differ depending on the browser that a user uses, cookie values and more. Since the ICP packet only contains the URL, Squid can only check the URL to see if it has the object, not both the headers and the URL. This can (very occasionally) cause strange problems, with the wrong pages being served.
- With cache digests, however, the bitmap value depends on both the headers AND the URL, which stops those strange hits of objects that are actually generated on-the-fly (normally these pages contain *cgi-bin* in their path, but some don't, and cause problems.)
- Cache digests can generate a small percentage of false hits: since the list of objects is updated only every 10 minutes, your cache could expire an object a second after you download the summarized index.

INTERNET ENGINEERING

CARP: Cache Array Resolution Protocol

- CAR Allows a group of Caches to work as a single logical cache by dividing the URL name space using distributed hash function.
- It uses a two-way hashing to map URL to Proxies.
- Request for a URL is directed to only one Cache thus it avoids delay.



INTERNET ENGINEERING

CARP Hashing

- Step-1: calculate a hash value for each Proxy (there are 4):
- Step-2: calculate a hash value for each URL

Proxy	Hash
Jericho1	13
Jericho2	8
Jericho3	5
Jericho4	28

www.microsoft.com
19

INTERNET ENGINEERING

Hash Combination

- Using Hash Combination function find a Hash Score for the URL for each Proxy.
- Pick the Proxy with highest score.

Proxy	Hash	www.microsoft.com: 19
Jericho1	13	5
Jericho2	8	9
Jericho3	5	7
Jericho4	28	4

INTERNET ENGINEERING

Load Distribution

- The Hash Combination function uses "load_factor" for each proxy to determine the load.

Proxy	Hash	www.microsoft.com	www.yahoo.com	www.msn.com	www.ibm.com
Jericho1	13	5	6	10	4
Jericho2	8	9	2	7	5
Jericho3	5	7	4	3	10
Jericho4	28	4	7	8	1

INTERNET ENGINEERING

Proxy Addition

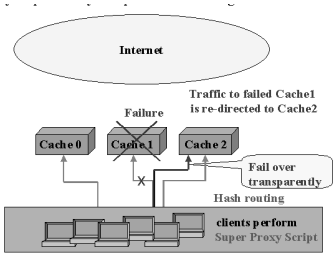
- New proxy takes off "1/5th" load from each.
- Other URL's stays in previous server. Thus the scheme offers stability.

Proxy	Hash	www.microsoft.com	www.yahoo.com	www.msn.com	www.ibm.com
Jericho1	13	5	6	10	4
Jericho2	8	9	2	7	5
Jericho3	5	7	4	3	10
Jericho4	28	4	7	8	1
Jericho5	14	2	4	4	6

INTERNET ENGINEERING

Proxy Failover

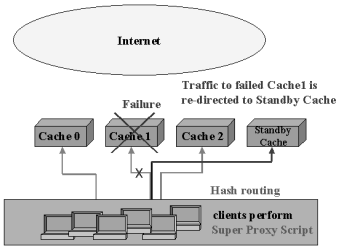
- If there is no reply from a proxy the URL gets mapped to the next highest scoring proxy.



INTERNET ENGINEERING

Proxy Deletion

- If there is no reply from a proxy the URL gets mapped to other standby proxy, or the loads can be redistributed.



INTERNET ENGINEERING

Array Membership Management

- Proxy uses an "array manager" to maintain a current list of the members of a particular proxy array and to make that list available to other systems which request it, such as downstream clients and proxies.
- Communications between array managers are handled through HTTP and remote procedure calls (RPC). RPC interfaces are used to handle modifications to the array table, such as membership, status, and parameters. HTTP is used to publish array information.
- Oone-stop shopping"—any one member of the array will have current information about every other member of the array. Therefore, a client need only query one randomly selected array member in order to properly route into the array.
- The membership lists contains information, including:
- The URL that an array manager should call in order to get the array information from a remote manager.
- Load factors to allow a different proportion of the requests for an array to be sent to different machines. Load factors could be a factor if new machines that have larger hard drives or significantly greater processing power have been added to an array.
- Time-to-Live (TTL) countdown until array members are checked again for status.
- Global parameters, such as how often any member of the array table should ask other members for an update in array membership.

INTERNET ENGINEERING

Example Proxy Array Table

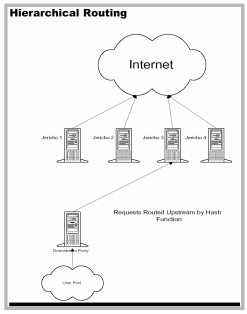
```
Proxy Array Information/1.0
ArrayEnabled: 1
ConfigID: 866749230
ArrayName: Test Cluster
ListTTL: 900

CATNET07 157.55.98.140 80 http://CATNET07:80/array.dll MSProxy/2.0 171 Up 100 3000
CATNET09 157.55.98.140 80 http://CATNET09:80/array.dll MSProxy/2.0 171 Up 100 1500
GSFGROUP 157.55.98.140 80 http://GSFGROUP:80/array.dll MSProxy/2.0 171 Up 100 1500
GENACCTS2 157.55.98.140 80 http://GENACCTS2:80/array.dll MSProxy/2.0 171 Up 100 500
```

- Two cases of Membership Management
- Upstream Table Management
  - A Proxy manages its own "impression" of the upstream tables. Whenever the TTL countdown expires (usually set for several minutes), the proxy queries for a new array table. A proxy within an array also watches all HTTP requests to any array member in order to determine the status of
- Local Table Management
  - If a request fails, the local proxy marks that proxy member as down in its table for a given TTL period and doesn't forward requests to that member until the TTL expires, and the next table query shows it is active.

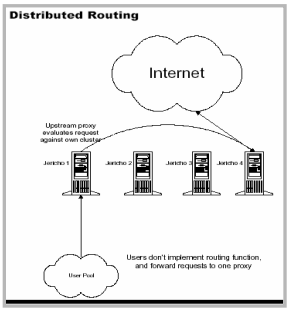
INTERNET ENGINEERING

Hierarchical Routing



INTERNET ENGINEERING

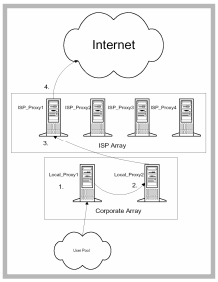
Distributed Routing



INTERNET ENGINEERING

Combination Routing

- Distributed and hierarchical caching can be easily combined. For example, all requests within a workgroup or branch office might first be resolved within its own array of proxies and then forwarded to the enterprise or ISP proxy array as needed. Examining the case of client à local proxy array à ISP proxy array à Internet, there may be up to four routing calculations per request:
- 1. The client forwards to local proxy #1. That proxy applies the routing algorithm against its own array and determines that local proxy #2 in its own array should handle the request and forwards it.
- 2. Local_proxy #2 applies the routing algorithm against its array and decides that it is the proper proxy to handle this request.
- 3. Local_proxy #2 doesn't have the object in its cache, so it applies the routing algorithm against the upstream array and forwards the request to ISP_Proxy #1
- 4. ISP_Proxy #1 applies the routing algorithm against its array and discovers that its the correct proxy to handle this request. Since it doesn't have it in its cache, it forwards the request to the Internet.



INTERNET ENGINEERING

CARP vs ICP

- ICP
  - Excellent for LAN parent/child configuration.
  - Messaging can induce congestion.
  - Negative scalability.
  - Not good for WAN (firewall stops UDP).
  - Duplicate storage (can waste space).
- CARP
  - Deterministic one shot routing.
  - Positive scalability.
  - Load distribution possible.
  - Hot document can still cause problem.

INTERNET ENGINEERING

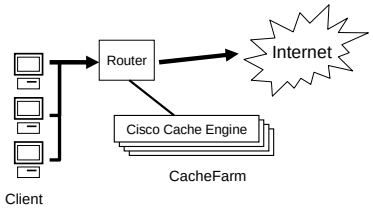
Transparent Caching

- A Router or Switch Interceptor Intercepts TCP/HTTP traffic and redirects it to a cache engine.
- The control is performed by Web Cache Coordination Protocol (WCCP).
- WCCP is capable of coordinating load balancing among multiple caches and ensuring availability of heart bit for reliable operation.

INTERNET ENGINEERING

3. CISCO CACHE ENGINE(2)

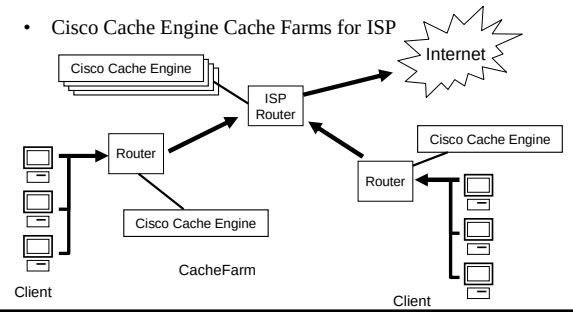
- Cache Engine connected directly to the router.



INTERNET ENGINEERING

3. CISCO CACHE ENGINE(3)

- Cisco Cache Engine Cache Farms for ISP

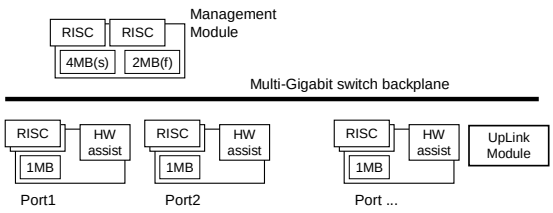


4. Alteon's ACEdirector(1)

- It is a general purpose redirector
- Examines all packet going through it, redirecting Web packets to a cache.
- For each packet examine the packet using SUREFIRE algorithm and send the packet to appropriate destination.
- In background, checks the health of the caches, exchanging keep-alive messages.
- It is high-performance processor intensive packet examination and manipulation equipment.

INTERNET ENGINEERING

4. Alteon's ACEdirector(2)



* Packet Examination and Redirection is distributed through the port.

Reference Material

- Primary source RFC-2068.
  - Specially read section:
    - Section 13
    - Section 14.9
- Cache Array Resolution Protocol (CARP) V 2.0
- Super Proxy Script
- Books:
  - Web Protocols & Practices, Addision Wesley, Krishnamurty & Rexford, 2001.
    - Sections 6 & 7.

INTERNET ENGINEERING

Research Perspective

- Design a new cache mechanism or suggest improvement of existing cache mechanisms for:
- Research Issues
  - How to Cache Multimedia Documents?
  - Time Sensitive Traffic (Real time systems)?
  - Dynamic Documents?
  - Distributed Cache

INTERNET ENGINEERING

Research Perspective: Prefetching

- Actively prefetch documents before their fast use, thus reducing the response time.
- Research Issues:
  - DNS Prefetching
  - TCP Connection Pre-staging
  - HTTP Prefetching
  - Reduction in wasted bandwidth.
  - Link prediction.

INTERNET ENGINEERING

Research Perspective: CDN

INTERNET ENGINEERING

- Keep the content over suitable locations in the Internet to serve then fast. Use direct mechanism to keep them up-to-date.
- Research Issues
  - Optimum replica placement.
  - Request routing to nearest content.
  - Content based routing.
  - Low cost content distribution.

Research Perspective: Active Proxy

INTERNET ENGINEERING

- Proxies can participate in active processing of content rather than passive storage.
- Research Issues:
  - Multiparty Service Composition.
  - Active Proxy Architecture.
  - Security of Code Execution.
  - Service QoS.