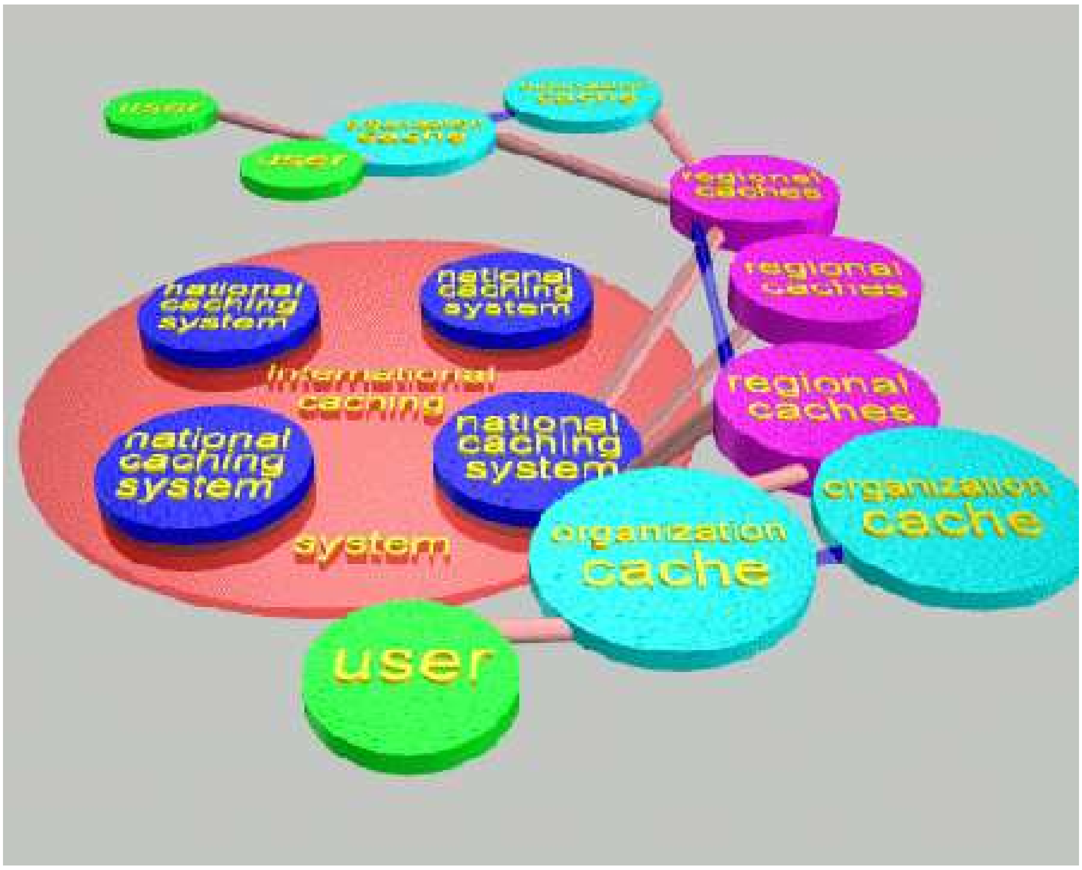
Squid Internet Cache

Lukas Fetz, Sigi Herzog

August 29, 1997

WWW Kommunikation erfolgt über TCP-Verbindungen; jeder Request bedeutet eine neue Verbindung (für jeden Client, für jedes Dokument). Verfügbare Bandbreite wird verschwendet; Performance degradiert



Web Caching

1. Ziel

- weniger Netzwerkverkehr
- Web-Seiten werden schneller geladen
- geringere Last am Ursprungsserver

Web Caching

2. Probleme

- Cache-Replikation kostet Massenspeicher
- Korrektheit der Daten
- geeignete Cache-Hierarchien
- unterschiedliche Caching Software

Web Caching

3. Anforderungen

1. Kooperation mit mehreren anderen Servern

2. Skalierbarkeit bez. der Web Entwicklung

3. Transparenz: Fallback im Falle von Serverfehlern

Caching Software?

- Squid Internet Object Cache, National Laboratory for Applied Network Research (http://www.nlanr.net/Squid)

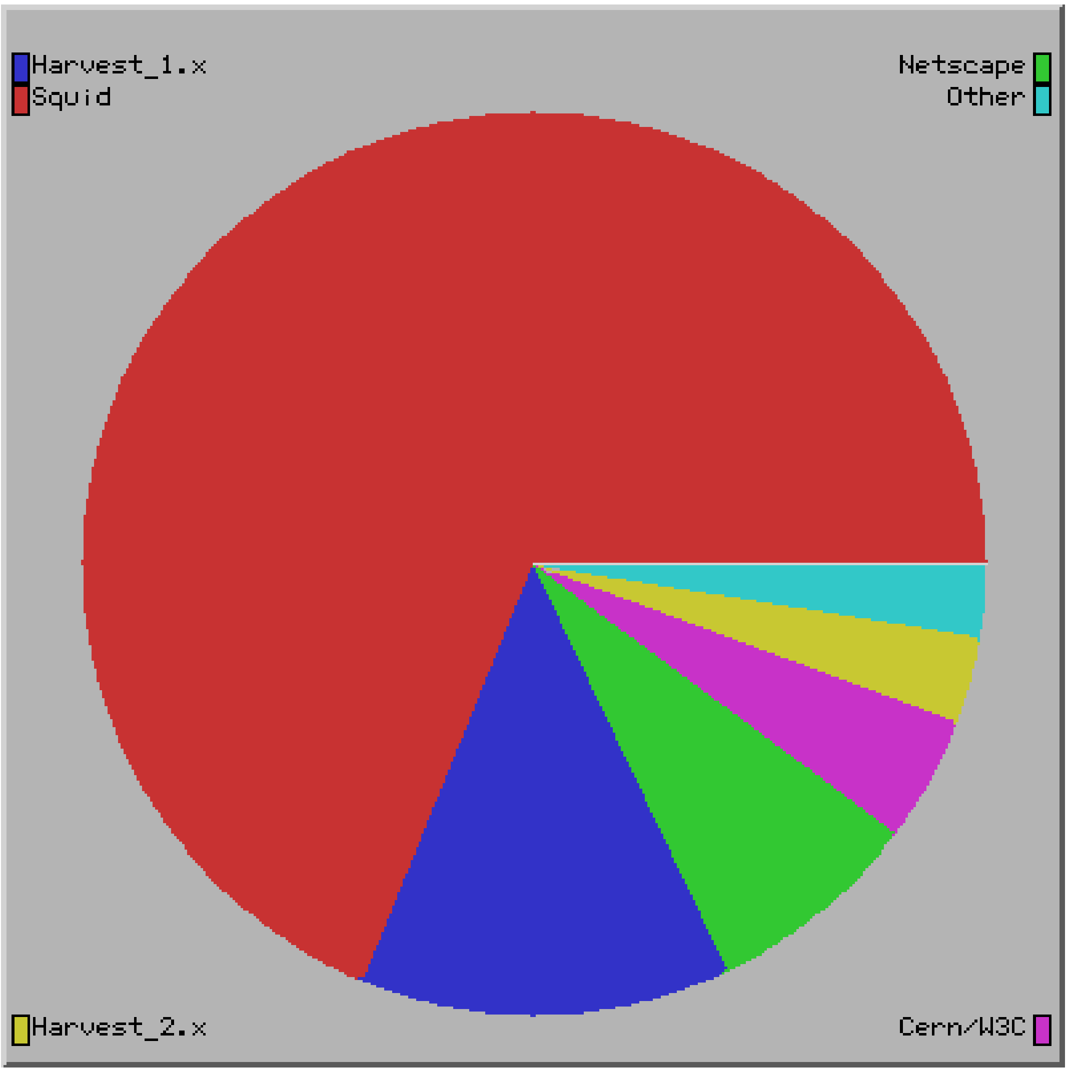
- NetCache, Kommerzielle Weiterentwicklung des Harvest-Servers (http://www.netcache.com)

- Netscape Proxy Server (http://www.netscape.com)

- Jigsaw, Referenz Caching-Server des W3C, unterstützt HTTP/1.1 und ICP

Caching Survey

Verwendete Caching Software.
Quelle: ICM Warschau, Oktober 1996



Warum Squid ?

- frei verfügbar
- läuft unter den meisten UNIX Systemen
- basiert auf Harvest v1.4 Code
- ICPv2 für Server Kooperation
- weit verbreitet

Protokolle - ICP

Dient zur Kommunikation zwischen Cache-Servern. Ist aus Performance Gründen als UDP-Service (verbindungslos) implementiert.

- data object location
- cache selection

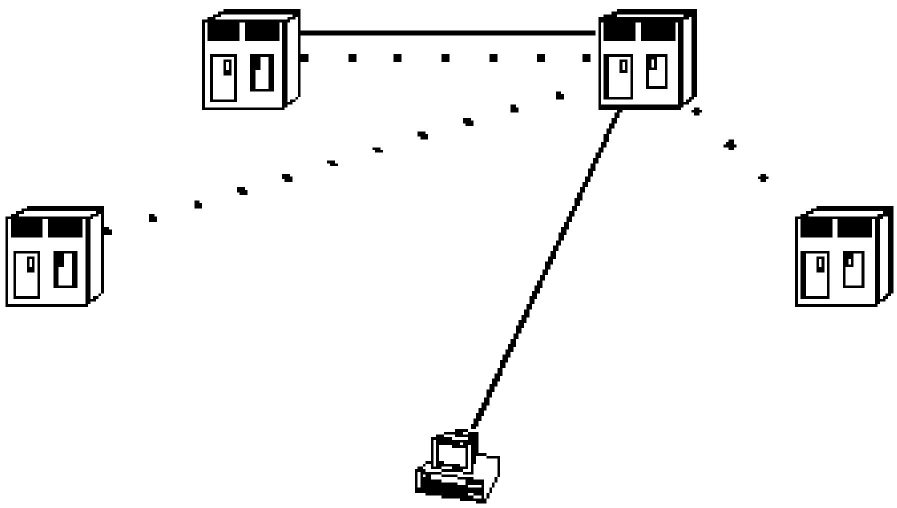


Fig.4 Simple client using a cooperating cache

UDP connection (discovery) . . . . .
TCP connection (retrieval) ————

Protokolle - HTTP

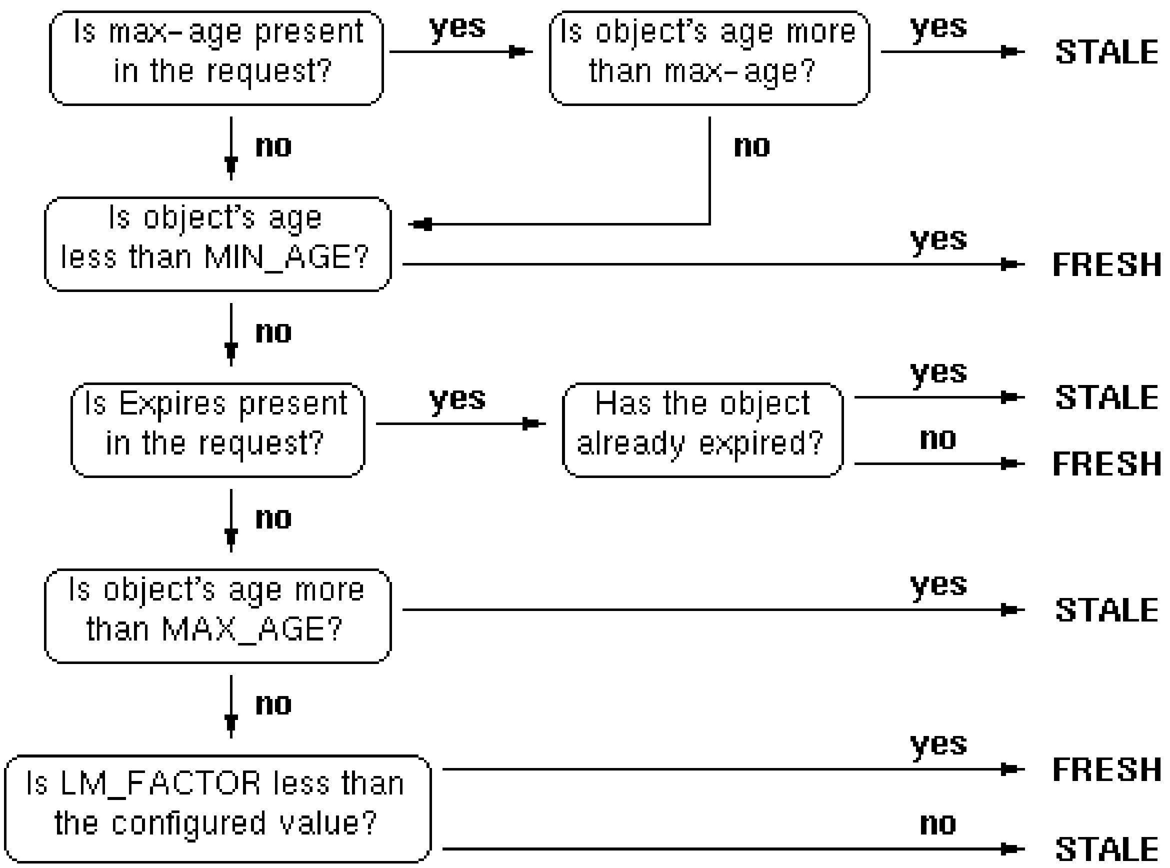
1. Expiration Model
   Weniger Requests an den Ursprungsserver (Serverlast)

   - Expires Entity Header
   - max-age Cache-Control (HTTP/1.1)

2. Validation Model
   Nur wenn sich was geändert hat wird das Dokument geschickt (Netzwerklast)

   - If-Modified-Since (conditional GET)
   - Cache Validator

Squid Refresh Rules



age = now - object date
lm_age = object date - last modified time
lm_factor = age / lm_age

MIN_AGE, MAX_AGE, LM_FACTOR are configuration variables and depend on the object's URL as well

Last-Modified: optional header containing the last modification date of an object
Expires: optional expiry time specified by HTTP servers
max-age: optional directive which specifies the maximum allowed age of an object

Squid Konfiguration

Die wichtigsten Squid Konfigurationsparameter. In der Datei *squid.conf* sind sämtliche TAGS ausführlich dokumentiert. Weitere Informationen gibt es unter:
http://ircache.nlanr.net/Squid.

4. Neighbour Selection

cache_host
cache_host_domain
local_domain
hierarchy_stoplist
cache_stoplist

```
cache_host parentcache.example.com parent   3128 3130
cache_host parent4com.example.com  parent   3128 3130
cache_host childcache2.example.com sibling  3128 3130
cache_host_domain parentcache.example.com !.com
cache_host_domain parent4com.example.com  .com
```

Squid Konfiguration

5. Cache Size

cache_mem

6. Logfile Pathnames, Cachedirectories

cache_dir
cache_access_log
cache_log
cache_store_log
debug_options

7. External Support Programs

Squid Konfiguration

8. Cache Tuning

9. Timeouts

10. Access Controls

acl
http_access
icp_access
miss_access

Squid Konfiguration

11. Administrative Parameters

cache_mgr
cache_effective_user
visible_hostname

12. Cache Registration Service

cache_announce
announce_to

13. httpd-Accelerator Options

14. Miscellaneous ....

Designrichtlinien

- *Harvest-based* Software (squid) verwenden
- Clients sollen den lokalen (first level) Cache verwenden
- Cache nahe an der Verbindung zum Internet plazieren
- Kooperation mit anderen Cachebetreibern (agreement)
- Logfile handling (anonymisiert, Logrotation)
- Hardware: Linux-PC (64MB Ram, 2-4 GB Harddisk), je nach Anzahl Clients..