XV_Analysis of Design alternatives for reverse proxy cache providers

Exercise on cache, miss, hit etc.
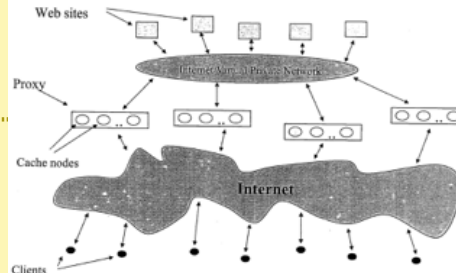
Modeling and evaluation

Reverse proxy cache geographically distributed, organized in a hierarchical manner, have limited customers, are implemented over cluster of workstation and are connected through a VPN to the WS.

Request management

Requests to the Web Server

RAM | disk

Cache node

Proxy

User incoming requests
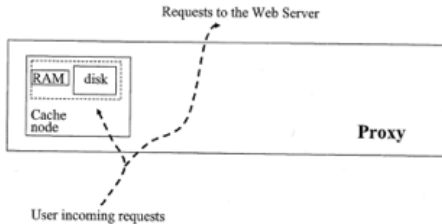
- Proxy configuration: no global memory management
- Cache content defined by access pattern (object popularity)

Architecture

Web sites

Proxy

Cache nodes

Internet

Clients

- Advantages:
  reduce load
  Improve THP
  Reduce latency
  Multiple sites
  share infrastructure.

The model has real constraints:
- bounded cache size and processing power
- popularity and update rates of the documents

It permits identification of archit. depending on:
- Resource assign policy
- Workload characteristics

Also ident of state and transient behavior of the arch.

Nomenclature

WS_k	k-th Web site
C_{WS}^k	total RAM capacity of WS_k
λ_k	arrival rate of HTTP requests to WS_k
n_k	total number of cacheable objects associated with WS_k
α_k	parameter of the Zipf-like distribution associated of WS_k
$p_{k,j}$	relative popularity of the j-th cacheable object of WS_k
$\mu_{k,j}$	update rate of the j-th cacheable object of WS_k
λ_k^{CN}	request arrival rate, associated with WS_k, seen by any single cache node
C^{tot}	total cache node RAM capacity
C_k	cache node RAM capacity destined to cacheable objects of WS_k
MR_k	miss ratio within the cache node RAM/disk for requests associated with WS_k
RHR_k	cache node RAM hit ratio for cacheable objects of WS_k
DHR_k	cache node disk hit ratio for cacheable objects of WS_k
N	total number of Web sites hosted by a cache node
NP	total number of Proxy sites
NCN_k	number of cache nodes within a Proxy site that are assigned to WS_k