# Web Caching
## *Questions and Answers*

*What is a cache?*

A cache is a collection of often-requested data that is stored locally and in such a way that it can be quickly located, identified and retrieved. It represents a subset of a larger collection of data. A cache is usually physically closer to the requester and its contents can be accessed more rapidly than those of the larger archive. Caches are built into most modern microprocessors in order to store frequently used pieces of information such as program code or data. Most web browsers store frequently accessed web pages in memory and/or on local magnetic storage devices. So anyone using a personal computer to surf the Web is relying on multiple caches to perform this task.

*What is a web cache? What is a caching proxy?*

A web cache is a collection of locally stored web pages, embedded images, and other types of digital objects retrieved over the Internet using the hypertext transfer protocol (http). The cache that is maintained by your web browser is a web cache. A web cache proxy server, or caching proxy is essentially a shared web cache. It is a web cache program running on a dedicated server that archives and returns documents frequently requested by a group of web clients. The proxy server returns an object from its archive if it locates a local copy of the object (a hit), or retrieves and archives the requested object if it does not find a local copy. It uses the absolute URL of archived objects to locate and identify hits. By returning local copies of objects whenever possible, a caching proxy can improve response time, reduce network traffic and increase the effective bandwidth available to end-users. Cache servers avoid delivering stale objects and free up space for new objects by expiring objects and by not caching dynamically generated content. Caching proxies typically implement a Least Recently Used (LRU) policy to determine when to remove objects from their archive.

*What is a distributed web cache?*

A distributed web cache is a collection of caching proxy servers that communicate with one another to locate local copies of requested objects. One mechanism for this communication is the Internet Cache Protocol (ICP). Each caching proxy can only store and track a limited number of objects. ICP allows caching proxy servers to request objects from nearby proxy servers or from a hierarchy of servers so that they effectively act as one huge cache.

*What are some other uses for caching proxies?*

Caching proxy servers can be used to restrict access to confidential resources, to block access certain sites or domains, to publish documents and other digital content from behind a firewall, etc. The library has a proxy server that serves as a gateway to restricted resources. Fee-based electronic resources that the library subscribes to are restricted to vt.edu users. The library proxy allows students and faculty who are using computers on other networks to authenticate with their PID and password and then functions as a gateway for authorized users. In practice, this server caches few documents, but it is still

functionally similar to a caching proxy because it requests documents on behalf of an http client.

*What are the benefits of using caching proxies?*
       The use of caching proxies reduces network traffic, speeds up access time for popular documents, and provides a higher effective network bandwidth.

*What are the risks?*
       It is possible that a caching proxy could return an outdated document if its header misrepresented or omitted last modified date information. Dynamically generated content cannot be cached so if the caching proxy resides on an internal intranet between end users and a set of web-accessible databases, those users will see little benefit. In the event of a caching proxy server failure, some browsers are not able to automatically bypass the cache server and must be reconfigured. Occasionally, users may have to use different proxy servers to take advantage of certain services.

*What are other sites doing?*
       The National Laboratory for Applied Network Research is developing a national caching prototype. The prototype is primarily based on squid (a cache server derived from the Harvest project), which supports the ICP protocol and is ideally suited for building distributed caches. Other cache servers that support the ICP protocol can be used as well. NLANR has established a cache registration database to help administrators locate nearby cache servers. Their goal is to facilitate the establishment of a global mesh of cache servers. (http://www.nlanr.net/Cache/Tracker/). Internet Service Providers are using caching proxy software to improve services. There are some national cache meshes setup in Europe.

*How might it work at Virginia Tech?*
       A top level caching proxy server for the university would communicate with lower level (departmental, dorm or lab) servers and with regional and national servers to locate documents. All web clients would be configured to access a local proxy. Autoconfiguration files would be published that could automatically configure popular web browsers to use the nearest web cache and to fail-over to the campus proxy or direct connection if a proxy server were down.

*What proxy caching software is available?*
       Netscape Proxy server is a caching proxy. It can be part of a hierarchy of caches but is unable to dynamically consult a pool of cache servers. Caching proxies that implement the Internet Cache Protocol are a better choice for a large organization. If a server or part of a network fails, ICP easily bypasses the outage. Squid and Jigsaw (a Java-based http1.1 test server developed by the World Wide Web Consortium) implement the ICP protocol.

*Is there a plan?*

The Computer Science Network Research Group has proposed a multi-level caching proxy hierarchy based on squid for Virginia Tech. There will be one server in the computing center designated as the top level VT web cache, called www.proxy.vt.edu. Few browsers will actually be configured to rely on it as their primary cache. Most departments and labs will purchase and install local cache servers that serve as the primary cache for departmental and lab workstations. As cache servers go online, they will publicize their existence to others so that they can query each other for local copies of requested objects. NRG plans to provide modified versions of squid that implements advanced removal algorithms that take in to account not only LRU, but also last modified date, size, and time it took to last download this resource from the remote host. They are evaluating use patterns and hit rates and will be working to further improve the distributed cache service. Distributed Information Systems will publish information about the caching effort with pointers to software, and maintain a cache registry.

*Are there other caching proxy projects underway here on campus?*

There will be at least one special purpose cache proxy based on Netscape proxy, for off-campus users who need to access restricted resources such as Firstsearch or copyrighted courseware materials. This service currently resides on a library server since the library had the most immediate need for such a resource. In order to use the proxy, the user must provide a valid PID and password. This proxy will rarely cache documents, as most requests will be database queries, which are fundamentally uncacheable. This project primarily relies on the ability of a caching proxy to act as a gateway. Soon, this service will be made available on a campus-wide basis by the Computing Center.

University Relations plans to install a few public access web stations around campus for visitors. These stations will be configured to use a restricted caching proxy that will only retrieve vt.edu documents. This service will be implemented using squid and can be integrated with the campus caching proxy hierarchy.

## References

The Cache Now! Campaign. <http://vancouver-webpages.com/CacheNow/>

Johnson, Tommy. A Campus-Wide HTTP Proxy Server Plan.
<http://csgrad.cs.vt.edu/~tjohnson/papers/vtproxy.ps>

National Laboratory for Applied Network Research. A Distributed Testbed for National Information Provisioning. <http://www.nlanr.net/Cache/>

National Laboratory for Applied Network Research. Squid Internet Object Cache.
<http://squid.nlanr.net/Squid/>

Wooster, Roland Peter. Optimizing Response Time, Rather than Hit Rates, of WWW Proxy Caches.
<http://scholar.lib.vt.edu/theses/materials/public/etd-34131420119653540/etd.pdf>

James Powell
Distributed Information Systems
May 15, 1997