

White Paper

LAN-Based Web Caching for Accelerated Web Access

Philip Romine
Director, Cachelink Development
Mangosoft, Inc

Caching Market Overview

A variety of web caching solutions exist on the market today. These solutions fall into several different product categories, which address different market segments.

Internet Service Solutions

Internet service solutions are targeted at businesses that provide Internet services to the masses. The target customers for this class of solutions are Internet content **producers** that include Internet Service Providers (ISPs), Application Service Providers (ASPs), Network Infrastructure Providers, and Web Server owners. The common thread between these different types of companies is that they will use a caching product to improve the experience of people accessing **their** service. Within this market segment, there are three broad types of solutions:

Internet Backbone and ISP accelerators are large scale caching products produced by companies such as Inktomi, Infolibria and Cisco. These solutions cost tens or hundreds of thousands of dollars and are designed for Internet Service providers. Cachelink is complementary to these server side offerings.

Network Services, offered by companies such as Akamai, Digital Island, and Mirror Image, employ content replication and caching technology to servers and networks. These companies use the resulting configurations as a platform to host and sell high performance solutions to Web site owners. Cachelink is not competing with these services.

Web Server Accelerators and Reverse Proxies are products used by Web Server hosts to improve access to their web sites. These products use various technologies, including caching. Novell's ICS software running on Dell and Compaq hardware can be employed as a reverse proxy. The reverse proxy is used to front end one or more web servers, improving the user experience for anyone accessing the web site. Cachelink is currently orthogonal to this technology.



Internet Client solutions

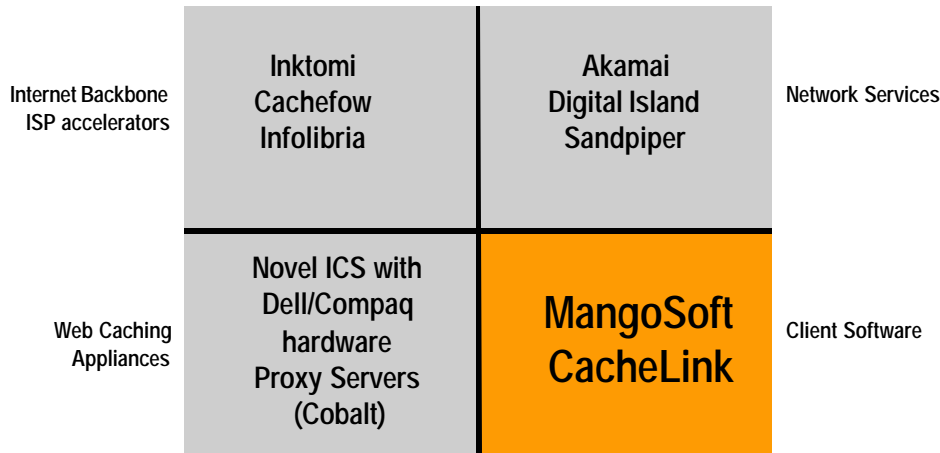
Internet client solutions are a completely different market segment, in that they are targeted at **consumers** of Internet/Intranet content and services. These products use caching to improve the Internet experience for clients accessing arbitrary web sites. The target customers for this class of solutions are much more general – businesses and organizations that use the Internet or Intranet in their daily activities. Within this market segment, there are two types of solutions, with Cachelink being a third, unique product:

Dedicated Cache Appliances are devices that perform the one function of caching Internet/Intranet content for clients on a LAN or LAN segment¹. This is Cachelink's primary competition. Companies competing in this market include Novell ICS software running on Dell and Compaq servers, Network Appliance, NetCache, and Cobalt Networks.

Proxy servers with embedded caches also compete in this market. These can be general-purpose servers, for example a Windows NT server running either Microsoft Proxy Server or Netscape SuiteSpot. They can also be dedicated devices, such as the Compaq Neoserver, whose shared Internet access feature contains a caching proxy server. Proxy servers typically perform many other functions besides content caching, such as Internet access multiplexing, firewalls, etc. – such additional features are not available in dedicated cache appliances.

Cachelink is a software-only solution that provides exceptional scalability, superior fault-tolerance, and

Internet Service Solutions



Client Solutions

tremendous performance at a price that is an order of magnitude lower than the competition.

¹ Note that some of these products also compete as reverse-proxy web server accelerators.



Price / Performance

Caching products generally present their performance in terms of the number of web requests they can serve per second (RPS). The following figures show Cachelink's performance in these terms. However, since Cachelink is a distributed caching service, the number of computers running Cachelink in the pool multiplies its performance – Mangosoftware has tested pools with over 100 PCs.

Mangosoftware's internal performance testing demonstrated a serve rate of 4400 requests per second on a small pool of eight PCs running Cachelink – on average, each PC served 550 requests per second. These PC's had Pentium CPU's ranging in speed from 200MHz to 333MHz.

As a comparison to other products, five fast PCs running Cachelink, for example, would be able to serve 2750 requests per second. The MSRP of Cachelink for five PCs is approximately \$300. According to statistics provided by the Second IRCache Bake-Off, the only known competitive products that can reach this serve rate are the Compaq C2500 and the Cisco CE7300 that sell for \$50,814 and \$130,995, respectively.

As another comparison, the Second IRCache Bake-Off provided an estimate of the serve rate provided by caching solutions per \$1000 investment. The winner in this category served 102 requests per second. Fifty seats of Cachelink can be purchased at MSRP for approximately \$1000, providing an aggregate serve rate of 27,500 requests per second – a value improvement of over 270 times.

Requirements

Computer Hardware and Software Requirements

For a computer to benefit from Cachelink, it must have Cachelink installed on it. Cachelink requires that each computer meet the following requirements:

- Windows 95, 98, NT 4.0, or Windows 2000.

- 20MB of free disk space.

Supported Browsers

During Cachelink installation, Cachelink can automatically configure popular web browsers to use it. Other browsers can be easily configured to use Cachelink as described in the README document included with the beta software. Cachelink can automatically configure the following browsers.

- Microsoft Internet Explorer 4, 5 or 6

- Netscape Navigator 4.0, 4.5, 4.6, or 4.7

Internet Connection Requirements

Cachelink makes use of your existing Internet connection(s) to populate the pooled web cache. Cachelink requires that each computer is connected to the Internet directly via any standard shared



Internet connection including DSL, T1, cable modem, shared LAN modem, etc. Cachelink does not support non-standard TCP/IP environments such as Artisoft I-Share or environments where each individual computer accesses the Internet via its own dedicated modem.

Cachelink caches only HTTP traffic. Other protocols are unaffected by Cachelink.

LAN Requirements

Cachelink benefits LAN segments that have between 2 and 250² computers connected to a single LAN using TCP/IP. Cachelink is compatible with other protocols, such as Novell NetWare, Microsoft Peer-to-Peer Networking, etc.

Technology

Overview

All major browsers, such as Microsoft Internet Explorer and Netscape Navigator, cache web page fragments that are referenced by the user. Mangosoftware's unique Cachelink technology complements the browser cache in each PC by linking them together into a single cache pool for all PCs connected to the LAN. Thus, access from any one system causes a cache copy to be available to all other systems. The effect is a virtual web cache appliance with no single point of failure. Cachelink's innovative distributed systems technology continues to provide cache service even when systems fail or leave the network.

Cachelink is safe, secure, and dependable. It is an anonymous cache. Unlike centralized proxy servers, it does not allow anyone to track what web sites a user visits. Cachelink is also secure and does not cache secure sessions, such as HTTPS. Cachelink is transparent to firewalls and virus protection technology and has no impact on these safeguards.

The system is extremely fast and has been carefully designed to optimize the performance of both cache misses and hits. It allows web page fragments, referred to as *trinkets*³, to be resolved simply and straightforwardly from the local LAN instead of from the Internet or Intranet. Mangosoftware's innovative distributed directory expertise has enabled a highly optimized design with extremely low LAN overhead. As with other web caching solutions, Cachelink adheres to the HTTP-1.1 rules and mechanisms governing content cache-ability and expiration.

Design Objectives

Cachelink's design objectives include:

- ☞ Decrease the need to access the Internet in order to display a web page or trinket that has already been displayed by another user.

² Mangosoftware is not aware of any scaling issues with the Cachelink product. However, it is possible to split large configurations into different Cachelink sharing groups. This is accomplished via the Cachelink Configuration program's advanced tab. Each group must have its own HTCLP port number assigned.

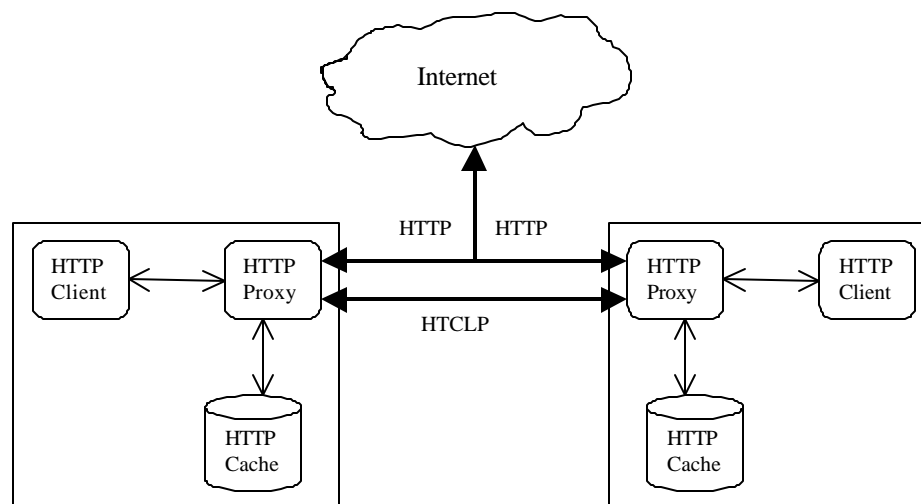
³ Trinket is the Cachelink name for a web page fragment, a component item of a web page that may be cacheable.



- ☞ Decrease the Internet access network bandwidth used by users who share the same access mechanism. This may have the beneficial side effect of decreasing Internet access costs in some cases. At the very least, it allows the saved bandwidth to be used by other Internet access applications.
- ☞ Increase the speed of displaying a web page.
- ☞ Increase the effectiveness of cache space across all users in an organization. That is, by decreasing cache item duplication, the space saved can be used by other cacheable items. This means that more data may be cached in the same amount of total cache space.
- ☞ Do not impact the performance or latency of cache misses.
- ☞ Cache data where it is referenced. Because a web page will most likely be re-fetched by the same user, Cachelink places trinkets closest to the user that originally requested them.
- ☞ Maintain user privacy.
- ☞ Do not interfere with existing network management policy or tools such as firewalls, filters, network monitoring tools, etc.

Cachelink Architecture

The following diagram shows two machines that are configured to share their HTTP caches. Although only two machines are shown, the design is scalable to any number of cooperating machines on the local area network.





Each machine is outfitted with Cachelink's **Proxy Agent**, an HTTP proxy server. Each HTTP client (e.g. Microsoft Internet Explorer and Netscape Navigator) application on the machine is configured to access the Internet via its Proxy Agent.

The Proxy Agent is the heart of the Cachelink product and performs many functions. It makes use of a Mangosoft proprietary protocol called the **Hypertext Cachelink Protocol (HTCLP)** to query cooperating proxies for the existence of specific cached data. It also uses HTCLP to retrieve a cached entry from a cooperating, local proxy. This protocol binds the computers on a LAN into a Cachelink network, allowing cached trinkets to be accessed in a similar manner from either the Internet or another cooperating node in the pool.

The Proxy Agent maintains a local Trinket Store and Index (TSI) using a persistent hash table mechanism. This replaces the browser's local cache during the installation and configuration of the Cachelink product. The TSI maps an URL to the associated trinket meta-data and data stream (file or file segment). The Proxy Agent is also responsible for searching its TSI for a trinket specified in an HTTP request forwarded to it from the browser client. If the trinket is found and not expired, it will be served back to the browser using normal HTTP response messages. If the trinket is not found, the Proxy Agent employs a proprietary multi-level search algorithm to attempt to locate and retrieve the trinket from another pooled node's cache. Note that each level of search result is subsequently cached, utilizing the pooling principal of performance based data migration. This innovative search algorithm is designed to efficiently determine if a trinket is cached and, if so, where it is located, without the overhead of a centralized directory.

For further information on this product, contact:

MangoSoft Inc.
1500 West Park Drive, Suite 190
Westborough, MA 01580
Tel: 508-871-7300
Fax: 508-871-7380
Email: info@mangoSoft.com

? 2000 Mangosoft Inc. All rights reserved.

Mango is a trademark of Mangosoft Inc. Cachelink is a registered trademark of Mangosoft Inc. Microsoft, Windows, Internet Explorer and other Microsoft products referenced are either trade marks or registered trademarks of Microsoft Corporation. Netscape and its products are either trademarks or registered trademarks of Netscape Communications Corporation. All other marks and names are the property of their respective owners, and no association or affiliation between Mangosoft Corporation and these companies is intended or implied.



Mangosoft Inc. 1500 West Park Drive Suite 190 Westborough, MA 01581
TEL: 888-88MANGO/508-871-7300 FAX: 508-871-7351 www.mangosoft.com