



A ViSolve White Paper

# **Optimized Bandwidth + Secured Access = Accelerated Data Delivery**

**Web Caching - A cost effective approach for organizations to address  
all types of bandwidth management challenges**

**March 2009**



**Disclaimer:**

The information in this document is believed to be accurate in all aspects, to the best of ViSolve's knowledge, at the time of publication and is subject to change without prior notice.

ViSolve Inc. is not legally liable for any errors or opinions that appear in this document



## Table of Contents

Executive Summary.....	2
Bandwidth Management Options .....	3
What is Web Caching? .....	4
Need for Web Caching .....	4
How does Caching work? .....	5
Squid – Web Proxy Caching Solution .....	7
Squid Feature Highlights.....	9
Squid Vs Proprietor Appliances .....	10
Business Case Study .....	12
Cost-Benefit Analysis .....	13
About ViSolve.....	14



## Executive Summary

What is one problem that is common to most organizations - ISPs, Educational Institutions, Corporate and anyone, for that matter, who has an IT data centre? – Bandwidth Management.

Effective bandwidth management is more important now, than ever, because availability of information is taken for granted. As more and more people start using internet, as more and more data are shared via internet, as more and more enterprises conduct business via online, the world of www has grown out of mathematical numbers, literally. Communications networks are growing more congested as consumers increase wireless usage and as companies generate and transport more data. As this bandwidth grows increasingly congested, service providers are constantly seeking relief. Carriers need innovative ways to break the logjams.

Bandwidth has become a rare commodity these days with more and more vendors fighting to gain supremacy over each other – to provide internet service to their customer/clients. With limited bandwidth and more traffic (number of requests sent by an end-user at any given point of time), the experience of browsing the web is fast deteriorating.

ISPs are struggling to meet bandwidth requirements while expanding their customer base. Educational institutions and enterprises are struggling to manage bandwidth and streamline data flow for right activities. So, is there an answer to all these bandwidth related issues? - Yes, there is – Web Caching



## **Bandwidth Management Options**

There are quite a few ways, depending upon specific business needs, to effectively manage bandwidth –

### **Saving Bandwidth**

Bandwidth could be easily saved by restricting/allocating access based on demand and for appropriate business needs. This initiative can directly bring in significant cost savings.

### **Traffic Management**

Most organizations struggle to manage data congestion on their networks which results in decreased productivity as bandwidth are often utilized for non-business usage like downloading music/videos or sharing files among employees.

### **Secured Content Management**

Ethical web usage not only reflects on the corporate governance aspect of an organization but it also plays a very vital role in business productivity – blacklisting unwanted URLs - and to create a congenial work environment. More importantly, this helps to protect data identify theft to a large extend.

### **Fast Content Delivery**

Managing benefits is not all about bringing benefits to the service provider. In hindsight, it massively improves the browsing response time for the end-user which could easily serve as a value proposition for the service provider, especially in the current competitive market landscape.

Almost all of these actions can be achieved by deploying a suitable caching solution.



## What is Web Caching?

In simple terms, Web caching is a technology that can significantly enhance end-user's web browsing experience and, at the same time, save bandwidth for service providers. In detail, a web cache is a temporary storage place for data content requested from the Internet. After an original request for data has been successfully fulfilled, and that data has been stored in the cache, further requests for those files (e.g., HTML pages, images) results in the information being returned from the cache, if certain conditions are met, rather than the original location.

Web Caching is the widely used technique, used by Internet Service Providers (ISPs) all around the world, to save bandwidth and to improve user response time. In short, web caching temporarily stores web objects – HTTP and FTP data – flowing into ISP's network.

This is not an entirely new invention in that Caching is an integral part of computer architecture, for example CPU cache speeds up an access to main memory, file system cache stores commonly requested blocks for faster access and so on.

## Need for Web Caching

We'll start with a simple analogy. This analogy is not a perfect match but it will give you the basic idea. Each morning before John goes to get the daily paper he asks his roommate, Bill, if he has already purchased one. If Bill already has the paper there is no reason for John to walk all the way to the store and spend money on the exact same information. He'll just read the paper that's already there. If a copy of the daily paper was already retrieved by Bill and is on hand, John saves his money and his time by not making a trip to the store.



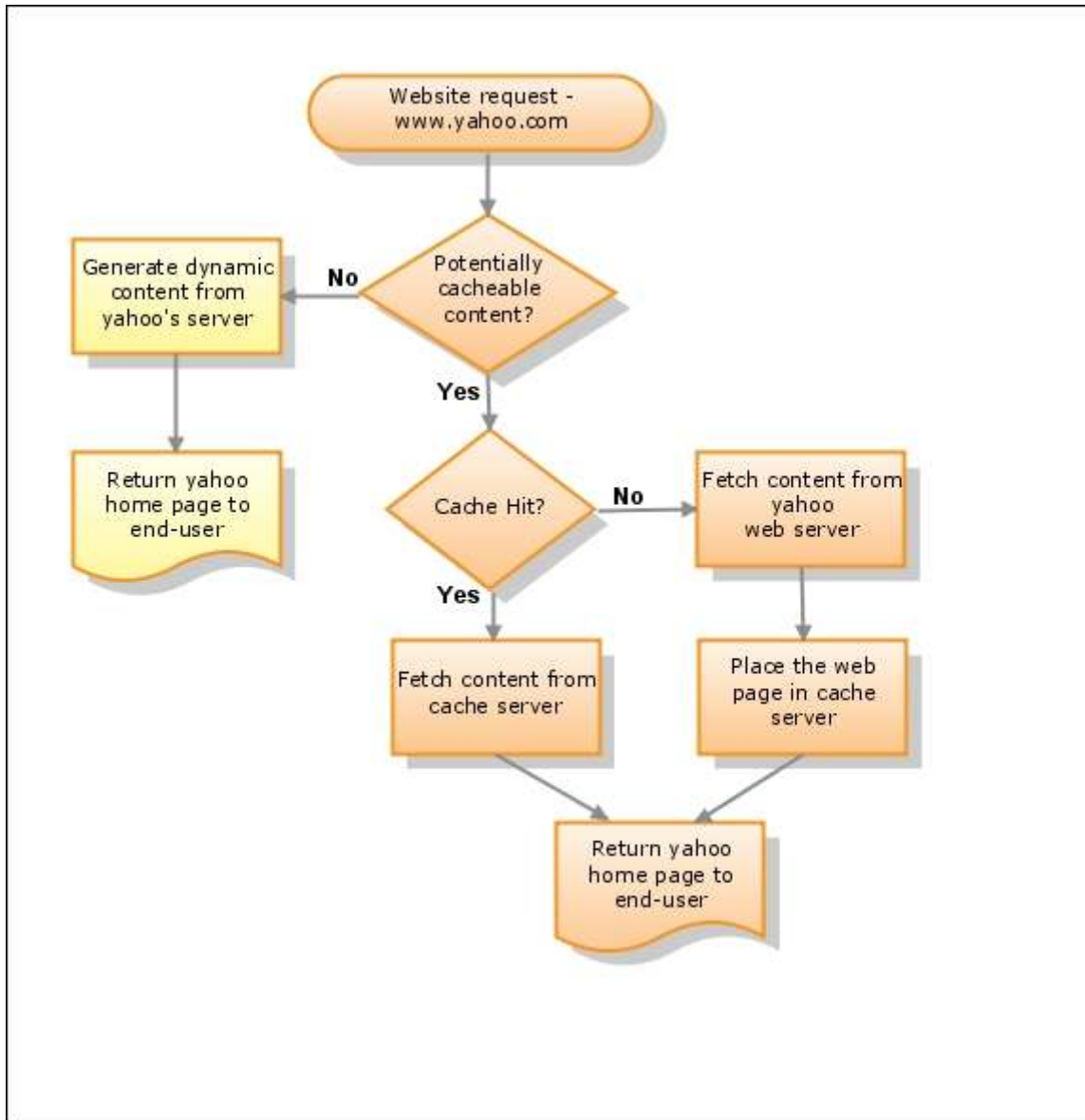
Web caching enhances web browsing in much the same way. When a user visits a site, say [www.yahoo.com](http://www.yahoo.com), web caching (if in use and available) will retrieve the page from yahoo's web server and store a copy of that page locally – in cache server. The next time a user requests [www.yahoo.com](http://www.yahoo.com) the web cache delivers the locally cached copy of the page (without fetching it from yahoo's web server). The user will experience a very fast download because the request did not have to traverse the entire Internet – all the way to where yahoo server is located. Also, the bandwidth that would normally be used to download the web site is not required and is free for other information retrieval or delivery.

## How does Caching work?

Web sites are continually updating their content. News headlines change, stock quotes change, weather changes. It may seem that caching is not worthwhile if it is returning outdated material. A traffic report that is two hours old doesn't do you much good. Fortunately there are checks and balances in place to ensure that the content you are viewing is current.

Web sites are made up of many small pieces that come together to make a complete page. A site might have logos, photographs, tables, text, and sounds. Each item will be cached as a different object, and some items may not cache at all. For example, when you access CNN.com frequently your cache may hang on to the CNN logo object, some advertising bars, and the rest of the stuff that makes up the basic look of the CNN Web site. But the news items will not sit in cache because they change so often. In this case your cache has made the CNN site much easier and faster to download because all the static graphics are already on hand and the only thing you need to complete the picture is the news content.

The following depicts how caching works from an end-user perspective.



So how does your cache know what to hang on to and what to let go? That depends on choices made by the Web developer as well as the way the user configures his cache. As mentioned above, Web sites are made up of individual pieces. Each one of these pieces is encoded with specific information that will tell your cache how to handle it. This information may specify,





“Don't cache this item,” in which case the cache will ignore it. The item may have a “max age” specified. This tells the cache that after a set amount of time the cache must check in with the web site for newer versions of that object. The “expires” field serves roughly the same purpose. The item might also have a “last modified” field. Last modified is another way for your cache to ask the Web server if the object has been modified since your last visit. If it has, the cache gets a new copy, if not the cache just hangs on to the copy it already has. The web site administrator controls each of these items. There are many cache products available. Each has lots of different configuration options to help ensure that your data is current. Caches can be configured to accept all, some, or none of the priorities that the web site administrator sets.

## **Squid – Web Proxy Caching Solution**

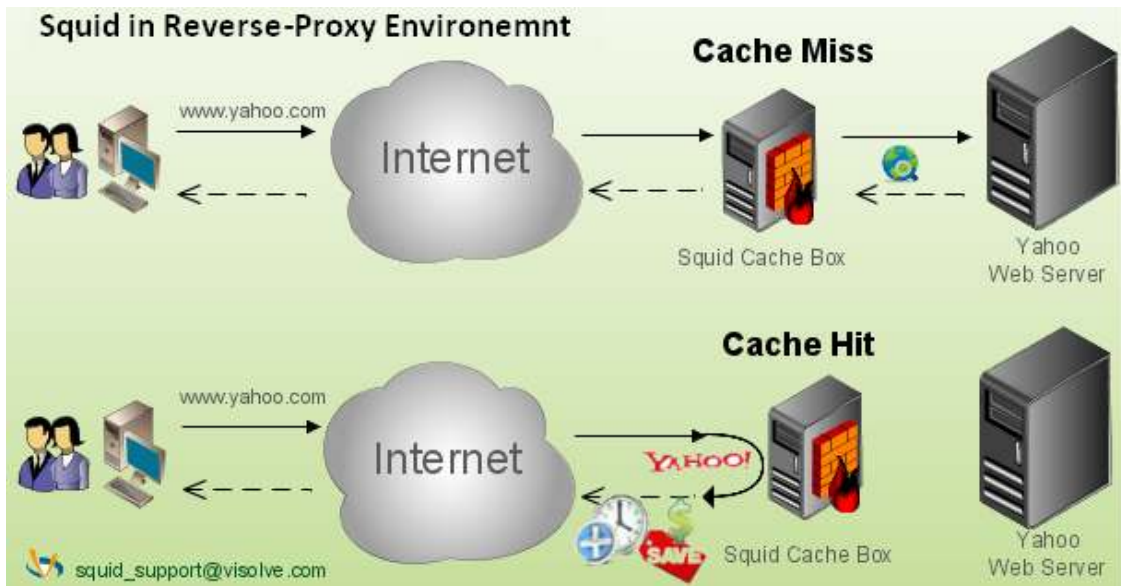
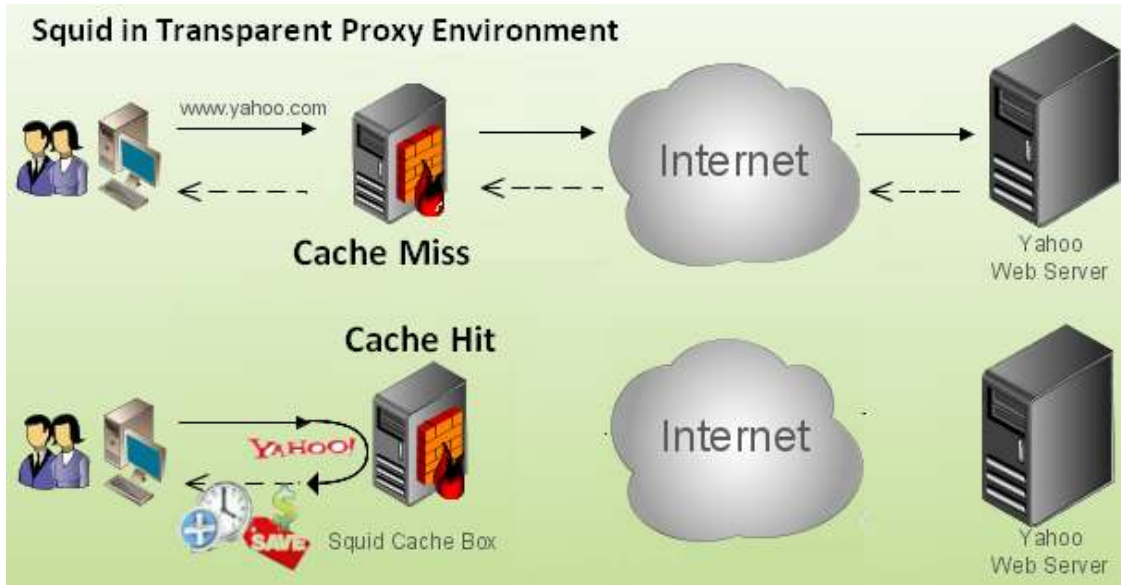
Squid is a typical example of a high-caliber product from the ever growing open source community. Squid is one of the two forks from the codebase of harvest cache daemon developed in early 90's. Squid is a fully-featured HTTP/1.0 (almost HTTP/1.1) proxy complaint offering rich access control, authorization and logging environment to develop web proxy and content serving applications.

- For ISPs – to save bandwidth and improve user experience
- For websites – Helps to scale applications without any massive investment in hardware and development time
- Helps content delivery providers to distribute content world-wide

For detailed technical information on how to implement Squid for different network environments, please visit <http://www.visolve.com/squid/whitepapers/index.php>

Optimized Bandwidth + Secured Access = Accelerate Content Delivery

The following picture depicts how Squid is deployed in transparent proxy environment (generally for ISPs) and in reverse proxy environment (for dedicated web servers).





## Squid Feature Highlights

In addition to Caching, Squid provides so many value-added features that compliment its core functionality so wonderfully that it provides a wholesome package to the deployment environment.

- Blacklisting URLs that are not related to business affairs
- Managing bandwidth for critical business needs based on IP address and time interval
- Implementing data security by integrating various authentication mechanisms
- Easier integration of different tools like content filter and anti-virus via ICAP

To understand this in detail, let us do a virtual case study for a school district and an IT organization where Squid is deployed primarily for web caching.

### **For School Districts –**

Most developed nations treat Cybercrime is a serious offence – especially in educational premises. The advent of internet is a great boon to the education system all over the world as it brought information as and when required and significantly changes the entire landscape of education. One of the most important factors of internet usage in educational premises is the nature of content students gets exposed to. Protecting students from unhealthy and unethical usage of internet is a mandatory, not optional, duty for all educational institutions.

Squid brings this functionality by default. With Squid one can easily blacklist URLs such as gambling, terrorism, religion, pornography, sports, etc., that are not meant for educational activities, thereby securing the use of internet.



### **For Corporate -**

We all know the significance of the role played by internet specifically in a corporate environment. Almost every employee uses internet 90% of his/her office hours. But strangely several industry surveys, conducted all over the world, reveal that only 10% of total internet usage is productive. What this means is, though internet plays a very vital role in sharing critical information and triggering innovative ideas needed for functioning of corporate, there is an equal downside that many organizations either tend to underestimate or do not know how to manage/tackle those unproductive exercises. And this is exactly where Squid come in very handy.

In general, for a corporate data center, Squid is usually deployed in reverse proxy mode to reduce traffic on the web server by delivering cached content. But there are other vital benefits also. With Squid, an IT manager can decide on the amount of bandwidth to be allocated for individual or user groups based on organization's policies. And what's more - bandwidth delivery can be managed based on time, IP address or even MAC address of a computer. For example, Squid allows you to restrict only 20% of bandwidth to employees at the time of a critical online business meeting thereby allowing 80% of bandwidth for that meeting. Also, Squid helps to integrate an organization's centralized authentication mechanism to achieve secured and authorized content delivery. Going one step further, Squid is not all about completely restricting employee freedom. If used intelligently, Squid can also motivate employees for better productivity, say; allowing certain websites like Sports, News, Entertainment, etc., during lunch or non-business hours.



## Squid Vs Proprietor Appliances

Squid has serious advantages over other proprietary caching solutions and appliances

- It is open source. It is free to deploy and use. No license is required
- Very low Total Cost of Ownership (TCO) compared with proprietary appliances
- Active and Stable community with developers all around the world contributing critical ideas and state-of-the-art features
- Can be customized as per changing business requirements as the source code is available for customization
- Can be deployed with almost all standard operating systems available in the industry – unlike appliances which requires a tailored operating system
- This is a one-time investment and there is no recurring cost
- When there is change in the underlying network architecture, Squid is both adoptable and versatile enough to fit into network changes without any complexity. And, most importantly, no need to make new investments each time there is a change in the network setup
- Squid, unlike proprietary appliances, is easily compatible with so many related open source products which can add significant value to a scalable organization, moving forward, to add appropriate plug-ins
- Reliability and Sustainability. There were so many proprietary caching appliance vendors who went out-of-business between 1999 to 2002 time frame when entire caching market took a hit, but Squid survived all that.



## Business Case Study

Data Telecom Services is the largest ISP in Madagascar. DTS has been providing internet service since 1996 through its X.25 link to France, has X.25 packet switching service (InfoPac) in 10 major towns in the country.

DTS has satellite links, which has more delay than leased line, for the Internet Backbone to Europe with customer profiles being xDSL, wireless (Wimax, CDMA, GSM), and also Dialup (PSTN) users. With the number of concurrent users in cache manager being 1800 clients and about 250 req/s, a web caching solution is needed as a web accelerator to serve a whole whack of traffic. Web Browsing takes about 68% of total bandwidth which requires a quick installation of a web caching solution that is functionally sound and cost effective.

The following are the business requirements of DTS –

- Deploying a caching mechanism that will save precious bandwidth and increase user response time
- Clustering cache servers with load balancing and fail-over mechanisms
- GUI based reporting tool to monitor web usage and bandwidth management

As the key endeavor is to save HTTP bandwidth usage and build a reliable failover mechanism, ViSolve recommend deploying a customized version of Squid – recompiling Squid in the underlying operating system to retain client’s original IP address – called as Tproxy (IP spoofing) to meet DTS’s business requirements.

## Cost-Benefit Analysis

Business Case – Data Telecom Services is the largest ISP in Madagascar. DTS serves its customers with a total of 60MB leased line bandwidth – 68% of it accounts to HTTP traffic.

Now, lets calculate the ROI metrics for Squid deployment for DTS

Cost of 60 MB Bandwidth per year for DTS =  $60 \times \$10000 = \$600000$

DTS utilizes 68% of bandwidth for HTTP traffic. So, cost of HTTP bandwidth } = \$408000 (40.8 MB)

Total Cost of Ownership (TCO) Items –

Hardware cost for 2 cache servers = \$6000

Initial deployment cost for 2 servers = \$4000

Annual support contract for 2 servers = \$10000 per year

-----  
Total = \$20000  
-----

The minimum industry average of total HTTP bandwidth saving for Squid } = 20%

So, minimum bandwidth saved for DTS = 8.16 MB (20% of 40.8 MB)

Therefore, minimum total cost saved =  $8.16 \times \$10000$   
= \$81600

Return On Investment = Total Cost Saved – TCO for Caching  
= \$81600 – \$20000  
= \$61600

**Therefore, the bandwidth cost benefit for DTS after Squid deployment is a whopping \$61600**



## About ViSolve

ViSolve is an open source based software consulting and support organization having operations in San Jose (USA) and Coimbatore (India). Established in 1995, ViSolve has made significant contribution for open source community, in general, and Squid, in particular, by assisting developers to help fix security vulnerabilities and testing. ViSolve is an acknowledged provider of Squid configuration manuals and whitepapers on various mode of Squid deployment thereby helping millions of online users searching for initial help to deploy Squid. For more details on ViSolve, please visit <http://www.visolve.com/index.php>

ViSolve is a distinct leader in providing commercial Squid support for the past 10 years and it ranks #1 in almost all search spiders for Squid related keyword searches. ViSolve, over the years, has undertaken several complex and mission critical Squid deployments all over the world – including few FORTUNE 100 organization. For more details related to Squid support, please visit <http://www.visolve.com/squid/index.php>

ViSolve has developed GUI based easy-to-use intuitive products based on Squid with an objective to enrich end-user experience of configuring Squid such a joyous activity. With ViCache+Filter, Corporates, SMBs, ISPs and Educational Institutions can see an average of 50% increase in network performance, enable customized web content delivery at the edge of the network (closer to the user) and could take the load off the high traffic Internet backbone. For more details on ViCache+Filter, please visit <http://www.visolve.com/vicache+filter/index.php>