



A Websense® White Paper

Proxy: Not a Product, Just a Feature

Putting Proxy In Its Place

Web proxies have been around nearly as long as the Web itself. In fact, for many years proxy caches were purchased and implemented as separate, standalone products. The subsequent evaporation of this discrete market segment, however, is indicative of the steadily eroding value of proxy technology. Due to the increasingly dynamic nature of Web content, caching is no longer as effective as it once was. In addition, Web proxy functionality has essentially become a commodity. Paying a premium for it, therefore, no longer makes any sense. IT organizations should instead focus their evaluations and investments on the value-added services that have come to use the Web proxy as a common operating platform. Ultimately, it is these higher-order capabilities and technologies that:

- Enforce acceptable use policies to reduce corporate liability and ensure productive use of Web-based resources;
- Protect PCs and the corporate computing environment from Web-based threats; and,
- Minimize corporate and compliance oriented risks due to data loss.

No Cash Warranted for Caching

A primary driver behind the once successful Web proxy market was its ability to cache Web content. The concept involved is a relatively simple one: by locally storing and serving Web pages and the individual objects that comprise them, proxy caches can improve Web performance in three distinct ways. First, latency is effectively reduced and the end-user experience enhanced because Web content is now closer to the clients that are requesting it. Second, serving content from a cache cuts the load on an organization's Internet circuits. This reduces the amount of bandwidth and associated infrastructure – such as routers, switches, firewalls, and intrusion prevention systems – that needs to be purchased and maintained. Finally, serving content in this way also reduces the service demands placed on the destination Web sites, not to mention the Internet at large.

It's important to realize, however, that these are only *potential* gains. The extent to which they are achieved actually depends on several applicable factors.

Limitations Have Been Present From the Outset

To begin with, the effectiveness of a Web proxy is fundamentally limited by:

- **The relatively small size of available caches.** Compared to the overall volume of content available on the Web – which continues to grow exponentially – no cache can ever be considered large. The impact is that some amount of content that could otherwise be served at some point is always being purged to make room for new/other content
- **Content that is inherently unable to be cached.** Requests that fall into this category include those where some sort of authorization is required or where cookies are present (although the latter is less of an issue with HTTP 1.1 which, unlike HTTP 1.0, facilitates the storage of user-specific content). Another substantial contributor to this category is dynamic content. This includes not only content that is changing constantly and, therefore is always unique, but also content that is changing so frequently as to make caching pointless. Notably, a fair amount of such content is present even in the absence of Web 2.0 services and technologies. These include, for example, Active Server Pages (ASP), Perl scripts, and Java Servlets.

- **Content that is selectively unable to be cached.** This refers to the fact that developers can take advantage of Cache-Control and Expires header fields within HTTP either to: (a) block specific content from being cached all together or (b) explicitly prescribe and therefore significantly shorten the duration that content is allowed to reside in a cache before it must be refreshed. Banner ads are a classic and very common example of content that falls into this category. In addition, a variation of this approach is also possible whereby client-side settings can be used to dictate that requested content be freshly obtained from the source provider.

The result is that even in a Web 1.0 world, cache hit rates were typically never greater than 35 to 50%. Furthermore, the reduction in bandwidth utilization that could be achieved was fractionally lower than these levels due to the combination of imposed refresh requirements and corresponding “adaptive refresh” feature sets.

The point is not that this was a poor outcome. Indeed, given what Internet services cost fifteen, ten, or even just five years ago, a bandwidth reduction of a third or more represented considerable savings – savings that could easily be used to justify investment in a standalone Web proxy solution. Rather the goal is to establish a baseline for the effectiveness of proxy caching and the factors that impact it in order to provide a foundation from which we can better comprehend and measure what came next.

Impact of the Dynamic Web on Proxy Technology

Based on the limitations identified above, the current chapter in the history of the Web proxy is not all that surprising. Simply put, the meteoric rise of Web 2.0 and the corresponding proliferation of dynamic content over the last handful of years have substantially undermined the benefits of proxy caching. Relevant considerations in this regard include the following:

- More than 80% of the top 100 most frequently visited websites utilize Web 2.0 techniques and technologies and offer content that is primarily dynamic in nature. Overall, it is estimated that the number of such sites has increased 10-fold in the past 5 years, driving a commensurate growth in the percentage of Web traffic that is inherently incapable of being cached.*
- Both users and developers are continuing – if not actually increasing – the practice of exploiting mechanisms at their disposal to control if and for how long an object can be cached. Indeed, many modern websites are going to great lengths to prevent Web caching on the basis that it corrupts metrics that are important to their business. This is especially true if the metrics involved factor into calculations of the revenue these companies are entitled to receive from their customers. Moreover, attempts by users to leverage so-called “advanced” proxy functionality to override these mechanisms is rarely worth the effort. Doing so requires non-stop adjustments to proxy settings to keep up with frequent site modifications and, even then, still results in a significant potential for: encountering proxy-site incompatibilities, serving stale content, negatively impacting the user experience, and increased help desk calls.
- The types of content that are inherently not able to be cached and the amount of traffic they represent continue to rise. One example is live video feeds. Another is the traffic associated with many software-as-a-service (SaaS) sites. Because the applications being hosted typically involve a lot of unique, user-generated transactions, the traffic being generated is not all that repetitive.

The net result is that cache hit rates and related reductions in bandwidth have declined precipitously over the past few years. Hit rates below 15 percent are now fairly common. Furthermore, this downward trend is unlikely to abate as the Web and the content that comprises it continue to evolve. The outcome is that proxy caching and, in turn, the need for a dedicated Web proxy, is no longer compelling.

Of course this is not meant to convey that bandwidth reductions of 20, 15, or even 10 percent are not worthwhile. Today's CIOs will gladly take whatever gains they can get. This conclusion only holds true, though, if the cost to realize the gain remains low relative to the gain itself. And herein lies the catch. The economics only work in this case if Web proxy capabilities entail little to no cost. In contrast, proxy products priced at a premium will not be worth the investment.

Fortunately, most solution providers understand this situation and have accounted for it in their product strategies. This is demonstrated by the fact that they only provide Web proxy capabilities as an embedded, zero-cost, zero-premium component of broader, multi-function solutions, such as a Web security gateway. From a practical perspective, this approach is the only one that is appropriate given that the Web proxy has essentially become a commodity. The associated capabilities, whether we're talking about caching or just acting as intermediary between source and destination, are simply too basic and too widely available to command the attention of IT departments – never mind a premium price tag.

What About WAN Optimization?

Another justification that is occasionally put forth for pricing proxy products at a premium is that they can also be used for WAN optimization. This claim is somewhat misleading though. The limitations identified earlier apply equally in this case as well – not to mention that WAN circuits are typically used to convey a much wider array of traffic types, all of which ideally require a measure of attention. This is why providers of WAN optimization solutions tend to incorporate much more than ordinary, object-based caching capabilities in their products. Greater emphasis is typically placed on byte-level caching, other advanced techniques, and application-specific optimization mechanisms – which is exactly where IT departments in the market for a WAN optimization solution need to focus their attention as well. When what's needed is WAN optimization, then the ideal approach is to purchase a corresponding solution from a WAN optimization vendor, not a proxy vendor.

The Real Role of Proxy

Given the steady demise of proxy caching it is natural to consider what other capabilities a Web proxy has to offer and, subsequently, what role it's suited to play for enterprise IT going forward.

Proxy is a Feature

Besides caching, another core capability of a Web proxy has to do with the “proxy” part of its name. Once again, the concept involved is fairly straightforward: being a proxy entails acting as an intermediary between a client (i.e., the source) and the Web-based resource with which it would like to communicate (i.e., the destination). Requests initiated by the source are actually terminated by the proxy which then connects to the destination on behalf of the client. Responses from the destination are handled in the same manner, just in reverse.

The main advantage inherent to this arrangement is that it helps maintain the privacy of user identities and internal addressing schemes. When used in conjunction with an authentication service, native or otherwise, it also provides a rudimentary method for controlling access to the Internet.

What needs to be acknowledged in this case is that “proxy” is indicative of little more than an architectural characteristic of a product; it's merely one of a handful of fundamental techniques for handling network traffic. Referring to it as a technology even seems unwarranted, at least from the perspective of how basic it is and the limited value that it *directly* provides.

Value is Derived from the Services that Ride on Top

As with Web caching, IT organizations need to treat the proxy component of solutions that incorporate a Web proxy as a core capability; it's a commodity that deserves neither significant attention nor any amount of monetary investment. Resources should instead be reserved for other, value-added services and technologies that take advantage of platforms featuring a proxy architecture to actually solve specific IT and business-oriented problems. In this regard, although Web 2.0 and the dynamic Web diminish the requirement for Web caching, they actually elevate and expand the need for protection of the Web channel of communications.

With the proliferation of Web 2.0 methods, mechanisms, and technologies, ordinary URL filtering and antivirus software are no longer sufficient. The highly dynamic nature of the Web now requires traditional countermeasures that rely on prior identification and classification of sites, content, and associated threats be supplemented with real-time content analysis, categorization, and security scanning capabilities. Furthermore, because Web 2.0 has transformed the Web from a "search-and-you-shall-receive" medium into a channel for bi-directional communications, it is now necessary to address the potential of the Web channel as an avenue for the unwanted exposure of confidential data.

An appropriate solution in this case is a Web Security Gateway, a product that uses a proxy-based platform to deliver and enable a comprehensive set of Web protection capabilities and technologies, such as:

- Real-time content inspection and classification technologies, used in conjunction with comprehensive reputation analysis techniques and a well-maintained URL database for URL-based filtering decisions;
- The ability to block "bad" elements from a site while still enabling good information to be displayed;
- Real-time anti-malware filtering that is based on multiple detection techniques and analysis engines;
- Advanced application controls for granularly tracking and limiting the use of specific protocols and services (e.g., IM and P2P), as well as individual functions within those services (e.g., file transfer);
- Comprehensive visibility and control over the flow of not just entire documents and attachments but individual pieces of data too;
- The ability to monitor Web activity and set usage and security policies on a per-user basis; and
- The ability to inspect and control SSL-encrypted traffic.

The stop-and-go nature of a proxy architecture does yield a significant advantage in this case. It affords these countermeasures the opportunity not only to conduct the necessary inspections but also to directly respond in a meaningful manner in the event that a misuse or other type of threat is detected. Web sessions can be blocked outright if that action is deemed appropriate. Alternately, individual elements can be modified in virtually any manner, thereby neutralizing the risk without having to completely impede otherwise legitimate activities and communications.

It should be clear, however, that what really matters is the scope and strength of the countermeasures themselves. Without these there is no Web security. Other characteristics are important as well of course, such as manageability, reliability, and scalability of the overall solution. But the fact that there is a Web proxy at its core is, for all intents and purposes, is a secondary or even tertiary consideration.

Beware of the Hardware Trap

Speaking of scalability, one final point concerns the need, or, more accurately lack thereof, to run a proxy-based solution on specialized hardware. Some vendors will point to high-end appliances, often bristling with an array of custom chips/processors, as another way to justify charging a premium for their products. For the most part, however, such systems are neither necessary nor advantageous.

As noted previously, proxying is a straightforward process. With the possible exception of handling cryptographic operations, there is little need for custom components. Moreover, rapid and continuous advances in processing technology and associated hardware almost always ensure that any gains obtained with a high-end appliance are short-lived. The unfortunate result is that customers of such products inevitably find themselves chained to an under-performing solution – they can't take advantage of state-of-the-art alternatives because they are financially locked-in to their original investment.

To be clear, IT departments with an affinity for appliances and the benefits they yield should still pursue this approach. It's just that appliances based on specialized hardware are rarely worth the premium price tags that accompany them.

Proxy is a Feature, Not a Product

The greatly diminished effectiveness of Web caching due to Web 2.0 and the proliferation of dynamic content provides the final confirmation of what the marketplace has already established: a Web proxy is a feature, not a product. Caching and the other underlying capabilities of a Web proxy – which are centered around acting as an intermediary between a client and the Web-based resource it's trying to access – still have a role to play. But they are no longer compelling in their own right. What matters more these days are the value-added services that rely on a proxy platform/architecture, such as those required by enterprises to establish adequate Web and Web 2.0 defenses. Accordingly, it is on these higher-order capabilities and technologies that IT departments should focus their attention and investments, as opposed to the underlying plumbing.

In this regard, the Websense Web Security Gateway is an ideal solution. By combining Web proxy as both a feature and the foundation for a full complement of market-leading technologies, the Websense Web Security Gateway delivers unparalleled visibility and control over the Web channel of communications along with comprehensive protection against associated threats.

For more information on Websense and the Websense Web Security Gateway, visit www.websense.com.