Blue Coat

and

LinchPin NETWORKS

Practical Strategies to Accelerate Business Applications Across the WAN

www.linchpinnetworks.co.uk
info@linchpinnetworks.co.uk
01284 830 841

EXECUTIVE SUMMARY

More business than ever is done in branch offices and remote sites, but fewer and fewer IT resources (applications, personnel, servers, etc.) are hosted there. Applications are being consolidated (sometimes centralized, sometimes outsourced) – cost is a driver, but compliance is the catalyst.

In many cases, a consolidated application translates to a poorly performing application. Long distances between users and applications, skinny/latency-prone network pipes, and applications and protocols stretched beyond their design limits mean poor application performance at remote sites. These issues are exacerbated by the introduction of additional bandwidth-hungry, latency-sensitive applications such as voice-over-IP (VoIP) and video.

The industry response is predictable – accelerate the traffic. While appropriate at a high level, the rush to a solution has left out some important questions – e.g., should everything be accelerated? If not, which applications are key? How are they evolving? What about encryption? How are networks evolving? Does MPLS or going directly to the Internet from remote sites change things? Given the significant deployment efforts associated with rolling out acceleration technology, these questions merit consideration. Any solution under consideration to accelerate applications should be considered with the answers to the above questions in mind.

Acceleration technologies range from compression to caching, to bandwidth management and protocol optimization. All of these techniques have benefits, but for a given application, some improve performance more then others. Each user and application combination has an ideal set of acceleration techniques – apply the wrong techniques and performance benefits can be nullified.

Enterprises (both private and public sector) need all of the techniques mentioned above for the array of applications deemed important to the business (file services, e-mail, web, secure web, video)– but what about the countless "applications" that run on the enterprise network that are not business-related, or worse – harmful to the business? Given that ˜30% of enterprise network bandwidth is consumed by unauthorized applications (web advertisements, inappropriate web surfing, P2P, Skype, spyware, etc.), removing the undesirable can be as important as accelerating the desirable.

Blue Coat's MACH^5 is a patent-pending framework of technologies designed to bring all of the appropriate acceleration techniques (bandwidth management, compression, protocol optimization, byte caching, and object caching) to all of an enterprise's key applications – including web, secure web, file services, email, and video. MACH^5 is delivered in Blue Coat appliances, which means that it enables organizations to manage all of their user/application interactions – to stop undesirable applications, throttle less-important applications, differentiate users and groups, and accelerate critical applications – even when encrypted.

SOLVING BUSINESS PROBLEMS, BUT CREATING APPLICATION PERFORMANCE ISSUES

For a variety of reasons (cost, customer intimacy, agility, focus), organizations continue to push people and offices out of headquarters and other large sites – meaning that fewer people are centralized, and greater numbers of personnel are out in remote sites. Unfortunately, efforts to reduce IT operations costs and comply with regulations, policies, and covenants spawn a variety of IT consolidation efforts – pulling most IT resources back into a select handful of sites (HQ, data centers, outsourcers/hosters). Given that bandwidth is always getting cheaper, one would think this isn't a problem, but several factors crop up to create application performance issues. First, the number of applications enterprises use continues to grow – some of which (e.g., VoIP, video) are very demanding of the network. Second, the distances between users and applications are quite large – large enough that the speed of light becomes a limiting factor when many round trips between user and application are involved. Third – many of the applications in question are built on protocols designed for LANs, and are very chatty – i.e., require lots of round trips between users and applications. Fourth, users continue to adopt flashy applications that have nothing to do with the enterprise's business, and may have significant network impact. Finally, while more bandwidth is available at lower prices than ever, many parts of the world are underserved by network providers. Even in the developed world, private WAN circuits remain far more expensive than broadband Internet. All of these factors mean that most enterprises have serious performance problems (both capacity and latency) with interactions between remote users and consolidated applications.

Different applications have different problems; some have latency issues (e.g., file services), some have bandwidth issues (e.g., some multimedia training/video), and some have both bandwidth and latency issues (e.g., web applications). Furthermore, different remote sites may have different network limitations (some long distances, some skinny/overburdened pipes, some both), and different groups/users have different roles and priorities within the organization. All of this adds up to one answer: enterprises need a variety of acceleration methods at their disposal, and may use different sets of methods for each combination of user, location, and application. Moreover, enterprises are not standing still – applications continue to evolve and develop at a rapid pace, and the mix of protocols moving across enterprise networks changes monthly. Two trends are emerging, however: the webification and outsourcing of enterprise applications, and the resulting encryption of that application traffic. Organizations continue to web-enable their applications, and are often looking to host part or all of these applications with a third party. Examples of third-party hosted applications include Salesforce.com, WebEx, Oracle OnDemand, CaseCentral, Digital Insight, and portions of the HR/benefits portal – e.g., retirement plan providers (Fidelity Investments), health insurance providers (Aetna), and payroll providers (ADP). Any acceleration solution under consideration should accelerate web applications, and increasingly, both internal and outsourced SSL-encrypted web applications.

Finally, networks are evolving too – enterprises are re-examining their WAN strategies, with some choosing to leverage the Internet (and site-to-site VPN) heavily. While many large organizations use site-to-site VPNs over the Internet in a fashion similar to the way private circuits work (i.e., backhaul to headquarters), by 2009, 50% of branch offices will connect directly to the Internet (Mark Fabbi, Gartner, February 2006). This means organizations will implement a split tunnel, where only the traffic destined for headquarters (or the datacenter) will move over the VPN, while the Internet traffic will go directly to the Internet.

ENTERPRISES HAVE OPTIONS

Organizations have several options to address these issues. First, they can upgrade their WAN circuits to higher-bandwidth links. Upgrading WAN circuits is expensive, in some parts of the world impossible, and doesn't address the latency issue. Second, enterprises can ignore these performance issues – resulting in employee dissatisfaction, and more importantly, broken applications, impeded business processes, and potential non-compliance with various regulations. Third, organizations can keep applications distributed – which would result in higher operations costs, and likely non-compliance with regulations, business agreements/covenants (e.g., VISA CISP/AIS), and privacy and security policies.

There is another option: accelerate the traffic. Not surprisingly, this is the option many organizations are now examining. With the array of options available – e.g., compression, caching, bandwidth shaping – it is difficult to compare different options. Moreover, different applications (e.g., e-mail, file services, web applications, e-learning applications) have different issues (latency, bandwidth) and therefore respond differently to different acceleration techniques.

When examining acceleration technologies, enterprises should take into account the rapid evolution of applications and networks – and incorporate all of their critical applications and network changes into their thinking.

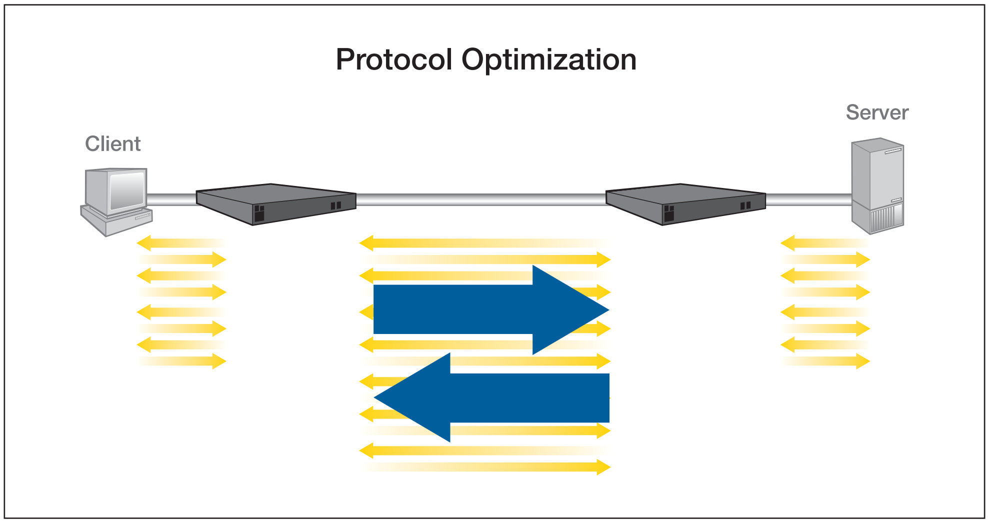
Acceleration techniques available include bandwidth management/traffic shaping, compression, protocol optimization, byte caching/dictionary compression, and object caching.

Bandwidth Management/Traffic Shaping

This technique assigns a priority to a particular application's (or user's) traffic. This priority has an effect both on the order the traffic is sent in, and in the amount of bandwidth the traffic is afforded. While this technique doesn't make traffic go any faster on the network, it does ensure that the network is available first for the highest priority traffic.

Protocol Optimization

Protocol optimization takes protocols that are inefficient over the WAN (e.g., CIFS, MAPI, HTTP, TCP, HTTPS) and makes them more efficient – typically by parallelizing traditionally serialized communications. There are other optimizations, depending on the protocol (e.g., TCP session reuse) that can make starting up/tearing down flows faster. These optimizations do not reduce the amount of bandwidth an application consumes, but can greatly accelerate applications (i.e., reduce latency) – the longer the WAN link, the greater the improvement. For example, protocol optimizations have an enormous impact over satellite links.



Byte Caching/Dictionary Compression

Byte caching is as it sounds – a low-level cache of small, sub-application-object pieces of information. Typically, byte caching/dictionary compression schemes observe repetitive patterns moving between two caches in application traffic, symbolize those patterns with a token, and send the token in lieu of the bulky traffic – tokens being typically a byte or two, symbolizing large blocks (e.g., 64KB). The cache on the far end matches the token with the original block of data, reconstitutes the traffic, and sends it on to the application or user (whichever is appropriate). Byte caching/dictionary compression is typically not application-specific, and operates at a lower level, reducing bandwidth of all TCP traffic. It does, however, have its limitations: it can never reduce bandwidth as much as object caching (because some data must be transmitted), doesn't reduce latency much, and doesn't offload source servers.

Object Caching

Object caching is very different than byte caching – in that it is protocol/application specific, and is an all-or-nothing affair. If the cache contains the object, the user is served the object from a local store – extremely quickly. Object caching, on a "cache hit" (which is where the object has been through the cache and then stored) reduces greatly the bandwidth used, and the latency – both to almost zero. If the cache does not contain the object (or contains an outdated version of the object), then for that particular transaction, object caching does nothing (although the next time that object is requested, it will be fast).

Compression

Compression uses a common compression algorithm (e.g., gzip, lz compression) to remove extraneous/predictable information from the traffic before it is transmitted. The information is reconstituted at the destination based on that same algorithm. It is important to note that there is no synchronization between the two ends – and the first time something goes through is just as fast as the second. This technique reduces the data transmitted over the WAN link, but has limitations on how much bandwidth reduction it can achieve by itself – and has minimal impact on latency.

ENTERPRISE APPLICATION ACCELERATION EVALUATION CRITERIA

Organizations should have a strategy for evaluating different acceleration approaches and techniques. Overall, enterprises should first prioritize applications needing acceleration. Second, organizations should examine each application (and its parent initiative) in detail, to assess the faults within that application, understand the best acceleration techniques for that application, and how to best bring those techniques to bear. Third, organizations should examine the trends, both in the application mix (e.g., web, in-house vs. third party hosted, HTTPS), and in networking architecture. Fourth, organizations should establish non-application specific overall solution criteria – deployability, manageability, reliability, and solution breadth/extensibility. Briefly, an enterprise set of criteria might look like this:

Enterprise Application-specific Criteria

- Application/protocol – File services, e-mail, e-learning/multimedia, web, secure/encrypted web, in-house or hosted – does it support the enterprise's key applications?
- Does the solution address the application(s) specific problem – network capacity and/or latency?
- Does the solution support all of the appropriate techniques for that application – bandwidth management, protocol optimization, object caching, byte caching, compression – and does it understand the application well enough to ignore inappropriate acceleration techniques?

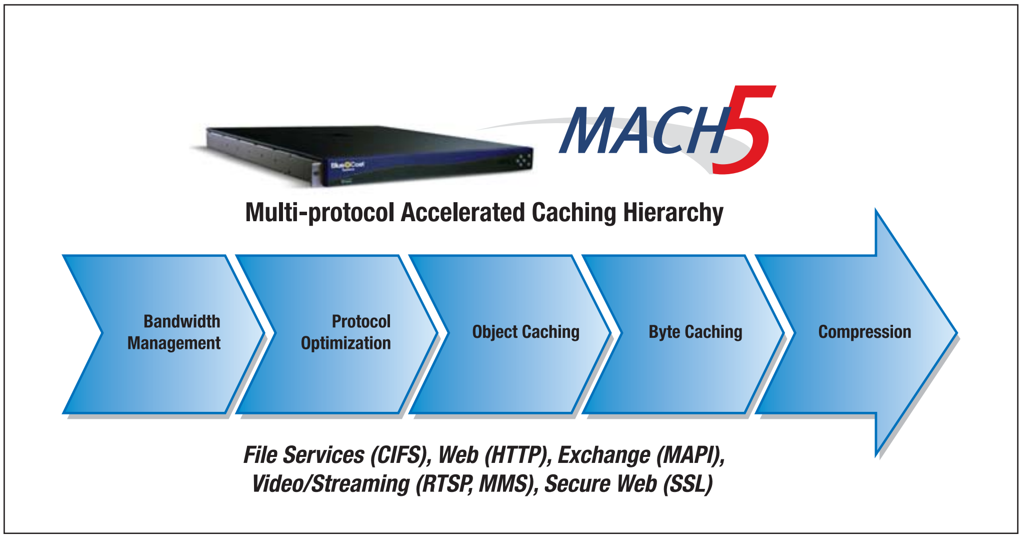
Enterprise Solution Criteria

- Solution scalability – does the solution scale down (are there appropriate form factors for small sites) as well as up (are there form factors for very large sites)? What about centralized management?
- Solution breadth/adaptability – does the solution support all key applications? Does the solution require additional components be rolled out to remote sites?
- Solution Investment Protection – when the enterprise rolls out new applications, will the solution accelerate it? Prioritize and optimize it? When the network changes, is the solution still effective?

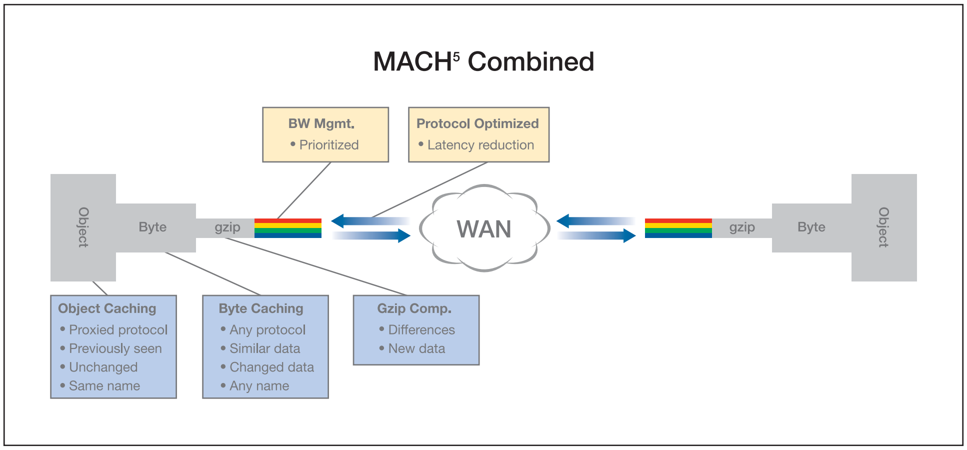
BLUE COAT MACH^5 – AN INTELLIGENT ACCELERATION FRAMEWORK

Blue Coat started out life as CacheFlow – an extremely successful web application caching company, one whose core technology was designed to accelerate applications. In 2002, CacheFlow changed its name to Blue Coat Systems – reflecting a change in focus toward security and control, but the underlying technology (and its application acceleration capabilities) remained. Organizations implementing Blue Coat appliances for increased security and control discovered an interesting dynamic – rare in the industry – of a single technology increasing security and performance at the same time. Most security technologies impede performance, and most performance technologies impede security. As organizations' infrastructure priorities have shifted over time (between performance-focused and security-focused), Blue Coat's core technology has remained the same. As enterprises (and their users) have continued to add applications, Blue Coat has continued to enhance the control over both security and performance for these new applications.

At the core, Blue Coat appliances use a proxy/cache architecture, which enables two key functions – visibility of user/application interaction, and control of that interaction. With visibility comes context and understanding, and with control comes the ability to accelerate desirable applications, while limiting the impact of undesirable applications. Blue Coat has assembled all of its acceleration capabilities into a framework called MACH^5 – the 5-part Multiprotocol Accelerated Caching Hierarchy.


MACH5
Multi-protocol Accelerated Caching Hierarchy
Bandwidth Management
Protocol Optimization
Object Caching
Byte Caching
Compression
File Services (CIFS), Web (HTTP), Exchange (MAPI),
Video/Streaming (RTSP, MMS), Secure Web (SSL)

MACH^5 incorporates all of the techniques available to the industry today – bandwidth management, protocol optimization, object caching, byte caching, and compression, and uses them in an integrated fashion. With MACH^5, all of these techniques work together – which is very powerful. For example, if the object cache contains an outdated copy of a document, the byte caching capability has patterns and tokens that require only the tokens, plus the changes to be sent. What little is sent is then compressed, and protocol optimized (reducing bandwidth consumed and latency/round trips). All of this is prioritized according the enterprise's preferences, such that the important applications get through first, with the bandwidth they need.



Blue Coat's native, proxy-based ability to understand user/application interaction ensures that the solution uses the right combination of acceleration techniques for that user, that application, and that interaction – even if the interaction is encrypted. Furthermore, the aforementioned ability to control that user/application interaction ensures that only valid users, interacting with valid applications, consume infrastructure resources.

There are hazards to applying inappropriate acceleration techniques to certain applications – i.e., the wrong technique for a particular application can either slow that application down, or do nothing for the application in question, while impacting other applications negatively – for example:

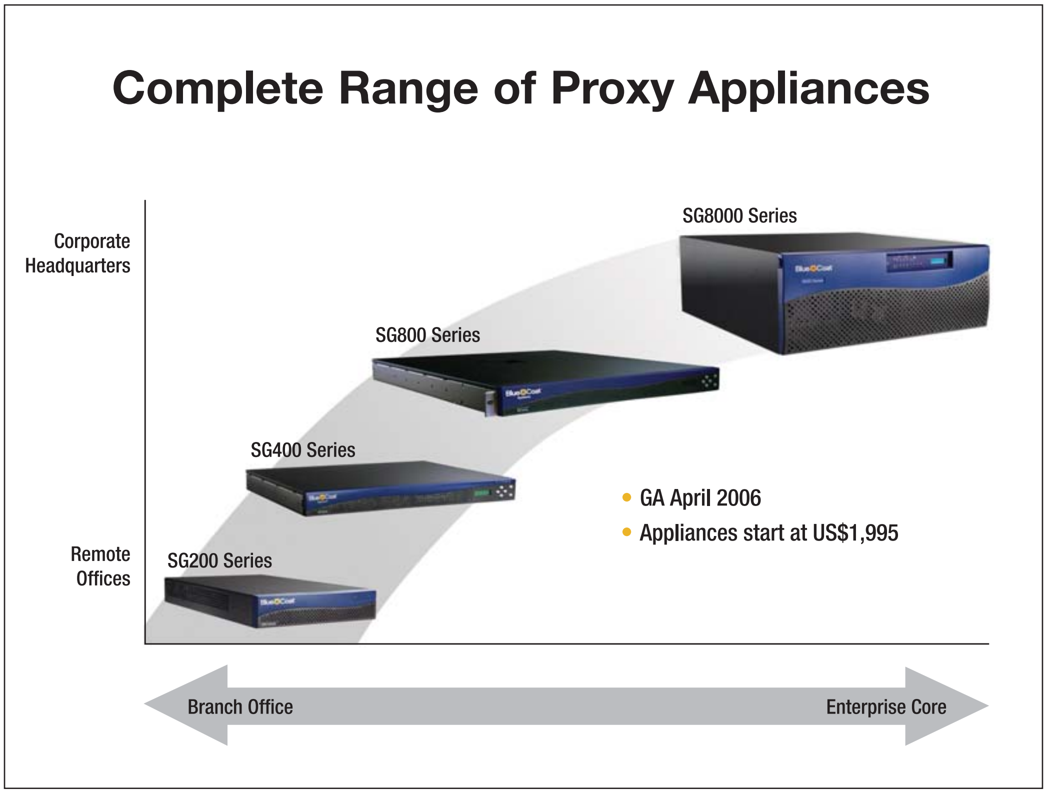
> With byte caching – when sending large amounts of non-repeating (e.g., streaming applications or OS or application patches) the data cannot be "compressed", and can force valuable information from other applications out of the cache. Effective solutions (like Blue Coat) will understand what the application is, apply appropriate techniques only, and avoid spoiling the byte cache with it.
>
> Like Blue Coat, any effective solution has to be able to adjust what and when you apply byte caching. Another example is a web (HTTP) application that supports gzip compression. It may or may not be a good idea to also use byte caching. For example, web pages that do not change can utilize both http compression and byte caching, but a more dynamic app can not use byte caching if http compression is enabled – gzip uses a stateful encoding mechanism, so if one byte changes in the original object, all bytes after that may be different in the gzipped version of the file.
>
> Blue Coat's SSL support is more important than it initially seems – it goes beyond just being able to accelerate HTTPS. Solutions that can't control all types of SSL traffic must pass port 443 traffic through unexamined and unaccelerated. This has two issues beyond the obvious – first, lots of bandwidth-hungry unauthorized or rogue applications hide there (e.g., Skype, P2P file sharing); and second, people within the organization often throw up internal SSL-encrypted applications on non-standard ports. Those have to be hunted down and routed around the byte caching to make the solution work.

Blue Coat accelerates two key application areas that few others can – secure web applications and video/streaming. Secure web application (SSL-encrypted HTTP) traffic makes up a growing percentage of enterprise network traffic (30%+ today, and growing) and comprises both internally (e.g., financials, HR) and externally (e.g., HR, collaborative, sales automation) controlled applications. Blue Coat's solution accelerates both internally and externally controlled secure web applications, including the ability to cache HTTPS-encrypted objects, while maintaining and enforcing an organization's privacy policy. Please see the "Gaining Visibility and Control of "Inside-Out" SSL Web Sessions" whitepaper for more detail on Blue Coat's ability to ensure privacy policy compliance while managing encrypted content.

Live (streaming) video and video on-demand is a growing area for most organizations, principally for e-learning/training reasons. Most organizations with a high degree of employee churn, and/or regulatory oversight have significant (and expensive) training needs. Many are turning to video for those needs, but quickly run into network capacity issues. Blue Coat appliances have been used in many enterprise networks to accelerate both streaming and video on demand applications, applying both object caching and protocol optimization (e.g., adaptive refresh, stream splitting) to accelerate the user experience and offload the network and servers.

BLUE COAT – PRODUCTION-TESTED

Blue Coat's Appliances are available in many different sizes – from the 200 series (appropriate for small remote offices), to the 8000 series (appropriate for the enterprise core), and everything in-between. Large deployments of Blue Coat appliances exist all over the world – accelerating applications and securing organizations at the same time. Blue Coat's reporting and centralized management capabilities – Blue Coat Reporter and Director, respectively – have proven themselves across hundreds of Blue Coat appliances in some of the largest organizations in the world.



CONCLUSION –
ENTERPRISE SOLUTIONS MUST ACCELERATE ALL KEY APPLICATIONS

In summary, organizations must examine all of their key applications when evaluating application acceleration solutions – file services, e-mail/Exchange, streaming/video, web, and secure web. Furthermore, organizations should look ahead – to what applications are coming, and how networks are evolving – given the large-scale nature of deploying remote office acceleration capabilities. Blue Coat's MACH^5 application acceleration technology draws on ten years of industry-leading application performance experience – accelerating all key enterprise applications (even encrypted application) – while maintaining control of non-critical applications.