A Riverbed Technology White Paper

# Five Ugly Truths about WAFS and Caching

## Why caching can't meet the enterprise application acceleration challenge

riverbed™

# FIVE UGLY TRUTHS ABOUT WAFS AND CACHING

## Introduction: The Enterprise Application Acceleration Challenge

A cache by any other name… is still a cache.

Whether they're referred to as Wide Area File Systems (WAFS), Web caches, or Web application accelerators, the fact of the matter is that caches take a myopic view to an important enterprise challenge.

That challenge is to end the tug-of-war in which IT managers are stuck in the middle. On one hand, distributed employees need full, fast application access. But on the other, management wants to consolidate IT infrastructure in order to reduce expenses and meet regulatory compliance. How can an organization balance these needs? This challenge spans almost all applications that organizations use, from common end-user applications such as MS Office, file sharing and document management down to less visible administrative applications such as backup and replication, software distribution, and database synchronization.

*Caching, in fact, is a poorly architected, partial solution to enterprise application acceleration that may actually introduce more problems than benefits into a network.*

Caching vendors claim that they solve these problems in one, easy-to-use device. They claim that their devices enable enterprise application acceleration, branch office IT consolidation, and bandwidth savings as well.

Caching, in fact, is a poorly architected, partial solution to enterprise application acceleration that may introduce more problems than benefits into a network. Caching devices, as we will explore in detail, *seemingly* improve throughput by short-circuiting the communication processes of applications and serving content locally. But this approach is limited in its efficacy, error-prone, and hard to manage.

This paper illuminates the 5 most dangerous shortcomings in the caching approach to enterprise application acceleration:

- Lack of breadth for application support
- Inability to handle common changes to information
- Data integrity weakness
- Inability to support disaster recovery
- Failure to really consolidate servers

Finally, this paper will present the Riverbed Steelhead appliance as an alternative to caches. By first understanding the shortcomings of caching, it will be easy to see why the Steelhead appliance is a more intelligent approach to meeting the application acceleration challenge. By the end of this paper, it will be easy to see how Steelhead appliances encompass the limited benefits of caching, but also give an enterprise much more acceleration power, breadth, flexibility, and manageability.

Read on to understand the ugly truths behind caching products and the troubles they can bring to your network. If you'd like to hear the real story behind caching, watch caching vendors squirm as they have to admit to these 5 deal-breaking limitations:

## Ugly Truth #1: "You're likely to need multiple caches in each of your offices."

Caches are built by implementing a local server that supports a particular protocol. Because of this, a cache's functionality is limited to the protocol that the server supports. Caches then try to store local copies of data, and wherever possible, serve a client request from its local copy as opposed to requesting information across the WAN.

For example, a web cache understands HTTP and can locally serve HTTP requests if it has a copy of the relevant web objects. However, that same cache is useless when user requests are in a file system protocol, mail protocol, database protocol, or the other important protocols that enterprise applications use. Likewise, a file cache (a.k.a. a WAFS device) may be able to serve some of the file requests it receives, but it can do nothing useful with email or web requests. If an organization wants to use caches to accelerate four different kinds of traffic (say CIFS, SQL, MAPI, and HTTP), they would need up to four caches and could need up to four devices in each office! How does that qualify as simplifying branch office IT infrastructure?

*Organizations may require up to 4 different caches in each office! How does that qualify as simplifying branch office IT infrastructure?*

Even caching products that claim to support more than one application often consist of multiple servers implemented on the same machine. Often times, the servers need to be configured and managed separately, with their own set of constraints and management tools. Adding a new protocol to the cache involves the development of a new server for that protocol, including engineering the necessary coordination of data and actions between the local server and the origin server. Not only can this dramatically reduce the performance of the "multi-protocol" cache, but it will most likely make the cache *even harder* to manage and configure.

## Ugly Truth #2: "If users make changes to files, caches are not very useful."

Do your users ever change the names of the files that they work on? Or edit files? If not, please skip this section. If, however, your users are like most, they probably make a lot of changes in the course of a day. "Presentation.ppt" becomes "Presentation_v2.ppt" and "Finances.xls" becomes "Finances- FINAL.xls." Sound familiar?

Well, those names *don't* sound familiar to caches, and this means that caches will not effectively support the way most people are accustomed to collaborating over via the WAN. Caching systems use the file as their object type, and the name of the file is the tag by which caches recognize it. When a user changes "Overview.doc" to "Overview – edited.doc" caches no longer recognize that this is data that has previously moved across the network. If a file name changes -- despite the fact that little or no data in the actual document may have changed -- caches can no longer optimize the transfer. If your organization is willing to use an iron fist in order to require users to never change file names, perhaps a cache would be a good fit.

> If a file name changes -- despite the fact that little or no data in the actual document may have changed – caches can no longer optimize the transfer.

Caches often support pre-positioning operations that allow content to be moved into the cache in advance of any requests. Unfortunately, a cache's approach to pre-positioning is not very effective either. Again, this is due to the fact that caches fail to optimize transfers when file names change or a small section of the data changes. These two scenarios explain the impacts of pre-positioning with caches:

### Best Case Scenario

Content is pre-positioned out to the edge, but soon afterwards the content on the origin server changes slightly. On receiving a client's request for the file, the cache recognizes that the origin file has changed, and has to request the file over the WAN again.
*Result of having a cache: No benefit*

### Worst Case Scenario

Content is pre-positioned out to the edge, but soon afterwards the content on the origin server changes slightly. On receiving a client's request for the file, the cache *does not* recognize that the origin file has changed, and serves old content. Users are now working on the wrong version of data.
*Result of having a cache: Lost productivity*

Caches simply are not designed to support users in real-world environments. What is the point of purchasing and deploying caching servers that can't even handle standard operations like file name changes and edits to documents?

## Ugly Truth #3: "Caches can interfere with basic data coherency."

Caches attempt to reduce WAN utilization by serving a local copy of data instead of requesting the original data from the server. A cache then uses its local data copy and its local server (see Ugly Truth #1) in order to spoof the client into thinking that it is actually interacting with the real (origin) server. Whenever this spoofing is successful, the client's request is effectively *absorbed* by the cache, taking the origin server out of the loop.

This act of absorption is an important characteristic of caches, and makes a cache an unpredictable component of your network infrastructure. Every time a cache serves data locally, there is the risk that the action the cache takes is different than what the origin server would have done if it had handled the request. The origin server – including its responsibilities to manage file permissions and file locking – is no longer a gateway to ensure security and data coherency for end users. Is a cache in a branch office really smart enough to take over the origin server's responsibilities?

> *Every time a cache serves content locally, there is the risk that the action the cache takes is different than what the origin server would have done if it had handled the request.*

 

Cache vendors will often highlight their ability to have both dramatic performance improvements, and to have one authoritative version of a file. They tend to gloss over the fact that, in a caching system, these two benefits are mutually exclusive. In environments where access control is essential, WAN disruptions take place, or where collaboration takes place on important shared data, caches add another risk to data coherency.

Some caching vendors say their proprietary technology has enabled both full coherency and simultaneous data availability, even when the WAN goes down. Such a combination is impossible, as has been proven by researchers at MIT and Berkeley. Riverbed's white paper "There's No Free Lunch with Distributed Data" explains their results.

With such a wide array of error-inducing possibilities, it is no wonder that caching vendors do their best to gloss over topics related to data integrity, and instead they focus on "dramatic" lab results created under contrived circumstances.

## Ugly Truth #4: "Caches can't accelerate backup and replication."

Organizations are putting a significant amount of time and resources into enabling fast, effective replication, business continuity, and disaster recovery processes. For disaster recovery or data replication, data needs to be rapidly moved from one site to another, but latency-induced effects at the TCP level often limit the effective throughput possible for any given WAN link.

With data center replication, these processes often include high-bandwidth links (45Mb/s to 1Gb/s) connecting datacenters a considerable distance apart. Because of the distance, transmissions on these high-bandwidth links have very high latency. With a round-trip time of 100ms, it is common for even an OC-12 network connection (with 622 Mb/s bandwidth) to only transfer about 5 Mb/s on a single TCP connection.

How do caches deal with this situation? They don't deal with it at all. Caches function by short-circuiting the communication process and serving content locally whenever possible. Rather than improving the speed with which data is immediately moved across the network, a cache will buffer data locally and move it gradually to the remote site over time. When the goal is to create a remote copy in case of a disaster, this quiet substitution of a local copy is worse than useless.

In addition, caches do little, if anything, to optimize TCP and the application protocols used to move information across the WAN. That means, for requests that a cache cannot serve locally, the data will not be optimized to move across the high bandwidth, high latency WAN connection efficiently.

When the goal is to create a remote copy in case of a disaster, this quiet substitution of a local copy is worse than useless.

When you think about it, it's easy to see that caches aren't really acceleration devices. They are just faking it with additional copies of data!

## Ugly Truth #5: "It takes a lot of work to implement and maintain a cache."

Here's an experiment to try: Go to your IT people, order them to take out each remote file server, and tell them to replace it with a different remote file server. Should you expect dramatic IT management improvements? Well, that's what caching vendors would like their customers to do. Instead of creating a system that eases manageability, caches require enterprises to replace a file server with a new file server that has the same set of management challenges as any other common server.

Caches require enterprises to replace a file server with a new server that has the same set of management challenges as any other common server.

Additionally, depending on the caching product, the user may have to reconfigure clients or servers in order to use the cache. This is just an additional level of complexity that caches may introduce to a network. Is that a chance you can take with your network?

But the real cost of management in a caching deployment is the opportunity cost. In implementing a caching system, an enterprise is giving up the opportunity to implement a system that would truly simplify branch office IT infrastructure.

## The Riverbed Steelhead Appliance: A Better Solution

The Steelhead appliance takes a widely applicable, more scalable architectural approach to enterprise application acceleration. Based on the Riverbed Optimization System (RiOS), the Steelhead appliance is designed from the ground up to provide enterprises with one, easy to manage device that can accelerate all key applications that run over the wide area network.

RiOS was created to solve application acceleration challenges in a very different way than caches. Caching was created as protocol-specific architecture, essentially only dealing with data in the "application silo" that they understand. RiOS, on the other hand, accelerates applications on three levels simultaneously:

1. Data Streamlining: Data Reduction for All TCP Applications
2. Transport Streamlining: TCP Optimizations for All Applications
3. Application Streamlining: Application-Specific Optimizations

Each of these approaches happens independently in the Riverbed Optimization System, meaning that all enterprise applications can benefit from data reduction and transport layer acceleration. Application layer acceleration is just treated as one piece of the puzzle in this architecture, while in the caching architecture it is a requirement that the cache understand the application protocol. The application-independent optimizations in RiOS mean that email, file sharing, document management, ERP applications, CAD applications, network-based backup, software distribution, web-based applications, and even custom-built applications see benefits.

The result of this approach enables massive acceleration for *all* applications that run over TCP – users see up to 100 times faster application speed and up to 95% less bandwidth utilization at the same time. The system is designed to intelligently accelerate applications while not creating the management problems that caches have created in today's networks.

### One Device for all Important Applications

The Steelhead appliance is designed to accelerate applications at multiple levels simultaneously. Steelhead appliances can eliminate data redundancies independent of the application, meaning that all applications that run on top of TCP will see massive reductions in the amount of data needed to be sent across the WAN when Steelhead appliance is deployed.

Data Streamlining functions by breaking up data into small application-independent segments, on average 100 bytes in size, and storing them in the Steelhead appliance on both sides of the network. Subsequent data transfers, regardless of the application, are segmented and compared to data already on disk for repeating data patterns. Matching segments do not need to be transferred across the network, only new segments (from completely new files or edited sections of existing files) need to be transferred. For most organizations, the end result of Data Streamlining is that 60 – 95% of data previously sent across the network no longer needs to be sent across the WAN. (For more information on Data Streamlining, download the RiOS Technical Overview from Riverbed.)

If an organization requested a cache vendor to support a new protocol, the vendor would have to redesign its cache, possibly including an additional server on the device with its own set of configuration and management tools.

Additionally, Steelhead appliances streamline the TCP protocol in order to reduce the number of round-trips required to send this reduced amount of data across the WAN. With Steelhead appliances, no additional steps are required to accelerate a new TCP-based protocol, and TCP-based optimization is immediately available any new protocol the customer chooses to use.

In contrast, if an organization requested a cache vendor to support a new protocol, the vendor would have to redesign its cache, possibly including an additional server on the device with its own set of configuration and management tools.

### Designed to Work the way Your Users Work

Steelhead appliances look at information on the byte level, instead of the file level, which enables the architecture to share data redundancies across different versions of files, different applications, and different protocols. The Data Streamlining process (described above) stores data in a way that is usable by any application.

The impact of this is far reaching. For example - your nightly backup windows are shortened by the fact that people send email attachments during the day; Access to your document management server can be accelerated by file sharing, even if it uses a different application protocol. And when users change file names or edit content, Steelhead appliances ensure that only their edits are sent across the WAN. Access is still accelerated.

When users change file names or edit content, Steelhead appliances ensure that only their edits are sent across the WAN. Access is still accelerated.

 

Caching's alternative? Tell all your users to never change file names.

## Maintains Data Coherency

A cache tries to spoof the origin server with a local server, and take over responsibilities such as permissions and file locking. Steelheads, on the other hand, let origin servers do their jobs. With Riverbed Steelheads in a network, origin servers are still responsible for deciding what data can be sent across WAN. Steelheads are just responsible for *accelerating* the resulting transfer. Using this approach, a Steelhead never introduces the coherency problems that a cache can introduce into your environment.

These advantages are especially important when dealing with complex situations such as people in multiple offices collaborating and modifying documents, situations with changing permissions, and environments with potential WAN disruptions. Steelhead appliances even have an integrated feature called Proxy File Services, which allow for certain, controlled data access in the event of a WAN disruption. Unlike disconnected operations in cache environments, the Steelhead appliance Proxy File Service capability is fully controlled by the IT manager, meaning that the Steelhead is not deciding which users can access particular data.

Steelhead appliances are just responsible for *accelerating* the resulting transfer. Using this approach, a Steelhead never introduces the coherency problems that a cache can introduce into your environment.

## Enables Data Center Replication

Because Steelhead appliances combine data optimization with transport optimizations, they can be powerful tools in the disaster recovery process. Recall that caches do not do anything to overcome the throughput limitations cause by latency. With a round-trip time of 100ms, it is common for even an OC-12 network connection (with 622 Mb/s bandwidth) to only transfer about 5 Mb/s on a single TCP connection.

But Steelhead appliances provide for TCP Streamlining as well as Data Streamlining. Using High-Speed TCP, an IETF-specified mechanism, a Steelhead appliance can enhance standard TCP dynamics in order to take advantage of the high available bandwidth while overcoming high latency. Using High-Speed TCP, a Steelhead appliance can achieve more than 600 Mb/s on a single TCP connection, while still maintaining TCP congestion control processes. No cache exists that can do that.

Using High-Speed TCP, a Steelhead appliance can achieve more than 600 Mb/s on a single TCP connection, while still maintaining TCP congestion control processes. No cache exists that can do that.

So Steelhead appliances can be genuinely useful in disaster recovery. They can allow for faster replication of primary data to secondary storage, and they can accelerate the recovery process as well. The transfer mechanisms, combined with the data coherency characteristics above, mean that the Steelhead at the affected site is not "in charge" of the restoration, but instead the origin servers define what data needs to be transferred, and the Steelhead appliances focus on accelerating that data transfer.

## Made for Streamlined Deployment and Management

Steelhead appliances take a transparent approach to application acceleration. The end result is less effort to implement the devices, and significantly less effort to manage the devices, when compared to caches.

Here is a partial list of Steelhead appliance deployment and management features. Compare this list to any cache's management features:

- Auto-detection of peers across the WAN
- Auto-interception of traffic
- QoS Support
- No requirement for tunnel configuration
- No IP address changes for clients or servers
- No need to identify file servers
- No "superuser" file server account
- No DFS changes required
- No router configuration required

 

In addition to these features, Steelhead appliances simplify ongoing management by providing a optional Central Management Console for bulk management and configuration, SNMP support, and redundancy and clustering features.

While managing a cache is just like managing a specialized, complex server across the WAN, Steelhead appliances genuinely simplify branch office IT infrastructure.

## Conclusion

Caches have a myopic approach to application acceleration. Caches are limited in the applications they can accelerate, may cause data coherency problems, don't function in all use cases, and require just as much work to manage as any other server. Why bother with all of that effort?

Riverbed Steelhead appliances can accelerate all applications that run over TCP, while at the same time streamlining deployment and ongoing management of branch office IT infrastructure. Steelhead appliances designed to have the flexibility and the intelligence to operate as organizations would like to operate, and not force IT staff or users to make dramatic changes to their normal business processes. Because of this unique approach, Steelhead appliances have become the technology of choice for enterprise application acceleration.

Riverbed Technology, Inc.
501 Second Street, Suite 410
San Francisco, CA 94107
Tel: (415) 247-8800
www.riverbed.com

Riverbed Technology Ltd. UK
200 Brook Drive, Green Park
Reading RG2 6UB
United Kingdom
Tel: +44 118 949 7002

Riverbed Technology Pte. Ltd.
350 Orchard Road #21-01/03
Shaw House
Singapore 238868
Tel: +65 68328082

Riverbed Technology K.K.
Shibuya Mark City W-22F 1-12-1
Dogenzaka, Shibuya-ku Tokyo
Japan 150-0043
Tel: +81 3 4360 5357