

Revolutionizing the Application Delivery Network



TMOS: The power behind the BIG-IP system



Getting a grip on the growing wave of dynamic, feature-rich applications is no small task in today's application delivery space. From security to availability to optimization, it is no longer enough to simply classify application traffic by TCP port or IP address. To get the job done, the network will have to become an integral component of the application lifecycle. Likewise, Application Delivery Networking (ADN) devices, typically known as Application Delivery Controllers (ADCs), need to better understand the nature of the applications they will be delivering. To handle these applications properly, application networking equipment needs to be able to fully inspect and act on all aspects of the traffic, not just the first few bytes or packets. It has to grasp the myriad data formats being exchanged and the context in which the entire application session is being delivered. It all adds up to the need for a smarter, faster, and more integrated ADC.

Multiple point products that don't share information or communicate add complexity to the mix. Point products that might provide security, optimization, and/or high-availability services for the application must each process and parse the traffic individually. The application's efficiency, availability, and responsiveness all take a hit when the same data is processed multiple times. In a worst-case scenario, this results in multiple sources of error, management, and troubleshooting. The outcome is a seemingly endless array of devices that need to be carefully and constantly managed to ensure that each is behaving as expected—even when they are integrated into a single platform.

Unfortunately, both the need for increased intelligence and the lack of consolidation work against performance in delivering applications. Increased intelligence means increased processor cycles and delays in forwarding traffic. Because this is currently done two, three,

Illustration by Doug Ross

or even a dozen times as the traffic moves from one device to the next, it amplifies the latency. Coupled with the growing performance demands of today's applications, typical solutions compound the performance problem and make it worse. Networking devices are taking more cycles to optimize applications that are also taking more cycles over the network. It's exponential performance degradation.

The answer to the current set of application intelligence, availability, performance, and consolidation challenges is an integrated solution that understands and can manage both the network and the applications. As the application environment changes, the ADC can adjust the network to accommodate; likewise, with network changes and fluctuations, the ADC can manage application access, availability, and performance in real time. This is often called "application fluency." Application Delivery Networking needs a champion in the ADC that can fulfill all of these needs and more.

Optimizing the network

TMOS, the software platform on which F5 ADN products are built, is how F5 is meeting these needs. "TMOS is designed with the notion of optimizing around transactions associated with application traffic in the network," says Karl Triebes, senior vice president of product development and chief technical officer at F5. TMOS was built specifically to address ADN challenges for today as well as tomorrow. It revolutionizes the way applications are delivered to users by understanding the typical applications deployed in the data center and how the network is designed and functioning. Then it manages this information all the way down to the hardware level.

A smarter platform

TMOS is based on a full-proxy design. Indeed, the first network-based firewalls were originally designed as full-proxy devices, because their designers were well aware that a full-proxy design provides more intelligence and understanding of the traffic being transmitted. Unlike more traditional (and much less feature-rich) packet proxies, TMOS' fully integrated session-based proxy provides line speed processing at layers 2-4 without sacrificing application processing at layer 7 (where all the application intelligence takes place). TMOS enables the ADN solutions to inspect and analyze all application data rather than bits and pieces of it—out of context—within individual packets. This provides TMOS with the intelligence necessary to dynamically adjust to application traffic and apply the most comprehensive services.

With F5 iRules, an event-driven scripting language based on the TMOS architecture, organizations can adapt and modify application messages and traffic in real time in order to meet their specific business and technical needs. "Every customer

application environment and network is vastly different,” Tribes explains. F5 developed a holistic way to solve this problem with iRules, which make F5 devices extremely flexible and give users the power to customize the way in which the BIG-IP system processes application traffic. “They can arbitrarily work with the protocol’s true programming language. But you don’t have to use iRules to get tremendous value from the BIG-IP product family. The capability is there to enable customers to be flexible and respond quickly to application, client, or network changes that can impact their ability to conduct business,” Tribes says.

Growing with the business

TMOS is also modular in design. Having more than a decade of experience in the development of the Application Delivery Networking space, F5 is fully aware of the constantly changing nature of making both the network and application secure, fast, and available. Making TMOS modular allows different pieces of functionality to adapt and change over time without affecting the rest of the system; it also allows the inclusion of new functionality by simply “plugging in” to the underlying software platform—like SSL-offload, caching, and compression, or even complete products, like F5’s BIG-IP WebAccelerator.

F5 also wanted to make sure that TMOS could easily grow with customers. That is another reason it was important for TMOS to be a modular platform, says Erik Giesa, vice president of product management and product marketing at F5. “It’s an adaptable platform for users, who can pick and choose which modules mesh best with their IT goals. It has also allowed F5 to adapt and add new modules as new technologies emerge,” he says.

Complete integration

True integration—enabling all TMOS-based systems to work simultaneously on traffic, even on the same hardware, and to communicate with each other—provides the TMOS architecture with significant advantages. For instance, it can run F5’s BIG-IP Local Traffic Manager simultaneously with BIG-IP Application Security Manager or WebAccelerator—all on the same box using the same management GUI. “Because the TMOS platform is integrated, users can easily change and add policies that impact security, application acceleration, or any aspect of the network,” says Jon Oltsik, senior analyst at Enterprise Strategy Group. “And users can make these changes without jumping through hoops.”

Faster and smarter

The integration of complete application awareness along with network intelligence enables TMOS to provide all of the benefits, intelligence, and modularity without sacrificing the speed and performance of typical nonintegrated application delivery devices. This is the final *coup de grace* of TMOS—intelligence, adaptability, and performance. “Other solutions use packet-level programming and processing, which does not match F5’s performance or intelligence,” Tribes says. “Their performance numbers are one-quarter of our best product and their level of intelligence is reduced to basic HTTP content switching.”

TMOS provides the underlying architecture on which to build the next-generation Application Delivery Networking products—with the intelligence, flexibility, and performance needed to solve today’s challenges as well as the adaptability and expandability to provide a foundation for the unknowns of tomorrow. When combined with the next generation of multicore, multiprocessor hardware platforms and coupled with

FULL-SESSION PROXY VERSUS PACKET PROXY

So what’s the difference between packet-based systems and full-proxy systems, anyway? Packet-based processing is the norm of most products in the market today. These systems take a “network” view of applications and only provide functionality based on the limited information available in the individual data packets. They don’t actually participate in the flow of data between the user and the application, other than to provide the most basic capabilities.

Session-based proxy systems, on the other hand, completely participate in the flow of information between the client and application, or between two applications—and do so at an application layer. A session proxy completely terminates a connection from the client and then creates its own connection to the application server. This provides several significant benefits:

- ▶ Because the session proxy must terminate the connection on both sides, it must have a complete understanding of the application protocols being used and how they are used for a particular session. This enables it to provide vital services, like protocol sanitation—preventing badly formed data from reaching the server where it might cause problems.
- ▶ Since the client is not attaching to the actual application server, the session proxy system provides server “cloaking.” The client is unaware that it’s communicating with a proxy device which, among other things, is critical for securing the application behind the session proxy. The client, especially a malicious one, will have increased difficulty in identifying the actual systems that run the application, instead believing it is communicating with the application server or service directly.
- ▶ The session proxy system can optimize connections, providing for the most efficient interaction in the following ways:
 - Clients transmitting data over the Internet and into the data center can be optimized for that medium, typically the WAN, whereas servers that are local to the TMOS-based device on the LAN within the data center can be optimized according to the specifics of the LAN inside the data center.
 - TMOS-based systems can provide connection pooling to the servers—enabling the data from several clients to use a small number of existing connections between the TMOS-based system and the server. This significantly reduces the overhead of connection setup/teardown for both the TMOS-based appliance as well as the server.

The primary reason for the use of packet-based systems is that, because they don’t have the capability or design to do application-layer intelligence, they have traditionally been much faster than proxy-based systems. TMOS breaks the performance barrier, allowing the best of both worlds: the intelligence of the proxy-based system and the performance of the packet-based ones.

code-level application intelligence—as exemplified through F5’s Application Ready Networks initiatives—TMOS is part of a complete solution helping F5 customers deploy their applications quickly and efficiently, and with the best security, performance, and availability in the industry. After all, in the end, it’s all about the application. ✨

Written by the F5 technical marketing team of KJ Salchow Jr., Lori MacVittie, Alan Murphy, and Paul Stalvig, with freelance writer Denise Pappalardo O’Connors