Comparative Performance Report

Application Delivery Controllers
Document Version: 2013

Inside:

COMPARING PERFORMANCE

The industry's most accurate, transparent and repeatable performance report covering the leading application delivery controllers (ADCs). This report provides valuable information to assist in comparing products from multiple vendors. Even if the vendor you are considering is not included in this report, the information is a valuable aid to understanding ADC performance.

This report includes:

- F5 Networks, BIG-IP 4200v (v11.2)
- F5 Networks, VIPRION B2100 Blade (v11.2)
- Citrix Systems, NetScaler MPX-11500 (v10.0)
- Cisco Systems, ACE 30 Blade (v5.2)
- Radware, Alteon 5224 XL (v28.1)

Letter of Introduction

The following performance report from F5 is a *comparative* and not a *competitive* report, which is an important distinction.

The guiding principles of a comparative report are that it must be accurate, transparent, and reproducible. It must provide as much factual information as possible so customers can make informed product decisions. All test methodology and configuration information must also be freely available so anyone can validate the veracity of the data produced. A comparative report requires complete transparency in how results were derived so that, as in science, the results are reproducible.

In contrast, the purpose and intent of a competitive report is for use as a sales tool, which does not assist in making an informed product decision. Reports of this nature tend to be official-looking documents released by third-party test firms that have been paid by a specific vendor. Customers can distinguish between the two types of reports by determining whether anyone has full access to the complete test methodology and the device configurations involved in the report. The methodology should mimic real-world use scenarios and not artificial ones that inflate numbers. If none of these attributes is present, it is a competitive report and should be viewed with skepticism.

Unfortunately, discerning between the two approaches isn't always easy for customers. Third-party reports have an air of authenticity and impartiality and appear to be objective comparisons between vendors. However, many customers are not aware that the vendor paying for the test often designs the methodology so their product "wins." The third-party test firm is not tasked with verifying whether the tests are meaningful or artificial, it simply runs the tests as directed and validates the results. While these reports appear to be meaningful and relevant, they can be very misleading, as evidenced by the following examples:

- One vendor released a report on their product's Layer 7 performance versus F5's. The numbers derived simply did not match our own performance benchmarking, which we go to great lengths to ensure is as accurate and factual as possible. As best we could discern (the test methodology was not freely available), it seems the methodology used generated HTTP (Layer 7) traffic, but it was processed only at Layer 4; no Layer 7 processing was actually done. However, the claims were touted as Layer 7 sessions per second, implying that actions such as content switching, redirection, and persistence were being performed. The report omitted this important detail. The results were misleading because Layer 4 processing is easier to perform than Layer 7 processing and yields higher results. Unfortunately, only after a customer bought the competitor's product would they become aware that they would get less than half the L7 performance from that product when doing real-world HTTP processing.
- Another report made some extraordinary SSL TPS claims about their product versus F5's. The results of their test made little sense to our performance experts. Like the previous example, the methodology was not freely available. After some intensive research, we were able to reproduce the "SSL" numbers reported by the third-party test firm paid to validate the results. The test seemed to have been designed to artificially inflate SSL TPS. Rather than measure the resource-intensive work of setting up SSL connections, the test measured how many HTTP requests could be pipelined over a single SSL connection. The test results were factual for what was measured, but incorrectly represented as SSL TPS.

These examples illustrate why such reports should never be viewed as truly comparative evaluations but rather as competitive sales tools. Such reports are the primary motivation for F5 producing the following comparative report. F5 stands behind all results in the following pages. We conducted these tests with our own performance experts and we intentionally did not pay for or hire an "independent" third-party test organization. However, we recognize we are not perfect and that honest mistakes are possible. F5 stands behind its testing methodology, but we welcome constructive input for improvement. We want our results to be open, honest, and repeatable.

Sincerely,

Karl Triebes, SVP
Chief Technology Officer, F5 Networks

OVERVIEW

The 2013 F5 Performance Report documents the performance of Application Delivery Controllers (ADCs) from the four top vendors based on their reported market share: F5 Networks, Cisco Systems, Citrix Systems, and Radware. To better reflect the consumers of this report, this document segments the results of the testing by price ranges, where the products compared are offered at a similar list price. It is hoped that this methodology will lead to easier comparison and consumption of the data.

The market for ADCs is very competitive, with nearly every vendor claiming a performance advantage in one scenario or another. Unfortunately, the claims from each vendor rest on differing definitions of performance criteria. Each vendor has their own interpretations of various terms (such as Layer 7 and connection), preferred configuration settings for the different devices, and the presentation of results. These factors significantly reduce the value of typical vendor performance claims. With the inconsistent definitions between vendors, especially in the context of published data sheets, performance metrics cannot be fairly compared between vendors.

Some vendors publish reports from performance tests that they have hired third parties to produce. The configuration files for the devices tested and the testing equipment are rarely made available. Methodology is also rarely publicly documented, and often when it is, the documentation is vague and incomplete. It is impossible to determine the fairness of the tests or their applicability to a customer's needs without this information.

For six years, F5 has produced the industry's' most transparent and robust performance testing methodologies guide[1], publishing it in the public domain. With this publicly available guide, customers are given the framework for evaluating the performance of multiple ADC products with consistent definitions and evaluation criteria. This report, like those before it, follows this methodology.

In the interests of transparency and independent test verification, F5 has published the configuration files for each device included in this report as well as the configuration files for the testing equipment. This allows anyone with the equivalent testing gear to reproduce these tests independently. The configuration files are available on F5's DevCentral web site[2].

TESTING PROCESS AND ENVIRONMENT

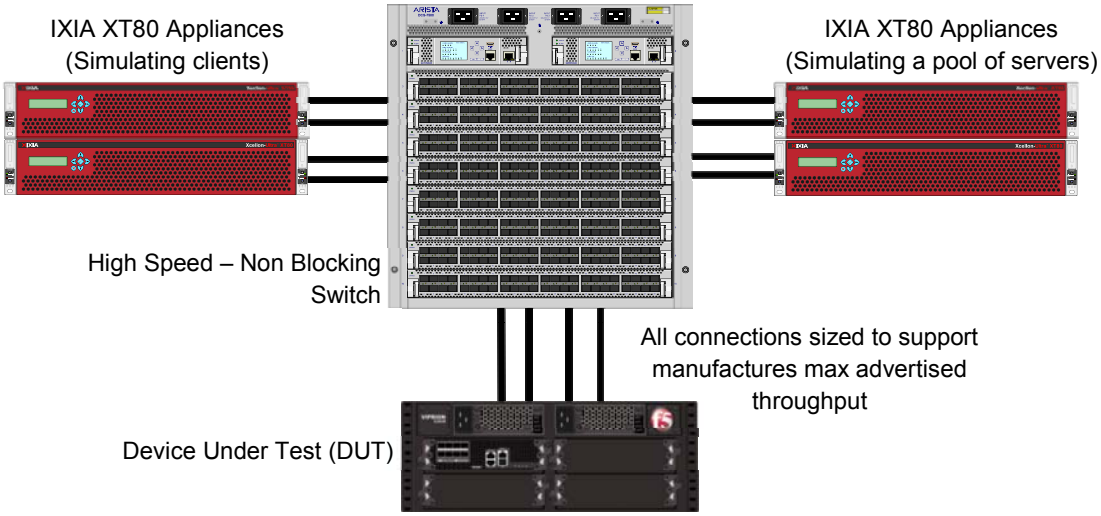
Each of the products in this report went through the same multi-phase testing process that F5 has used in previous reports. This process consists of the following phases:

1. Preliminary Testing: Create and validate the configuration for each Device Under Test (DUT) so that all DUTs manage the network traffic the same way.
2. Exploratory Testing: This determines the best test settings for each device and reveals how well it performs in each type of test. The DUTs configuration is finalized during this phase.
3. Final Testing: Each type of test is run multiple times. Testing is repeated until there are at least three good runs that consistently produced the best results. It can take many runs of a test to reach this standard of consistency.
4. Determine Best Results: The three best test runs for each type of test are examined in detail to identify which one produced the best overall performance. The results of that best run for each type of test are what is used in this report.

In total, more than 1,000 test runs were conducted in order to produce this report.

[1] http://devcentral.f5.com/downloads/f5/creating-performance-test-methodology.pdf
[2] http://devcentral.f5.com/downloads/perf/PerformanceReport2013_configs.zip

The following diagram illustrates a high-level view of the environment in which the devices were tested.



All tests used IXIA XT80 appliances to generate and measure the network traffic that simulated clients and servers.

Products Tested

The following table lists the products tested for this report and the vendor's published performance specifications for each device.

Vendor Published Performance Specifications (based on information published by the vendor on public web sites or datasheets)

Vendor	Device	L4 CPS	L4 Throughput (Gbps)	L7 CPS	L7 RPS	SSL TPS (2K Key)	SSL Bulk Throughput (Gbps)	Compression (Gbps)
F5	VIPRION 2400 chassis with (1) B2100 blade.	400,000	40	150,000	1,000,000	10,000	9	10.0
F5	BIG-IP 4200v	300,000	10	100,000	850,000	9,000	8	8.0
Cisco	ACE 30 Blade	500,000	8	200,000	*	*	6	6.0
Citrix	NetScaler MPX-11500	*	8	*	1,200,000	15,000	6	3.5
Radware	Alteon 5224XL	480,000	8	190,000	*	11,200	4.1	3.6

** We were unable to find numbers published by the vendor for these categories*

It is always difficult to determine which devices are the best fit for a given series of tests. A test that compares one vendors' high-end product against the entry level product of a competitor is essentially useless. For this testing, it was decided to test products in similar price bands and run comparisons from that perspective. This fits well with the idea that there is a given budget constraining the purchase of an ADC and offers comparisons that can be zeroed in on. No selection of devices is perfect, however, so we simply offer an open and honest explanation of what the grouping criteria were. The following is a guide to the chart labels that are used throughout this document.

Segment One ($40,000 to $60,000 USD List Price)

- F5 Networks, BIG-IP 4200v
- Citrix Systems, NetScaler MPX-11500

Segment Two ($60,000 to $80,000 USD List Price)

- F5 Networks, VIPRION B2100 Blade (in a C2400 Chassis)
- Cisco Systems, ACE 30 Blade (in a 6509 Chassis)
- Radware, Alteon 5224 XL

L4 Tests

LAYER 4 OVERVIEW

Layer 4 (L4) performance is a measure of basic TCP/IP load balancing, a baseline configuration with the minimum set of features enabled. L4 performance is most relevant for applications that deal with a lot of bulk data, where little application awareness is required. Load balancing FTP and DNS servers are common scenarios for L4 load balancing. All devices tested are configured with options for a L4-only (or TCP only) mode. These modes limit the features available, but will often produce much better performance with that limited feature set. Since the device is generally doing very little processing, L4 tests often show the highest connections per second and throughput results possible for a given Application Delivery Controller. This makes L4 testing appropriate for use in baseline testing; as it is very unlikely that performance under more complex scenarios (i.e. with additional features enabled) will be higher than the baseline L4 results.

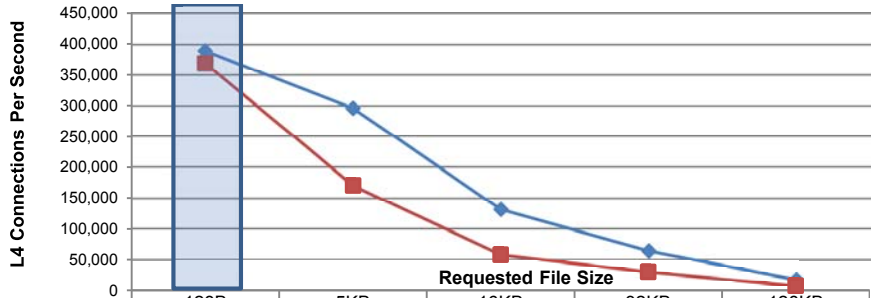
Layer 4 Connections per Second

The L4 Connections per Second (CPS) test is an indicator of raw packet processing power. In short, this is a gauge of how fast TCP connections can be created and closed. While the test set included all standard file sizes, this metric is normally indicated using 128 bytes of data, replicating small data transmissions such as those used by AJAX requests. Many of the publicly available numbers will show much different numbers than this by utilizing either zero byte files, or even resetting the TCP connection instead of shutting it down. Neither of which is common in the real world nor recommended by the standards.

L4 CPS Test Results

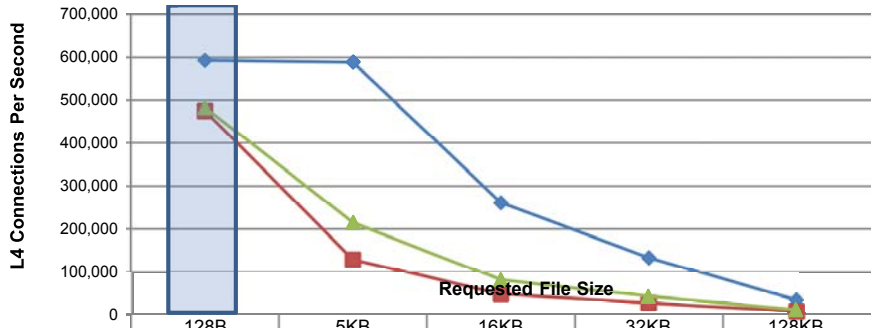
By any measure, the VIPRION B2100 can handle the largest number of connections due to specialized layer 4 offload hardware. In segment one, the NetScaler MPX-11500 comes in a close second to the BIG-IP 4200v, but as expected, both devices come in lower than the segment two devices.

Segment One



	128B	5KB	16KB	32KB	128KB
BIG-IP 4200v	389,799	295,678	131,309	63,940	17,229
NetScaler MPX-11500	369,616	170,145	57,783	29,651	7,528

Segment Two



	128B	5KB	16KB	32KB	128KB
VIPRION B2100 x 1	593,612	589,170	261,087	133,635	33,588
ACE 30 Blade	475,615	128,870	47,651	26,539	6,935
Alteon 5224 XL	483,455	214,946	81,397	42,067	10,852

L4 CPS Analysis

Since Layer 4 only shows part of the problem for an ADC, understanding what the tests are about is important. In an L4 scenario, the device is being used to quickly route packets based on layer 4 information. High performance financial trading systems often use L4 switching as a rapid mechanism to hook users to the correct servers. In this scenario, transferring a small amount of data and closing the connection is common, and this test adequately models such a scenario.

Due to their software based license limitations, at 128 bit file size, the results for the NetScaler MPX-11500 and Alteon 5224 XL devices probably show the limits of the appliance, where the device is not approaching its licenses throughput limits. At larger packet sizes, the device becomes limited based on licensed throughput, and the connections per second are being restricted due to licensing. What this means to customers is that they should carefully set their expectations for software license limited platforms since not all metrics will scale in the same ratio as the measurement that is license limited.

Layer 4 Throughput

The best use of L4 throughput tests in the modern ADC marketplace is to garner an indication of the maximum throughput that the device can handle. Since ADCs provide processing far beyond just shoving packets through, the L4 throughput tests should be considered an absolute maximum, and not an indication of the throughput you are likely to see from the device under normal usage scenarios.

As with all throughput measurements, layer 4 throughput is measured using large file sizes and multiple HTTP requests per TCP session to minimize TCP overhead. This is normally the metric that is used in categorizing ADC's and is used in marketing and data sheets as the device's throughput.

Generally, a devices layer 4 throughput is limited by the available bandwidth, either in external ports or internal devices since there is little processing needed in forwarding layer 4 traffic.

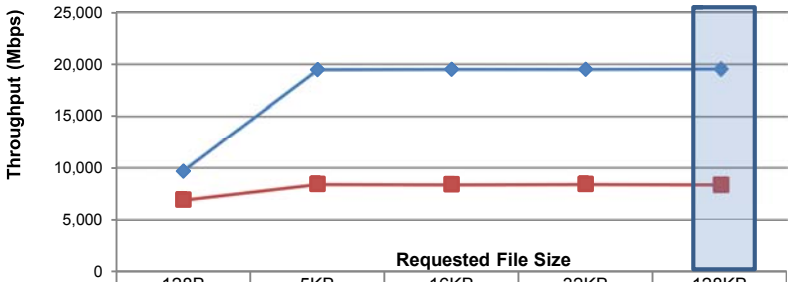
L4 Throughput Test Results

The Alteon 5224 XL and NetScaler MPX-11500 both performed at a level above their licensed throughput. Since the licenses were more than likely the cause of the leveling off of these devices' throughput, they can only be marked "performed better than advertised", and comparisons to other, non-rate limited devices should be avoided.

The VIPRION B2100 takes this test hands-down, with a max throughput of over 38 Gigabits per second. This is more than twice the maximum advertised throughput of ether the ACE 30 or the Alteon 5224 even with their highest license levels. As licensed, the ACE 30 performs at right around 20% of the throughput of the VIPRION B2100.

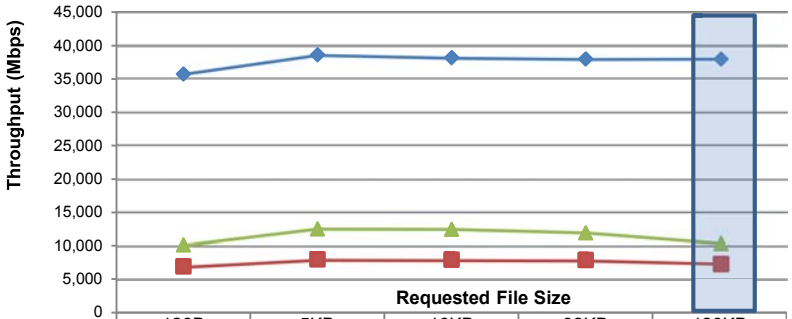
Considering the high performance of the VIPRION B2100 and BIG-IP 4200v, and the considerably lower throughput capability of the ACE 30, if Layer 4 throughput is important to an organization, replicating this test with all devices licensed at the level required is advisable.

Segment One



	128B	5KB	16KB	32KB	128KB
BIG-IP 4200v	9,700	19,485	19,517	19,507	19,546
NetScaler MPX-11500	6,946	8,446	8,419	8,436	8,424

Segment Two



	128B	5KB	16KB	32KB	128KB
VIPRION B2100 x 1	35,731	38,582	38,169	37,964	38,009
ACE 30 Blade	6,874	7,930	7,871	7,823	7,389
Alteon 5224 XL	10,080	12,492	12,466	11,922	10,370

L4 Throughput Analysis

This test was run at 100 requests per connection to provide very minimal TCP overhead. The license limiting of two devices makes this test less easy to analyze and less useful overall than most of the tests in this report.

Note that the F5 appliances performed significantly higher than the rest of the devices. Since throughput numbers are often limited by internal bandwidth between components within the ADC, the benefits of purpose designed hardware that is designed to move large amounts of data within the appliances shows up clearly.

It is worth noting that the Alteon 5224 XL device far surpasses the licensing limit of 8 Gbps throughput, showing a measured 12 Gbps of throughput. We believe this is an artifact of how Radware implements its licensing limits.

L4 Performance Envelope

Performance tests are designed to measure the maximum performance with a given set of circumstances and settings. When measuring requests or connections per second, you could use smaller file sizes to maximize the number of requests that will fit in the bandwidth available. When measuring throughput, larger file sizes can be used to minimize overhead and reduce the effect of transactions on the measurements. All of these measurements can be considered the best that the device can do in those circumstances.

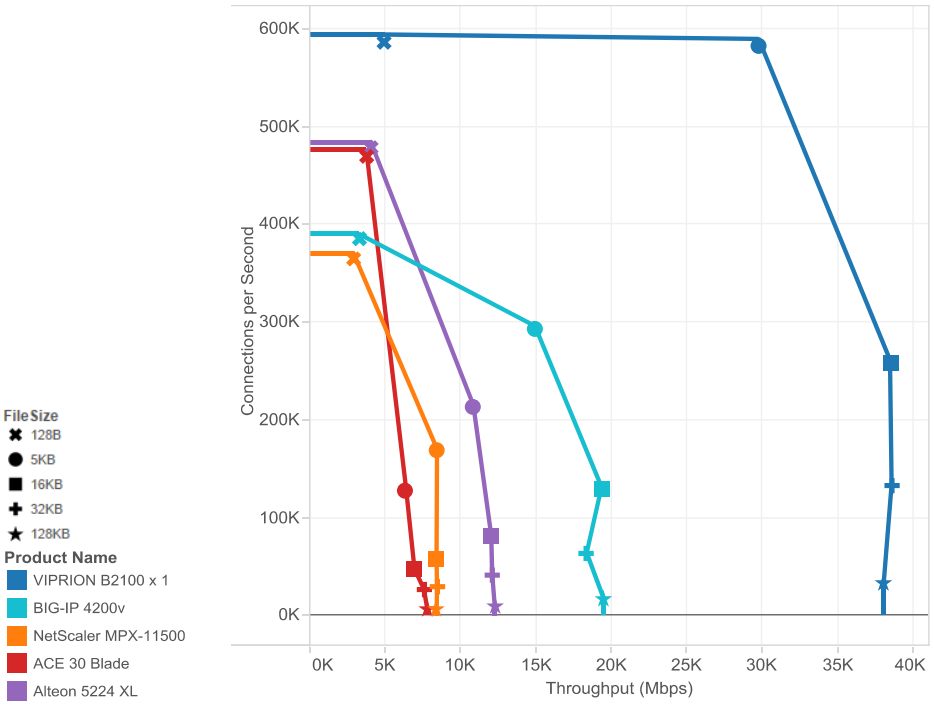
In the real world however, network traffic is very seldom a matter of transferring all very small files, nor is traffic made up of all large files to transfer. The real world of networking and applications is mixed. Web pages and web applications are made up of small and large image files, medium sized html and script files, and small requests and acknowledgements, as well as many other files and packets of all sizes. Some of these require complex handling by an ADC; others do not require any special handling. Because of this, no one measurement will completely represent a customer's real world network.

In an effort to try to provide a better way of comparing device performance for real world applications, we have come up with performance envelopes. These charts display connections or requests per second on the vertical axis, and throughput on the horizontal axis using multiple file sizes. Since these numbers indicate the maximum performance in that direction for that file size, this can be seen as the outer edge of the performance of the device. The actual performance would be somewhere within this envelope, depending on many factors including file size and processing required.

A larger performance envelope indicates devices that will perform better in all areas. A performance envelope that is shorter, but wider will perform better with larger files or packets, but may not do as well with smaller files. A taller but thinner envelope will perform better with smaller packets and files.

One thing to look for is the relative performance at the 16 kilobyte file size. According to Google research, the average size of files transferred is approximately 16 kilobytes[3], indicating that this is may be reasonable place to look for the average performance of these devices.

This chart shows layer 4 performance using L4 CPS, and Throughput. There are similar charts for SSL performance as well as layer 7 performance.



Because the ACE 30 has a high connection rate that drops rapidly as throughput increases, this chart shows that the ACE 30 is tuned to handling fast connections. The Alteon, on the other hand, is more tuned for throughput as shown by its higher throughput at larger file sizes.

L7
Tests

LAYER 7 OVERVIEW

L7 performance is relevant to most applications being deployed today. The most important difference between a L4 test and a L7 test is that the Device Under Test (DUT) must inspect the application-layer data transferred between the clients and servers in order to take some kind of action such as load balancing it to a specific set of servers based on the content (such as small images going to one server and dynamic web pages going to another), or changing the content for security or functionality purposes. Because every client request must be inspected, additional stress is placed on the DUT over other types of traffic. Additionally, features such as HTTP request multiplexing and TCP connection reuse can be used to reduce the impact of the traffic on the servers. Because of these reasons, Layer 7 information tends to more accurately reflect the real world performance of an ADC than layer 4 performance metrics do. IT advisory firm Gartner characterizes an advanced platform ADC as one that operates on a full layer L4-L7 basis, delivering a broader range of functionality and achieving full application fluency[4].

[3] According to Google (https://developers.google.com/speed/articles/web-metrics) the average (90th percentile) size of a GET for the top internet sites is 16.75KB.
[4] Gartner, "Magic Quadrant for Application Delivery Controllers," Joe Skorupa, Neil Rickard, Bjarne Munch, 30 October 2012.

In our Layer 7 (L7) performance tests we measure basic HTTP-aware load balancing; every HTTP request is inspected and then load balanced to one of a group of servers. To insure that each request is being inspected, this test performs a simple inspection of the HTTP URI to identify requests for images which would be directed to a different group of servers. Since no images are actually included in the test, the performance numbers are not affected by differing server pool numbers.

Layer 7 Requests per Second

The number of requests that an ADC can process has a direct impact on the perceived performance of a corporate website in high-traffic scenarios. Not all connections are transferring a lot of data all the time, but high traffic does imply that many connections are being made. Testing maximum requests per second offers an indication of performance under high-load situations, but like any testing, the amount of data being transferred in each packet will influence actual performance in a production environment.

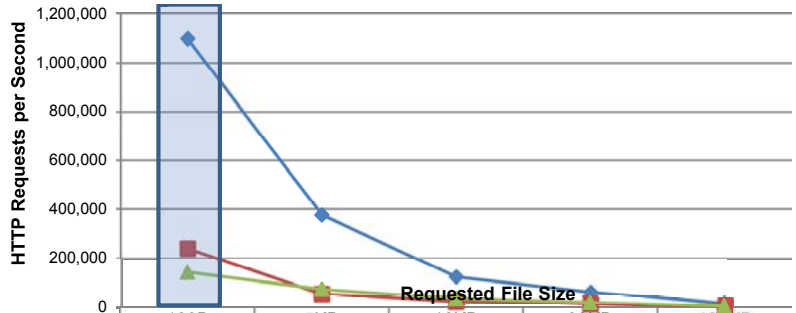
The L7 Requests per Second (RPS) results use 10 HTTP requests per TCP connection. This follows common practice with modern web browsers and applications to attempt to send as many requests per connection as possible, which uses one multiplexed connection to represent multiple connections from a single web browser.

L7 RPS Test Results

Segment One

	128B	5KB	16KB	32KB	128KB
BIG-IP 4200v	859,702	416,464	136,770	69,268	17,507
NetScaler MPX-11500	648,750	179,759	59,056	29,862	7,520

Segment Two



	128B	5KB	16KB	32KB	128KB
VIPRION B2100 x 1	1,098,621	376,647	124,515	63,066	15,795
ACE 30 Blade	240,352	54,690	24,026	12,955	4,806
Alteon 5224 XL	144,010	75,028	34,788	19,161	5,358

L7 RPS Analysis

It is normal and expected to have a performance curve with results on smaller packet sizes showing many times faster than results on larger packet sizes. This is due to the increased network efficiency transporting files large enough to fill up packets. Normally this curve goes down sharply and is generally flat with little difference after file sizes of between 1,000 and 5,000 bytes, at which point the differences become very minimal. In the case of the ACE 30 and Alteon 5224 appliances, this curve is actually hidden due to the scale imposed by the VIPRION's high

performance numbers. In this test, the ACE 30 and Alteon 5224 devices performed worse than either device in the lower price band, indicating that they may have problems with processing L7 traffic.

Layer 7 Throughput

Layer seven throughput is the throughput that a device can achieve when processing traffic at layer 7. Similar to layer 7 requests per second, this measures the ability of an appliance to keep up with traffic while also performing the functions of an application delivery controller.

There are two ways of measuring Layer 7 throughput. One way is to follow similar measures as with layer 7 requests to insure that the DUT is performing layer 7 processing, and then measure the traffic generated at layer 2 by the processes involved and report that number. The second way is to actually measure the layer 7 content carried by that traffic without any of the overhead of the TCP/IP or Ethernet framing (often called goodput). Neither of these is particularly better than the other; however it is important to know the difference.

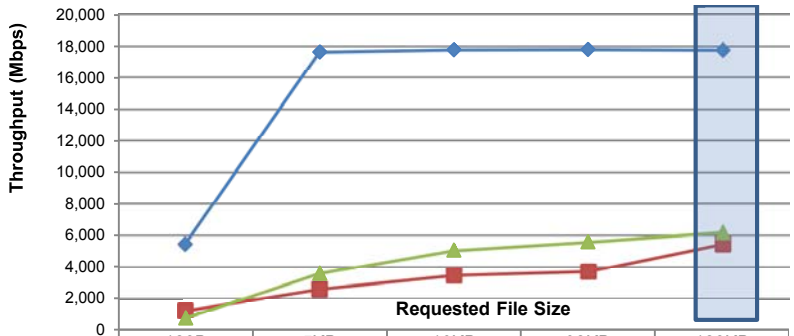
F5 has chosen to measure Layer 7 throughput using the first method (layer 2 throughput while processing traffic at layer 7). This is consistent with the testing performed by most other device vendors.

L7 Throughput Results

Segment One

	128B	5KB	16KB	32KB	128KB
BIG-IP 4200v	4,260	19,489	19,515	19,528	19,543
NetScaler MPX-11500	3,216	8,447	8,422	8,431	8,427

Segment Two



	128B	5KB	16KB	32KB	128KB
VIPRION B2100 x 1	5,445	17,633	17,776	17,791	17,750
ACE 30 Blade	1,190	2,588	3,452	3,708	5,442
Alteon 5224 XL	726	3,576	5,053	5,579	6,193

This is one of the tests where licensing played a key role in the results. The NetScaler MPX-11500, ACE 30 and Alteon 5224 devices were licensed for 8 Gbps of performance, which is less than the maximum the device can handle. In the case of the NetScaler MPX-11500 device, the 8 Gbps limit was the limiting factor in performance results. For the ACE 30 and Alteon 5224 devices, 8GB was never hit during the testing, so the license limit should have played no part in the test results. Since we do not know specifically how Cisco or Radware limit performance with licensing, it is unknown how much this would be affected at other licensing levels. This is a good example of some of the potential problems inherent with limiting performance with software limits.

L7 Throughput Analysis

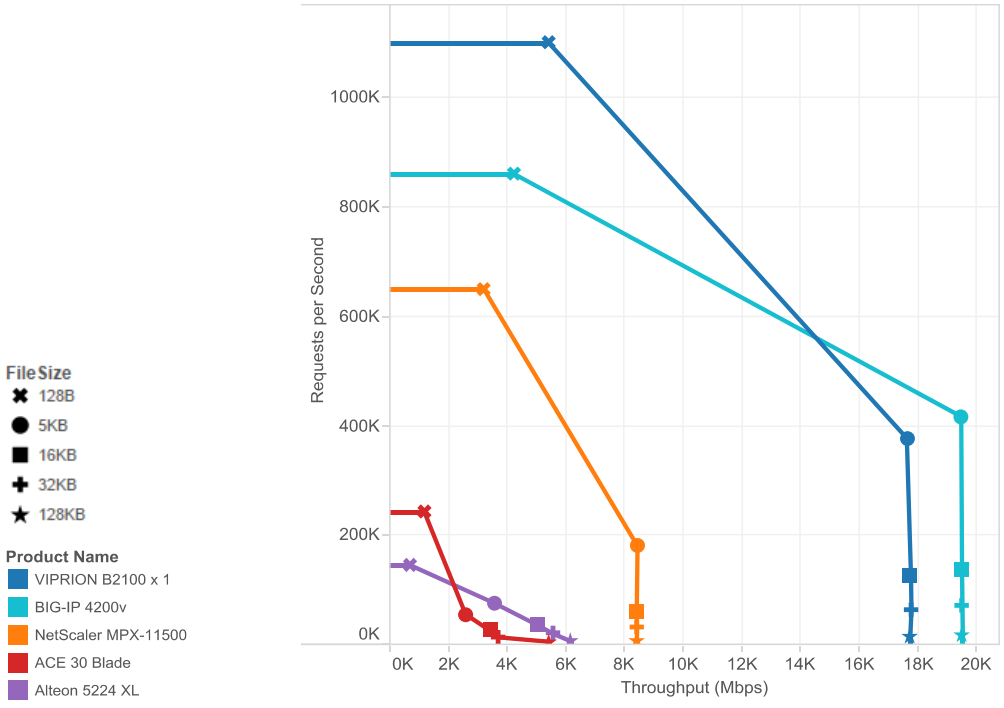
Most vendors do not specify the throughput at layer 7, so judging their performance versus advertised throughput is not an option. This leaves us to the above comparisons and some generalizations about the platforms and issues with the testing. The ACE 30 was the worst performing of the series, at just under 5 Gbps even though it was licensed for 8 Gbps.

As expected, the file size had an influence on the overall throughput, but the way that devices responded to increasing file sizes is interesting. The F5 products tend to hit a maximum performance early (at smaller packet sizes) and stay there, while other devices had different patterns. The NetScaler device performance graph looks much like the VIPRION B2100, with a leveling off at the licensed throughput limit, while Alteon 5224's performance is more of a curve going up slowly as packet size increases, and the ACE 30 almost stair-stepping up as packet size increases.

The performance curve flattening out early with the BIG-IP 4200v, VIPRION B2100, and NetScaler MPX-11500 platforms in an indicator of the additional horsepower built into the platforms to handle application processing tasks. Generally the performance curve would look similar to the Alteon 5224, NetScaler MPX-11500, and ACE 30 graphs. When the curve flattens out it is normally indicative of something limiting it. With NetScaler MPX-11500, this limiting factor was the license. With the F5 platforms, the limiting factor was the available bandwidth. Any additional tasks that a device would have to handle would generally directly impact the throughput and transactions that a device can handle, however in this case, these devices would have additional overhead to handle processing without impacting throughput as quickly.

The application layer task selected for this test was among the simplest of the tasks that could be performed by an ADC, it was selected for compatibility across multiple vendors' products to insure a fair comparison. However the F5 platforms were designed to support much more complex tasks than this and were designed to be able to handle these tasks with minimal impact on overall performance.

L7 Performance Envelope



You can see in the layer 7 performance envelope that the Alteon will handle more throughput than the ACE 30; however it will not handle the larger numbers of packets inherent in smaller file sizes.

Layer 7 Cookie Persistence

In many applications it is important that requests from any one user are always sent to the same server. This is called persistence. Persistence is frequently required when the application maintains state for each user such as with shopping cart applications. ADCs use a number of techniques to implement persistence depending on the application, with cookie persistence being one of the most commonly used methods. With Cookie persistence, the ADC inserts a cookie in the HTTP header sent to a client... This cookie identifies which server should be used to handle requests for that client. The client returns this cookie to the ADC with all subsequent requests which are then forwarded to the same server.

This test measures the ADCs performance when required to do the extra work of creating, inserting, tracking, and then parsing received cookies. The measured metric for this is Requests per Second as the cost of inserting and processing cookies impacts the number of requests that an ADC can handle.

While Cookie persistence is an F5 patented technology, nearly every ADC vendor in the marketplace has licensed the technology from F5 for their own implementations and all of the vendors in this test implement this feature.

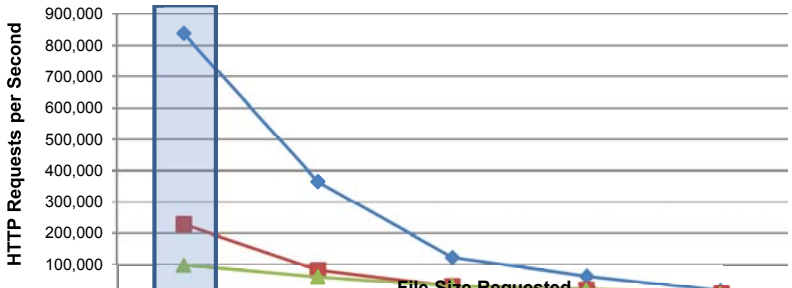
L7 Cookie Persistence Results

The layer 7 test for cookie persistence performance is intended to measure the ability of the ADC in processes that require it to maintain and use large tables of lookup information as well as updating information in the header of the http packet. Since updating information in the data portion of a packet will require updating many of the layers of the packet including lengths and checksums, as well as buffering and reading the traffic, this processing is a good indicator of the abilities of the ADC to handle advanced content management tasks.

Segment One

	128B	5KB	16KB	32KB	128KB
BIG-IP 4200v	699,967	403,178	135,432	69,104	17,486
NetScaler MPX-11500	428,648	178,067	58,830	29,879	7,558

Segment Two



	128B	5KB	16KB	32KB	128KB
VIPRION B2100 x 1	837,336	365,080	123,436	63,166	15,863
ACE 30 Blade	229,112	83,010	31,320	19,747	5,974
Alteon 5224 XL	98,423	60,981	31,354	18,315	5,437

L7 Cookie Persistence Analysis

The process of inserting cookies in the HTTP header demonstrates the limited layer 7 processing capacity of the ACE 30 and Alteon 5224 XL. All the other devices outperform them. As should be expected, all devices had lower performance in this test than in the layer 7 Requests per second measurements. This is an indicator of the complexity of dealing with content at layer 7.

SSL Processing Tests

Secure Sockets Layer (SSL) encryption is used around the world to secure communications between users and applications. SSL is a standard encryption protocol available in every major operating system, web browser, smart phone, and so on. SSL technology helps make online shopping secure, enables secure remote access (SSL VPN) and much more – SSL is ubiquitous in commercial and consumer networking security solutions. SSL provides security using a combination of public key cryptography to share the cryptographic keys, and symmetric encryption (commonly RC4, 3DES, or AES) to actually encrypt the files. Both the key exchange and the various encryption algorithms are computationally-intensive, and require specialized hardware on the server side to achieve acceptable performance or large scale in nearly all commercial uses of SSL.

SSL Transactions per Second (TPS) performance is primarily a measure of the key exchange and handshaking capability of a device. Normally measured with small file sizes, this measures the handshake operations that occur at the start of every new SSL session. This operation is computationally-intensive and all major SSL offload vendors use specialized hardware to accelerate this task. For larger responses and file sizes, the computational cost of the handshake operation is less relevant. Because the operation only occurs once at the beginning of a session the overhead is much less. A more balanced metric for comparison of performance is the throughput of encrypted traffic, also known as symmetric encryption or bulk crypto. Bulk crypto is a measure of the amount of data that can be encrypted and transferred in a given second.

There are different approaches to handling SSL in devices. Some devices will use specialized hardware just for the SSL handshake and key exchange, then use the standard CPU for ongoing encryption. Others will use the specialized hardware for both. By reserving the hardware for the handshake and key exchange, vendors will realize a higher SSL TPS measurement, but their bulk throughput will be reduced since there is no crypto hardware available for this process. Vendors who chose to use the available SSL hardware for both processes, provide increased SSL bulk performance along with very good transactional performance, and will provide better overall performance in most situations. F5 has chosen to use the SSL hardware for all SSL operations.

As usual, tests were conducted across a range of file sizes to demonstrate the performance in a range of situations.

Tests were run using 2048 bit key sizes, which is the size that is recommended by all reputable security agencies and using AES256-SHA ciphers, which is one of the most common cypher algorithms available.

SSL Bulk Encryption

This test is useful to evaluate the possible impact that offloading encryption/decryption to your ADC might have on overall performance. By freeing CPU time on the web server, performance of the overall application will improve. In most implementations, this is the most useful measure of overall SSL performance.

This test is generally measured with larger file sizes and uses one hundred encrypted HTTP transactions for each SSL session to better replicate a normal encrypted page or session. This simulates approximately 2 encrypted web pages[5].

SSL Bulk Encryption Test Results

This test was interesting because of variation across file sizes. The NetScaler MPX-11500 performs very well at smaller packet sizes, but levels off after 5 KB packet size and actually falls behind the BIG-IP 4200v as packet size grows. The ACE 30 shows performance degradation for larger packet sizes – above 16KB, overall throughput goes down.

This suggests which devices utilize encryption hardware to accelerate bulk SSL performance, those devices that quickly level off, or actually drop as the file size goes up would be using the devices' general purpose CPU to encrypt

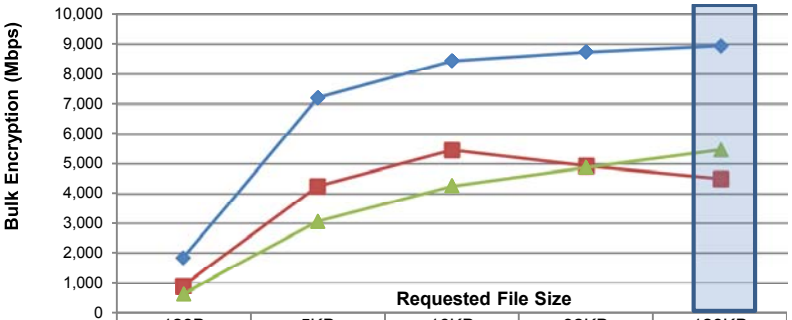
[5] According to Google (https://developers.google.com/speed/articles/web-metrics) the average web page has 42 GET requests.

the traffic. In non-encrypted scenarios, the CPU is used less for larger file sizes since they generate fewer packets, however if the CPU is encrypting these files, it actually gets busier as the file sizes go up, reducing amount of available CPU time for normal tasks and reducing the overall performance of the ADC. Devices with that show a normal distribution of traffic across file sizes show that the general purpose CPU is not being affected by the encryption process.

Segment One

	128B	5KB	16KB	32KB	128KB
BIG-IP 4200v	1,270	5,136	7,286	8,925	9,137
NetScaler MPX-11500	2,196	6,297	6,710	6,756	6,828

Segment Two



	128B	5KB	16KB	32KB	128KB
VIPRION B2100 x 1	1,844	7,208	8,435	8,736	8,949
ACE 30 Blade	902	4,254	5,459	4,926	4,501
Alteon 5224 XL	631	3,069	4,259	4,876	5,464

SSL Bulk Encryption Analysis

There appears to be a systemic limitation of the ACE 30 above 16KB that reduces throughput. This is important if your requirements include streams that transfer larger file sizes of encrypted data. Backups, video and even a decent number of HTML pages fall into this category. Know your data characteristics before picking this platform.

For throughput, the VIPRION B2100 is the overall best performing product, but for very small data transfer sizes, the NetScaler MPX-11500 shows slightly higher throughput. The Alteon 5224XL has a performance curve similar to the VIPRION B2100, but the Alteon delivers roughly similar performance to the ACE 30, without the reduction in performance above 16KB file sizes.

SSL Transactions per Second

Some in the industry point to SSL Transactions per Second (TPS) as the key benchmark for SSL, but there are several reasons why it isn't the best view of overall application performance. Primarily, this measures the ability to set up and tear down SSL connections, which is important, but does not represent the overall performance of SSL in normal use. We prefer to use bulk SSL measurements (throughput) as our key measure because it better represents the entire process – the ability to set up and tear down connections, the throughput those connections can manage, and the time it takes for a connection to do what needs doing and then close/tear down. There are certainly a few situations where a higher transaction rate is more important than a higher overall performance, however these are not common and we encourage customers to better understand their needs before relying only on a transaction rate measurement.

This test uses a single http request per SSL connection, setting up the connection, sending one request and tearing it down again afterwards. This is the standard way of measuring transactions per second, but it is not consistent with normal web traffic patterns where a single SSL connection would carry many application requests.

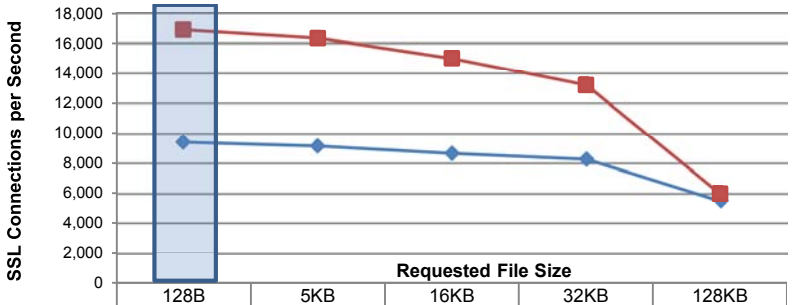
SSL TPS Test Results

For this series of tests, the NetScaler MPX-11500 performs better than the others, with BIG-IP products closely behind.

The performance curve for all devices tested was similar, with the MPX device dropping off faster in terms of connections per second for larger file sizes, but still maintaining a slight lead over BIG-IP products.

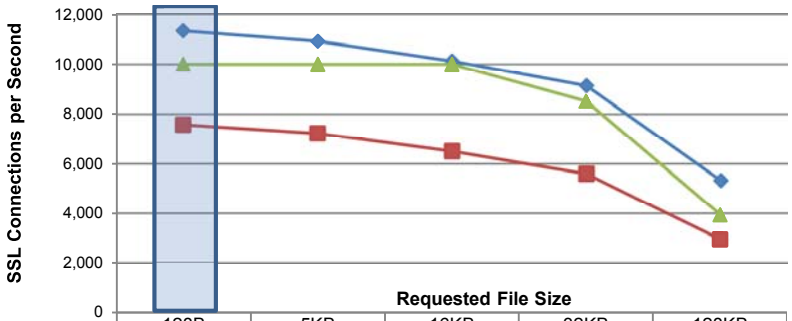
The ACE 30 comes in a distant last on all measured results, barely achieving half of the throughput of the top performing device.

Segment One



	128B	5KB	16KB	32KB	128KB
BIG-IP 4200v	9,426	9,162	8,659	8,280	5,495
NetScaler MPX-11500	16,900	16,348	14,996	13,229	5,952

Segment Two



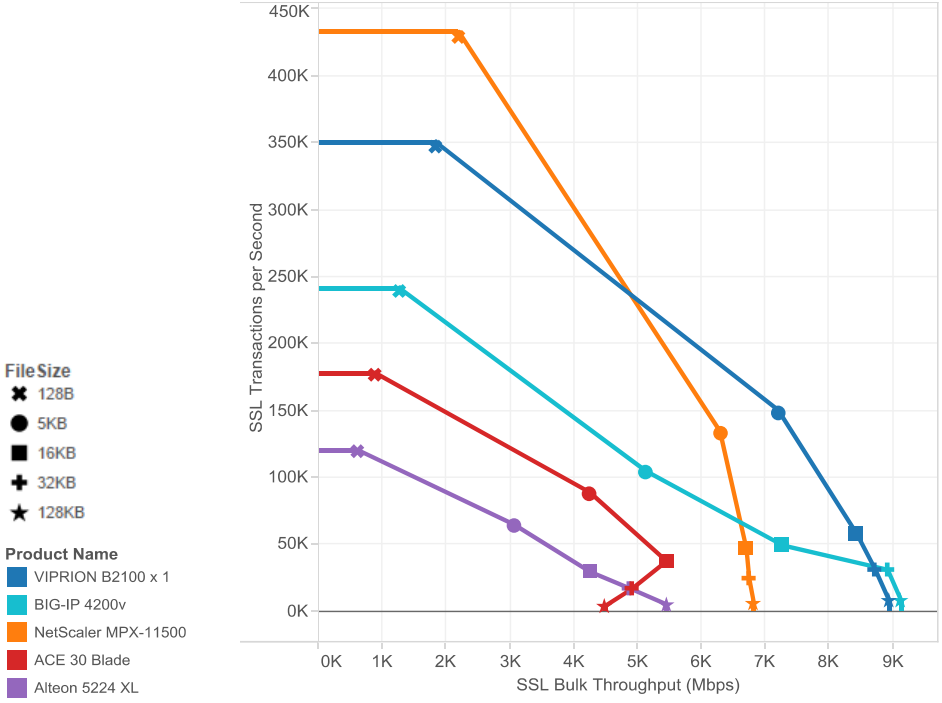
	128B	5KB	16KB	32KB	128KB
VIPRION B2100 x 1	11,345	10,938	10,133	9,144	5,330
ACE 30 Blade	7,576	7,231	6,497	5,602	2,939
Alteon 5224 XL	10,001	10,001	10,002	8,522	3,939

SSL TPS Analysis

As noted in the introduction, this test shows the speed with which an SSL connection can be set up and torn down, but does not test with a significant amount of data through the connection. Since encrypting the data being passed over the connection is a critical portion of the overall performance of SSL, SSL Bulk provides a better measure of the performance possible for a given device. While the NetScaler MPX-11500 device manages to maintain an edge over all other devices when simply starting and stopping connections, the SSL bulk tests show that this edge is less meaningful in real life situations.

In short, this test is useful if your traffic consists of a high volume of starting a connection, dropping a tiny bit of data through it, then closing the connection. If SSL connections are utilized as they are in most enterprises – Web Page delivery, VPN, VDI – then SSL Bulk is the test that will have the most meaning.

SSL Performance Envelope



As with the layer 7 performance envelope chart, this chart combines the request per second metric with the throughput metric to provide a view of the overall SSL performance envelope of the devices.

This chart demonstrates that while the NetScaler MPX-11500 device is higher (supports a larger number of smaller transactions) it does not perform as well with the 16k average sized files, referenced in the L4 performance envelope, as do either of the F5 platforms.

This also shows the interesting drop-off in performance that the ACE 30 has with larger file sizes.

Comp
Tests

COMPRESSION

HTTP Compression performance is a measure of the standard compression algorithms supported in all modern web browsers. In situations where the bandwidth between clients and servers is limited, compression can provide significant performance benefits to end users. Compression can also help companies to achieve cost savings by reducing the bandwidth required to serve web-based applications.

The benefits of compression are widely understood, but compression is not universally used because it's very computationally intensive for servers. As a result, HTTP compression is commonly offloaded to Application Delivery Controllers.

The most important metric when measuring compression is throughput. More data sent from the servers directly correlates with more compression work for the DUT.

In this test, 10 HTTP requests are sent per TCP connection to simulate common web communications.

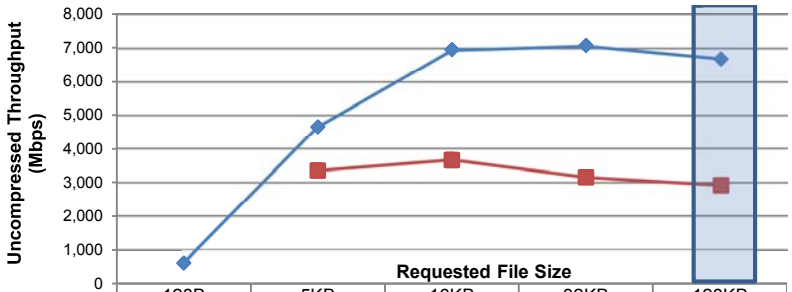
Compression Test Results

At very small packet sizes the ACE 30 performs admirably in this testing, bested only by the VIPRION B2100.

In our testing the NetScaler MPX-11500 would not compress small packet sizes. This is not documented, however it may be intentional since the performance gain in compressing smaller files is often less than the overhead it costs to compress them. Rather than present erroneous results, we started their graph with the next larger file size and utilized their actual results for all other test sizes.

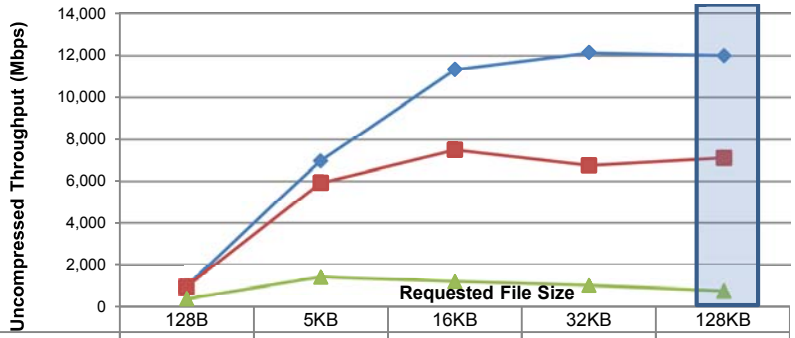
The Alteon 5224's performance was artificially limited based upon the license level of the appliance tested, so the results reported in the overview are the advertised performance numbers and should be treated as "untested", the detailed graphs below show the tested performance, but this should not be taken as their best performance.

Segment One



	128B	5KB	16KB	32KB	128KB
BIG-IP 4200v	611	4,643	6,942	7,066	6,687
NetScaler MPX-11500		3,351	3,664	3,153	2,924

Segment Two



	128B	5KB	16KB	32KB	128KB
VIPRION B2100 x 1	954	6,974	11,333	12,140	12,026
ACE 30 Blade	936	5,903	7,489	6,752	7,105
Alteon 5224 XL	370	1,403	1,197	1,010	753

Compression Analysis

The performance curve of the NetScaler MPX-11500 is flatter than expected and this is, at least partially, because it does not have compression hardware. We observed during testing that at all file sizes, the NetScaler MPX-11500's CPU was at 100% utilization. The performance of the Alteon shows the licensing limitation kicking in quickly.

The BIG-IP 4200v performed at less than its advertised numbers, this was due to the test design limiting the number of simulated users in order to keep a level playing field and maximize the performance of the other appliances. The BIG-IP algorithm is designed to utilize the CPU of the appliance for compression in addition to dedicated hardware when larger numbers of connections are present. Since most appliances are not run at 100% utilization there is normally available CPU for this, but during this test we did not utilize enough simulated users to activate this feature.

The advantage of having dedicated compression hardware is demonstrated when comparing the results for the BIG-IP 4200v to the NetScaler MPX-11500. As noted above, the NetScaler MPX-11500 was at 100% CPU utilization at all file sizes. The BIG-IP 4200v was at only 50-75% CPU utilization with files sizes of 16KB and larger.

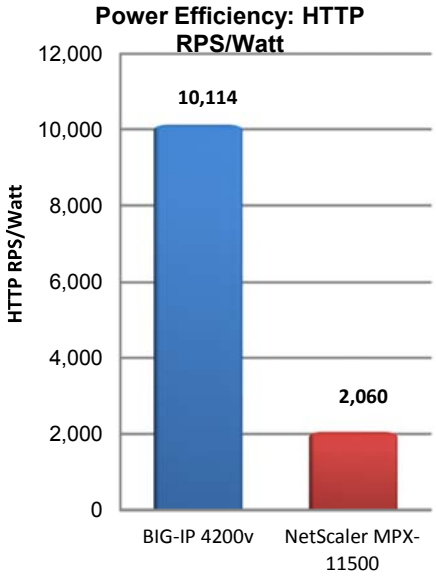
POWER EFFICIENCY

The amount of electricity used by data center equipment is being scrutinized more than ever due to the increasing costs, both direct and indirect, of every watt consumed. A simple way to evaluate the energy efficiency of products is to calculate the work units produced per watt consumed.

The power draw of each device in this report was measured under three conditions. The first measurement was with the device configured and connected as it was for all tests but with no tests running. This is the minimum electricity the device will consume when powered on and connected to the network. The second measurement was taken while the device was under full load during the layer 7 test using 10 HTTP requests per TCP connection. The third measurement was taken while the device was under full load during the layer 4 throughput test. The power measurements for the two chassis-based devices (VIPRION B2100 and ACE 30) include the chassis and all installed components.

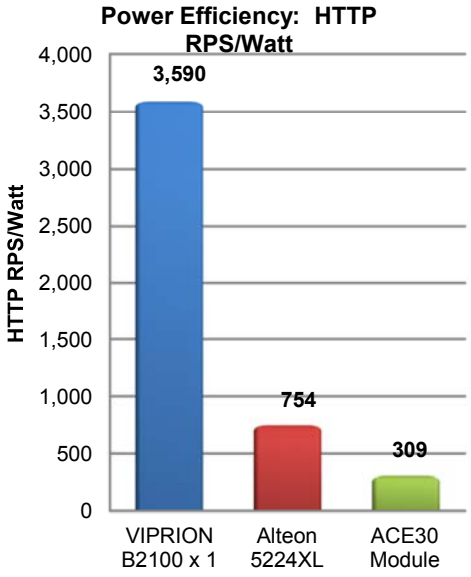
The efficiency of a device was calculated two ways. The first method takes the measured L7 Requests per Second of the device, and divides it by power drawn at full load. The second method takes the measured L4 Throughput and divides it by the power drawn at full load. The difference in power draw between the two full load tests was within the margin of error of the measuring equipment so the same power draw numbers are used to calculate the efficiency of both full load tests.

Segment One



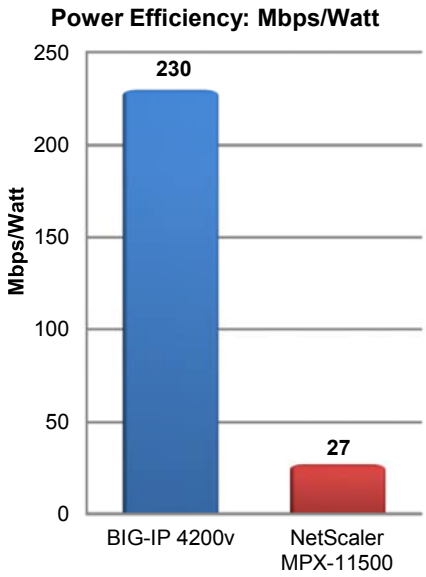
Device	Power Draw		Efficiency
	Idle	Full Load	RPS/Watt
BIG-IP 4200v	65	85	**10,114**
MPX-11500	301	315	**2,060**

Segment Two



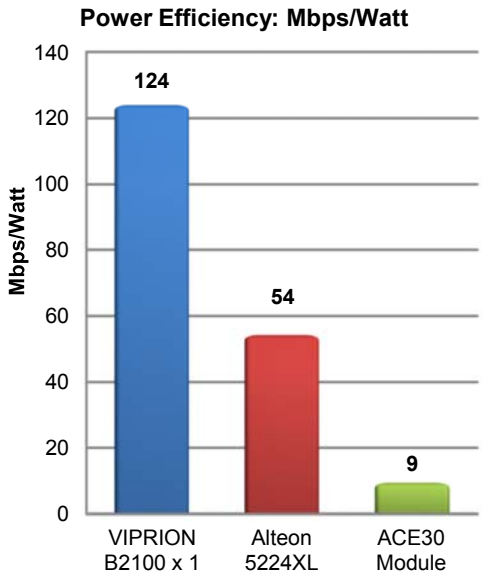
Device	Power Draw		Efficiency
	Idle	Full Load	RPS/Watt
VIPRION B2100	251	306	**3,590**
Alteon 5224XL	191	191	**754**
ACE 30	761	779	**309**

Segment One



Device	Power Draw		Efficiency
	Idle	Full Load	Mbps/Watt
BIG-IP 4200v	65	85	**230**
MPX-11500	301	315	**27**

Segment Two



Device	Power Draw		Efficiency
	Idle	Full Load	Mbps/Watt
VIPRION B2100	251	306	**124**
Alteon 5224XL	191	191	**54**
ACE 30	761	779	**9**

Power Efficiency Analysis

Higher RPS/watt and Mbps/watt calculations indicate more efficient ADC operations. Using the RPS/watt metric, the F5 devices have 174%−3273% greater energy efficiency compared to the other vendor's products.

The two methods used to calculate power efficiency demonstrate how much the way a performance metric is defined can affect the results. As an example, the BIG-IP 4200v is 5x more power efficient than the NetScaler MPX-11500 when calculated (or "defined") as RPS/watt. However, the BIG-IP 4200v is 8.5x more power efficient than the NetScaler MPX-11500 when calculating it as Mbps/watt.

The ACE 30 and required switch chassis draw significantly more power than the VIPRION chassis because of architectural differences between the two chassis-based products. Both chassis had fans and two power supplies installed. The ACE 30 was installed in a Catalyst 6509E chassis which also required a Supervisor Module and a four-port, 10 Gbps Ethernet module. In contrast, the VIPRION 2400 chassis required only the VIPRION B2100 blade.

APPENDIX A: ADDITIONAL TESTING DETAILS

Devices As Tested

Vendor	Device	Software Version	Uplinks to Switch	Simulated Users*
F5	VIPRION 2400 Chassis w/one B2100 Blade	11.2.1-HF1	4 x 10 Gbps	1024
F5	BIG-IP 4200v	11.2.1-HF1	2 x 10 Gbps	768
Cisco	ACE 30 Blade	A5(2.1)	2 x 10 Gbps	768
Citrix	NetScaler MPX-11500	NS 10.0 Build 70.0.nc	2 x 10 Gbps	1024
Radware	Alteon 5224 XL	28.1.8	2 x 10 Gbps	512

** Results reported are for tests with the SimUsers set as shown. Please see Appendix B for the reasons that the SimUsers is different for different platforms.*

Load Generating Equipment

The load testing equipment used to generate and measure traffic to and from the ADCs was IXIA XT80 appliances. The application IxLoad (ver. 6.0-EA) was used to conduct the tests.

Common Test Settings

The following test settings were changed from the defaults in the Ixia software:

- File Sizes: 128B, 5KB, 16KB, 32KB, 128KB
- Number of Client IP addresses: Equal to number of Simulated Users
- Connections per user: 1
- HTTP Requests per connection: 1, 10, 100 (depending on type of test)
- Simulated Users: 512, 768 or 1024 (depending on DUT – See Appendix B, question 7)
- Servers: 72

Test Specific Details

Test	HTTP Requests per Connection	Layer 7 Processing	File size for standard measure
L4 CPS	1	No	128 Bytes
L4 Throughput	100	No	128 Kilobytes
L4 Envelope	1 and 100	No	128 Kilobytes
L7 RPS	10	Yes	128 Bytes
L7 Throughput	10	Yes	128 Kilobytes
L7 Envelope	10	Yes	128 Kilobytes
L7 Cookie Persistence	10	Yes	128 Kilobytes
SSL Bulk	100	No	128 Kilobytes
SSL TPS	1	No	128 Bytes
Compression	10	No	128 Kilobytes
Power Efficiency	10	Yes	128 Bytes

SSL Performance Test Specifics

- SSL Ciphers: AES256-SHA
- SSL Certificates: 2048 bit Key

Compression Test Specifics

- Compression method: gzip
- All files were moderately compressible (approximately 75%) HTML

DUT to Switch Connections

In all the tests for every device, all the interfaces connected from the DUT to the switch used link aggregation and 802.1Q VLAN tags. The link aggregation was manually configured with LACP disabled. Each DUT was connected to the switch with enough interfaces so the available bandwidth met or exceeded its specified performance capability.

APPENDIX B: QUESTIONS AND ANSWERS

1. Why don't the test results match the specifications published by the vendor?

There are several possible reasons why the results published in this report do not match the vendor's performance specifications. One of the most common reasons is vendors using different definitions for a performance measurement. L7 HTTP Requests per Second (RPS) is another measurement vendors have defined differently. F5 defines L7 HTTP RPS as: Full TCP connection establishment (three way handshake – SYN, ACK, SYNACK), HTTP request & response (complete HTTP transaction) and TCP close (FIN, ACK, FIN, ACK), with the device inspecting each HTTP request and making a decision based on it. If only one HTTP Request is sent per TCP connection, F5 calls that test connections per second (CPS). If more than one request is sent per TCP connection, it is a measurement of RPS.

Some vendors define L7 HTTP RPS as the number of HTTP requests transmitted across connections of which the device is only doing L4 load balancing (no inspection or decisions based on HTTP content). In this scenario, the device is unaware of the HTTP requests or responses, as it is only processing data up to L4. F5 specifically does not classify this layer 4 processing as a layer 7 test.

In order to ensure devices are inspecting the HTTP requests, we created a rule (or policy) on the device being tested. This rule examines the URI (the part of the URL after the name of the server) of the HTTP request. If the URI ends in .png, then it directs the request to a different pool of servers than all other requests.

Another common reason a device may not reach its specified performance level is differences in test settings. There are many configurable parameters on both the device being tested and on the testing equipment that can affect performance. For many of these there is not one "correct" setting for a given type of test. Multiple settings may be equally valid.

Finally when the device performs better than the vendor indicates, this may be the vendor making certain they can achieve the numbers they publish. With F5 as an example, both devices can do a far L4 CPS than the marketing literature indicates. But the numbers in the marketing literature are the guaranteed numbers. If something in a future release causes the performance to degrade beyond those marketing performance specifications, steps will be taken to regain performance at least to the marketing number. So in this case, marketing numbers are a guarantee of sorts, while performance numbers in this test are the maximum the device can do *at this time*. Please see question #2 below for more about marketing numbers that do not match performance test results.

2. What are the impacts of license limiting?

It is standard practice within the industry to sell a device capable of a standardized performance level, and then utilize software licensing to limit certain aspects of that device's performance. There are numerous things for users to be aware of when utilizing this type of scenario.

In our testing, the level of software licensing and how well it was implemented appears to have varied widely. Some devices actually do set a hard limit when throughput passes the licensed level, others do not. In some use cases the limit seems to have been more reliable than in others for some devices.

The number of scenarios that can be impacted by software licensing are every bit as large as the number that can be impacted by an undersized device. In the testing performed in this document, it is stated when software licensing may have been the cause of a device under-performing and in throughput tests that is relatively obvious. It is less obvious in other tests, and though it is noted here for the reader, in some scenarios discovering that software licensing could be at fault for performance lower than expected was non-trivial.

Devices Over-performing due to software licensing can cause problems as well if the customer is not aware of the details of the licensing. If a device is limited using throughput, but connections per second is not limited, and they are using that device in a scenario that is dependent on connections per second. When that customer decides they need to upgrade their license to the next level, they may see little or no improvement in CPS for the money they paid for the upgrade.

Many vendors sub-license all sorts of items on the device also – throughput is one license, SSL another, compression another, all at varying bandwidths. While this was the industry standard in the past, that world is changing, and we are moving toward an industry with all core ADC functionality having one license level. Testing and evaluation will be much easier when this is true, until then evaluating how a given network will be impacted by the various levels of licensing is imperative.

It would pay to study these test results and do your own test in a prepared environment before accepting software licensing.

3. How is compression throughput affected by the compressibility of files?

How compressible the files are has a dramatic influence on the throughput of an ADC when compressing traffic. In these tests, we used files that were approximately 75% compressible.

Compression of files is a balancing act, files that can be compressed further take more processing to compress, but are smaller and can be transferred faster. In addition, files that are small enough to lack enough compressibility can add processing time with little gain. Most ADC's have parameters to tune their options to control which files should even try to be compressed. We used consistent settings on these options for all platforms.

4. How did you decide what file size settings to use in the tests?

In our previous Performance Reports, we have tested with different file sizes. We adapted the file sizes used to perform this testing based upon feedback from readers. If there are file sizes more appropriate to a given scenario, please do provide F5 feedback for the next iteration of this report. Based on previous feedback, we chose to use file sizes of 128B, 5KB, 16KB, 32KB and 128KB.

5. Why was the Ixia SimUsers ("simulated users") setting changed for some devices?

Extensive preliminary tests were conducted to verify DUT configurations and identify settings that might be limiting performance. We know from experience that devices usually perform best when the Ixia was set to one number of simulated users (SimUsers) compared to other numbers. During this preliminary phase, the L4 and L7 tests were run with SimUsers setting of 512, 768, 1024, 1536, 2048, 3072 and 4094. If performance degraded from the 1024 to the 1536 settings, and again from 1536 to 2048 SimUsers, then tests were not run at the 3072 and 4096 SimUsers settings.

Because the Ixia equipment is trying to make connections and requests as quickly as it can, one simulated user generates a load similar to several hundred or several thousand real users. We found that there was not a consistent setting that performed well on all platforms. The results presented in this report are from tests with the SimUsers setting at which each ADC performed best during the preliminary testing.

The decision to adjust the test settings for each device was made in order to represent each one at its best, and was necessary to compensate for the varying performance ranges of the tested devices

F5 Networks, Inc. 401 Elliott Avenue West, Seattle, WA 98119 888-882-4447 www.f5.com

F5 Networks, Inc.
Corporate Headquarters
info@f5.com

F5 Networks
Asia-Pacific
apacinfo@f5.com

F5 Networks Ltd.
Europe/Middle-East/Africa
emeainfo@f5.com

F5 Networks
Japan K.K.
f5j-info@f5.com