

“TCP Wirespeed”: Testing TCP Throughput to 10 G

Managing TCP/IP-based Applications in the Ethernet World

As Ethernet speeds rise from 1Gb/s to 10 Gigabit Ethernet (GigE), networks continue to grow increasingly complex. Traditionally, the network provider has been responsible for Layers 1-3 of the network Open Network Initiative (OSI) stack while the end-users’ IT department manages has been responsible for the application layers (5-7). Refer to Figure 1 which illustrates the various layers of the OSI Model.

The gray area surfaces in the management of Layer 4 for Transmission Control Protocol/Internet Protocol (TCP/IP)-based applications such as HTTP, FTP, and e-mail. Network providers must prove that more than just the Layer 3, IP network is functioning properly because IP performance can be excellent, while TCP performance can be poor. Improperly tuned TCP causes degradation in application performance and end-customers often blame the network. The problem may be in the network provider’s domain (lack random early discard queuing, improper prioritization, etc.) or within the end customer’s domain (i.e., poorly tuned TCP window, lack of CPE traffic shaping, etc).

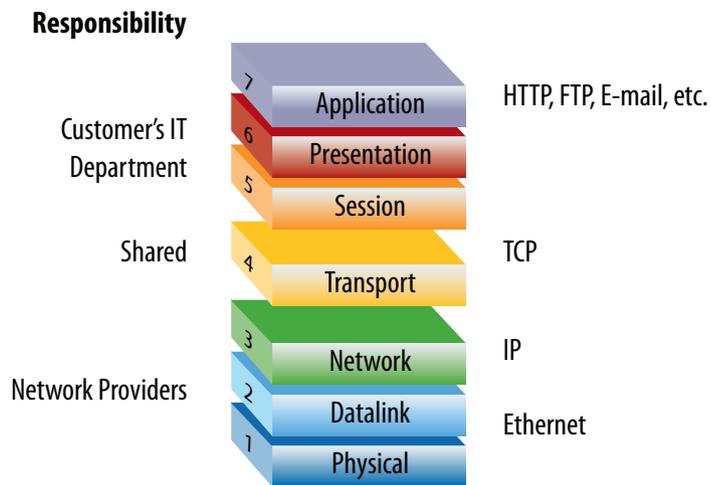


Figure 1: OSI Model

To remain competitive, network providers must continue to evolve their test capabilities as new technologies emerge. Those who can test Carrier Ethernet and IP field networks using application-level tests for Layer 4 (L4) throughput will dramatically reduce finger-pointing between the end-customer and the network provider. Figure 2 highlights an enhanced workflow for turning up L2/L3-based IP services, emphasizing application/service-level verification.



Figure 2: Workflow for Application-Centric Turn-Up

Providers must be equipped with a field communication test tool that can test various TCP parameters and optimize TCP performance against network conditions such as latency. Improper TCP configuration coupled with network latency is one of the biggest culprits resulting in poor application performance. Most applications ride on top of TCP, such as FTP and HTTP, whose performance is directly related to proper TCP performance. Providers must test TCP first, tune the TCP configuration, and then test FTP and HTTP.

The Fundamentals of TCP Window Sizing

The TCP window size refers to the amount of data bytes that a sender will place onto the network before an acknowledgment is required from the receiver. In the example in Figure 3, the TCP window is "full" after sending 17 kB which occurs in about 3 ms; no more data can be sent till the ACK is received.

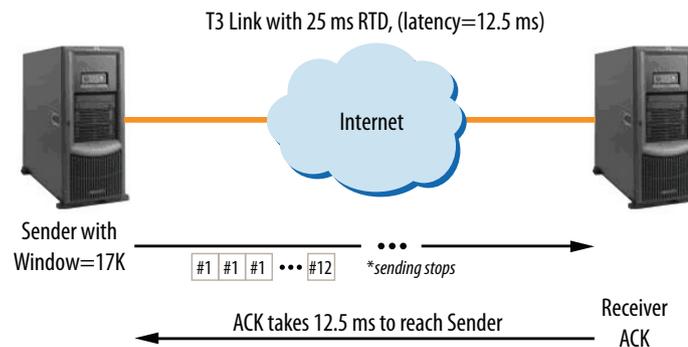


Figure 3. Improper TCP Settings Reduce Application Throughput

To further illustrate, improper TCP settings on a T3 link with 25 ms round-trip delay (RTD) can easily reduce application throughput by 80 percent (6 vs 45 Mb/s). Conducting tests with multiple window sizes, to find one that is ideal for maximum throughput, simplifies tuning and testing efforts and lets users quickly determine if low TCP throughput was the result of improper window configuration.

The primary concept for TCP throughput testing involves the bandwidth delay product (BDP). With Layer 3 tests (below the TCP layer), the latency of the network does not affect the achievable Layer 3 throughput. In other words, a 45 Mb/s link with 25 ms delay could achieve nearly 45 Mb/s at Layer 3.

But for TCP (Layer 4), the latency is key along with the TCP window size setting. Again, the TCP window size is the amount of data bytes that a sender will place onto the network before an acknowledgment is required from the receiver.

The BDP is the product of a data link's capacity (in bits per second) and its end-to-end delay (in seconds). The equation for the BDP is as follows:

$$\text{BDP} = \text{RTD} * \text{Link Speed}$$

The BDP calculation provides the TCP window as the ideal number of bits, dividing by 8 will calculate the ideal TCP window. In the example of 45 Mb/s link and 25 ms round-trip delay, a window size of ~140 kB is optimal. For frame sizes of 1514 bytes, this equates to about 95 frames sent to the receiver before a TCP acknowledgment is required.

The JDSU TCP Wirespeed Option: Simplifying the Complexity of TCP Testing

The JDSU T-BERD*/MTS-6000A supports "stateful" TCP sessions at a line rate of up to 10G and can support up to 64 TCP sessions. Another benefit of the T-BERD/MTS-6000A is that it is compatible with the commonly used *iperf* software tool, but eliminates the limitations and inaccuracies of PC-based *iperf* performance.

As mentioned in the previous section, the TCP window must be properly tuned with the latency of the network in order to achieve full throughput. The establishment of the proper window size can be a complicated and tedious task, but the JDSU T-BERD/MTS-6000A "Walk the Window" script automates the process to allow less experienced technicians to conduct the tests.

After activating the Walk the Window script, the user simply configures the starting and ending window sizes, the number of windows to test (between the starting and ending window size), and the duration of each trial test, as Figure 4 shows.

Walk the Window		
Window Sizes		
Starting Window Size (KB)	1	
Ending Window Size (KB)	100	
Number of Window Sizes	4	
Duration Each Size (secs)	15	
Server Info		
Server IP Address	192.168.11.2	
Server Listen Port	5001	
Cancel	Prev	Start

Figure 4. TCP Walk the Window Script Configuration

The test is then run between the two end-end JDSU T-BERD/MTS-6000 or a 6000A and an *iperf* server. After completion, the user is presented with an easy-to-understand test report that clearly highlights the window size versus throughput performance, as Figure 5 shows. For this test, the link speed was 1 GigE and the RTD was 1 ms. This example shows that the optimum TCP window size is 128 kB which achieves maximum Layer 4 throughput of ~950 Mb/s.

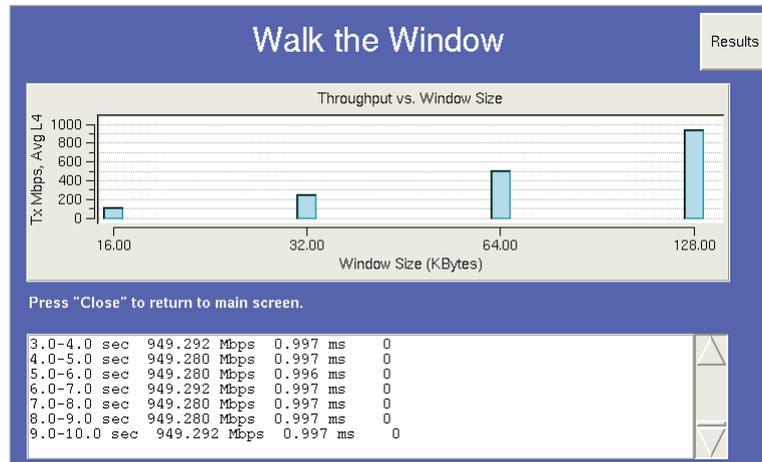


Figure 5. TCP Walk the Window Script Results

After establishing the optimum window size, TCP throughput testing should be conducted with a mixture of background traffic such as UDP to verify customer throughput under realistic network conditions. Router queue settings commonly cause customer throughput issues and cannot be diagnosed unless a network is under load. TCP Tail Drop issues are surprisingly widespread in networks and many times go undetected until a customer exceeds a particular traffic threshold.

TCP Tail Drop Explained

In properly configured routers, the buffer mechanism should be set to periodically drop packets when the router becomes congested. Routers use various "Early Discard" algorithms to alleviate congestive situations.

If a router is configured improperly, the buffer will fill and cause the router to drop the "tail" of the buffer. This drop causes a drop of a sequential block of packets that can adversely effect the built-in retransmission mechanisms of TCP. The net result is that the multitude of TCP hosts (such as servers and PCs, for example) begin to retransmit in synchronicity, which can cause an endless loop of router congestive overload and discards. The user then experiences an incredible slowdown in TCP sessions such as HTTP and, in many cases, is the cause of the long "spinning hourglass" on Windows workstations.

Automating the TCP Test Process

Even though RFC2544 was meant to benchmark network equipment and to be used by network equipment manufacturers (NEMs), network providers also use it to benchmark operational networks in order to provide Service Level Agreements (SLAs) to their business customers. Ultimately, network providers realize that a successful RFC2544 test result does not guarantee end-user satisfaction, thus requiring, an automated means for testing TCP performance.

Network providers (and NEMs) are wrestling with end-to-end complexities and must standardize the methodology to validate end-to-end TCP performance, as this is the precursor to acceptable end-user application performance. JDSU is developing a standard methodology for testing TCP layer performance and guidelines for expected TCP throughput results for observation in the network under test.

The next subsections highlight the important elements of the JDSU TCP testing methodology.

End-to-End TCP-Only Installation Testing

The T-BERD/MTS-6000A TCP Wirespeed option, when used in the configuration shown in Figure 6, provides results that characterize the achieved TCP throughput at various TCP window size settings. In addition, this tester provides resultant retransmission counts (due to packet loss) and offers much greater confidence for providers that customer applications will achieve expected throughput.

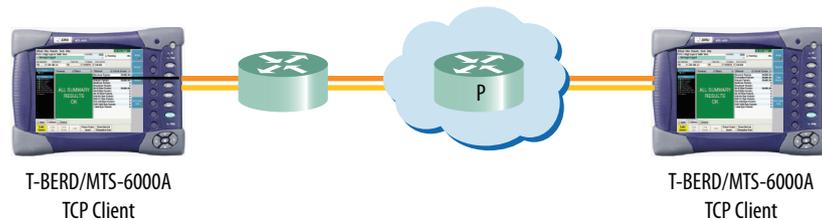


Figure 6. T-BERD/MTS-6000A acting as both the TCP Client and TCP Server

End-to-End TCP and Streams Installation Testing

Figure 7 shows how operators can verify proper CoS/QoS prioritization in the network by conducting an end-to-end test with concurrent TCP and background streams to prove that customer application sessions (TCP) will receive expected priority versus background traffic types such as UDP streams.

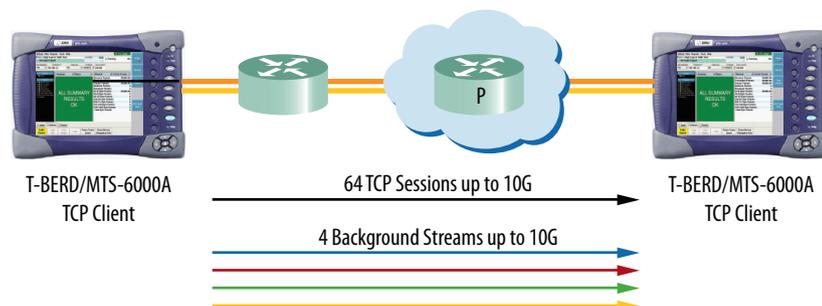


Figure 7. Using the T-BERD/MTS-6000A to Verify Proper CoS/QoS Prioritization using Background Streams

TCP Throughput Testing with an "iperf" Host

Figure 8 shows the T-BERD/MTS-6000A MSAM acting as a Client communicating with an iperf Server running industry-standard, open source TCP test software. This test lets network operators go one step further in terms of throughput segmentation by either using a test server or the actual customer server to conduct TCP throughput tests. The T-BERD/MTS-6000A can be considered a "golden" data source because its hardware TCP implementation transmits TCP data at the full line rate.

Cisco's "ttcp mode"

Most Cisco devices support the ttcp TCP throughput utility which is also compatible with the T-BERD/MTS-6000A. The 6000A test unit can be used to sectionalize TCP performance across Cisco devices as it is also compatible with ttcp.

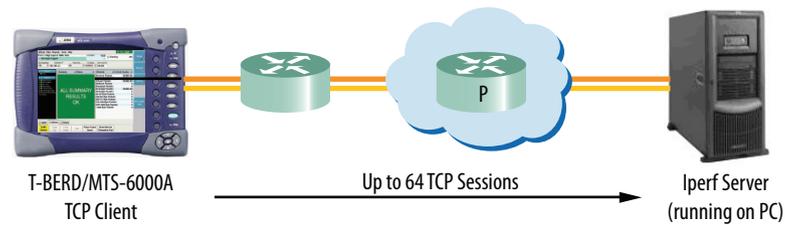


Figure 8. T-BERD/MTS-6000A MSAM as the TCP Client Communicating with an iperf Server

Conclusion

As wire speeds continue to rise and new applications emerge, networks increase in complexity. For network owners to remain competitive, they must evolve the turn-up approach as new technologies emerge. Network operators must conduct turn-up tests that emulate actual customer traffic in order to prove that their networks deliver SLAs at the TCP layer, which will eliminate finger-pointing when problems arise.

Applications such as file downloads, e-mail, and Internet access can perform poorly when TCP is running poorly. L4 tests are required to validate and ensure network SLA and often require the expertise of an expert network engineer. However, employing the use of the T-BERD/MTS-6000A can equip a Tier 1 technician to conduct TCP throughput tests at line rates up to 10G that can support as many as 64 TCP sessions. It is compatible with the industry-standard iperf software tool which allows the T-BERD/MTS-6000A to be used interoperably with iperf-based client and servers.

The T-BERD/MTS-6000A can be used to verify proper CoS/QoS prioritization, conduct TCP throughput tests, or to automatically characterize TCP performance at various TCP window sizes. Network operators and service providers who can verify performance accurately and ensure SLAs can reduce customer churn and eliminate finger-pointing when issues arise.



Test & Measurement Regional Sales

NORTH AMERICA TEL: 1 866 228 3762 FAX: +1 301 353 9216	LATIN AMERICA TEL: +1 954 688 5660 FAX: +1 954 345 4668	ASIA PACIFIC TEL: +852 2892 0990 FAX: +852 2892 0770	EMEA TEL: +49 7121 86 2222 FAX: +49 7121 86 1222	WEBSITE: www.jdsu.com/test
---	--	---	---	--