

Understanding Switch Latency

White Paper

June, 2012

Contents

<u>What You Will Learn</u>	3
<u>Overview</u>	3
<u>Different Latency Measurement Methods</u>	3
<u>How Do Measurement Tools Calculate Latency?</u>	4
<u>Switching Method: Cut-Through or Store-and-Forward?</u>	5
<u>Latency Measurement</u>	6
<u>Average Latency</u>	6
<u>Measuring the Latency of a Cut-Through Switch</u>	6
<u>FIFO and LILO Latencies of a Cut-Through Switch</u>	6
<u>LIFO Latency of a Cut-Through Switch</u>	7
<u>Measuring the Latency of a Store-and-Forward Switch</u>	7
<u>LIFO Latency of a Store-and-Forward Switch</u>	7
<u>FIFO Latencies of a Store-and-Forward Switch</u>	8
<u>Which One Is Faster - Cut-Through or Store-and-Forward?</u>	9
<u>Other Factors to Consider</u>	9
<u>Traffic Pattern and Traffic Rate</u>	9
<u>Intrinsic Latency on Measurement Tools and Cable Latency</u>	10
<u>Choosing the Right Measurement Tools and Software</u>	10
<u>Conclusion</u>	11
<u>About the Author</u>	11

What You Will Learn

Ethernet switch latency is becoming a very critical factor when the performance of the switch is evaluated, especially for high performance networking. Low latency switching is a key element to enable upper layer application to get their job done as quickly as possible, regardless it's for high frequency trading or high performance computing. Due to the advanced switching architecture design and improved silicon technology, the Ethernet switch latency is decreasing from double-digit milliseconds to sub-1 microsecond in the last several years. In this white paper, you will learn the switch latency definition, different switching methods, and other important factors that can drastically change the latency measurement results. This paper will equip you with all the necessary knowledge to understand those variables and get the most accurate latency measurement for all Ethernet switches under different configurations.

Overview

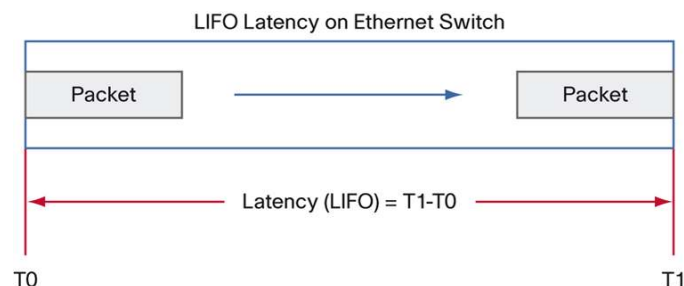
Ethernet switch latency is defined as the time it takes for a switch to forward a packet from its ingress port to its egress port. The lower the latency, the less time the packet needs to stay in the switch waiting to be processed, the faster the switch, the quicker the packets can be sent to the intended destination host - and ultimately the faster the response time the upper application can has. It's very important to understand the terminology of switch latency in various switching methods and the methodology to obtain the most accurate latency measurements. Many variables can affect the accuracy of the latency measurement results, such as the position of the time stamp in test packets, the traffic rate and pattern, etc. Sometimes an incorrect setting of a single variable can drastically change the latency measurement results. This paper will analyze all aspect of the latency measurement and help you understand all the necessary details.

Different Latency Measurement Methods

You can measure latency in four ways.

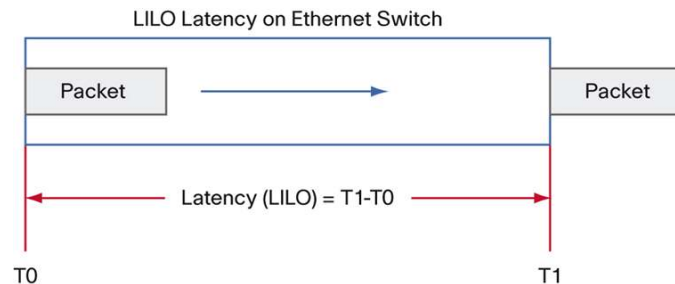
- **Last-in, first-out (LIFO):** Defined in RFC 1242, the latency measurement timer starts as soon as the last bit of the packet gets into the switch (T0) and stops when the first bit of the packet leaves the switch (T1). Figure 1 shows a diagram of LIFO.

Figure 1. LIFO Latency Measurement



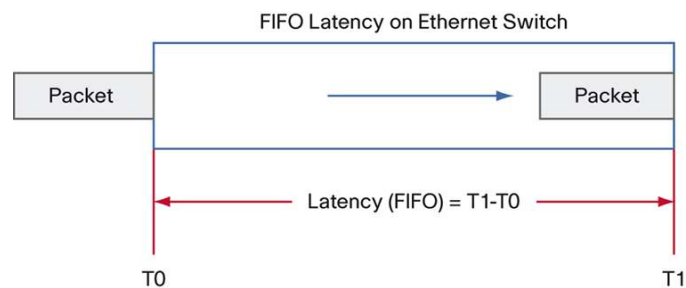
- **Last-in, last-out (LILO):** Defined in RFC 4689, the latency measurement timer starts as soon as the last bit of the packet gets into the switch and stops when the last bit of the packet leaves the switch. Figure 2 shows a diagram of LILO.

Figure 2. LIFO Latency Measurement



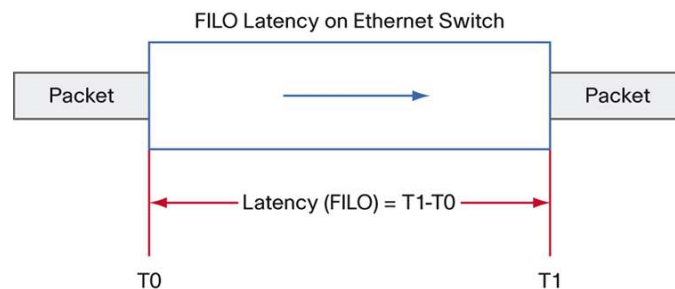
- **First-in, first-out (FIFO):** Defined in RFC 1242, the latency measurement timer starts as soon as the first bit of the packet gets into the switch and stops when the first bit of the packet leaves the switch. Figure 3 shows a diagram of FIFO.

Figure 3. FIFO Latency Measurement



- **First-in, last-out (FILO):** This method is currently not defined in any RFC. The latency measurement timer starts as soon as the first bit of the packet gets in the switch and stops when the last bit of the packet leaves the switch. Figure 4 shows a diagram of FILO.

Figure 4. FILO Latency Measurement



How Do Measurement Tools Calculate Latency?

LIFO Latency

The measurement equipment usually places the time stamp at the fixed location of every packet it generates. Spirent or IXIA put the time stamp at the end of the frame just before Frame Checksum Sequence (FCS) and will not change its location regardless of what latency measurement method is chosen, meaning that measurement tools will always measure the LIFO latency. Because all other latency measurement results are actually derived from LIFO, Then the question is how can we derive all other latency results based on the LIFO results?

FIFO Latency

Under normal circumstances, the LIFO latency is the same as the FIFO latency. This fact can be verified with the LIFO and FIFO latencies measurement results shown in a later section of this paper. Because FIFO is much more popular in all test tool configuration and documentations, going forward in this paper, we will use FIFO latency to represent both FIFO and LIFO latency measurement.

LIFO Latency

LIFO latency can be calculated as follows:

$$\text{LIFO} = \text{FIFO} - (\text{Packet size in bits/Link speed})$$

Packet size in bits/Link speed is the serialization delay, which is the time it takes to clock a packet in or out of a given transmission media. For example, the serialization delay of a 64-byte packet on a 10-Gbps link is $(64 \text{ bytes} \times 8 \text{ bits}) / 1 \times 10^9 \text{ bps} = 51.2 \text{ nanoseconds}$. Because we are calculating the time it takes for a single packet to traverse the wire when its first bit hits the ingress port, we should not add a preamble or interframe gap in the formula. This formula can also be verified by comparing the latency results between FIFO and LIFO latencies measured for the same switch (refer to Tables 1 and 2).

Table 1. Serialization Delay on 10-Gbps Link

Packet size (bytes)	64	128	256	512	1024	1024	1518	4096	9216
Serialization delay (microseconds [us])	0.051	0.102	0.205	0.41	0.819	1.024	1.214	3.277	7.373

This formula indicates that for any type of switch, the LIFO latency is always lower than all other latencies. The most interesting observation is that negative latency numbers will result if you measure the latency of a cut-through switch with LIFO.

FILO Latency

FILO latency can be derived with FIFO based on following formula:

$$\text{FILO} = \text{FIFO} + (\text{Packet size/Link speed})$$

The same link serialization delay listed in Table 1 is also applied here.

So the question is: Which one should be used? The answer depends on the switching method of the Ethernet switch, which is covered in the next section.

Switching Method: Cut-Through or Store-and-Forward?

Given that there are multiple ways to measure the switch latency, which one should be used to evaluate a switch? The answer depends on the switching method of the switch: cut-through or store-and-forward.

- **Cut-through switching:** The Ethernet switch starts to send the packet forward at its input port as soon as the destination is learned, normally at the first 6 bytes, long before the entire packet is read into the switch. Obviously this type of switch can forward packets much faster and provide lower latency. The most common way to measure latency of a cut-through switch is with FIFO.

- **Store-and-forward switching:** With this type of switching the entire packet needs to be stored in the ingress of the switch before the switch can forward it. Obviously the switch will take more time to forward the packet compared to the time for the cut-through switch, because it has to store the entire packet before it starts forwarding it. The most common way to measure latency of a store-and-forward switch is LIFO. The following sections explain the latency measurement results with a different method.

However, in order to have a consistent measurement results, the same measurement method (preferable FIFO) should be used if you're comparing the latency of different switches, regardless it's cut-through or store-and-forward switch, the details are elaborated in the later section.

Latency Measurement

Average Latency

For any given high performance switch, latency was measured among hundreds of billions packets, It make sense to use average latency to fairly represent the switch's latency characteristics, though minimum and maximum latency can also provide a different view on latency characteristics. Average latency is used to represent switch latency throughout this paper for our discussion.

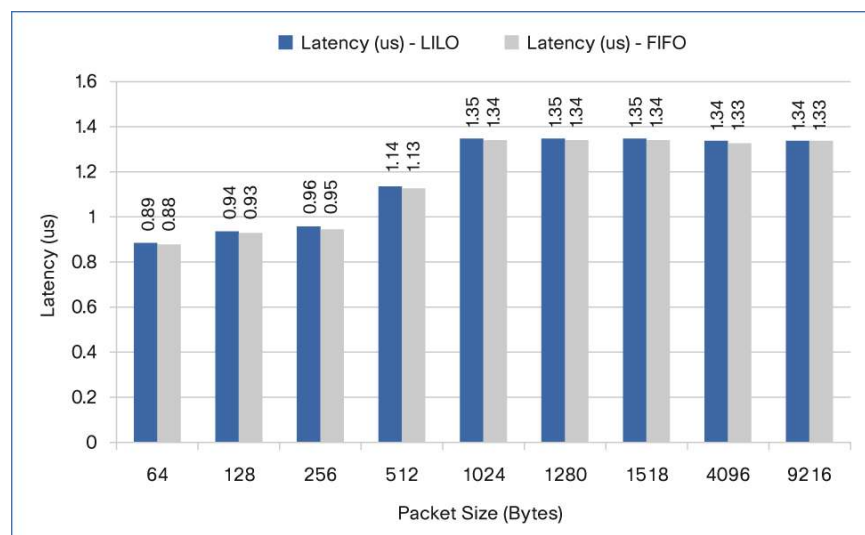
Measuring the Latency of a Cut-Through Switch

Consider the ultra-low-latency Cisco Nexus[®] 3064PQ Switch with sixty-four 10 Gigabit Ethernet ports as an example to illustrate how to correctly measure the switch latency. Four ports of the Cisco Nexus 3064PQ are connected to the Spirent Test Center; full-meshed test traffic is configured as 100-percent line rate, meaning that each port will transmit and receive traffic from all other ports at 10Gbps. Each iteration will run through a fixed packet size, starting from 64 bytes all the way to 9216-byte jumbo packets.

FIFO and LIFO Latencies of a Cut-Through Switch

Figure 5 graphs the FIFO and LIFO latencies for a cut-through switch.

Figure 5. FIFO and LIFO Latencies for Cut-Through Switch



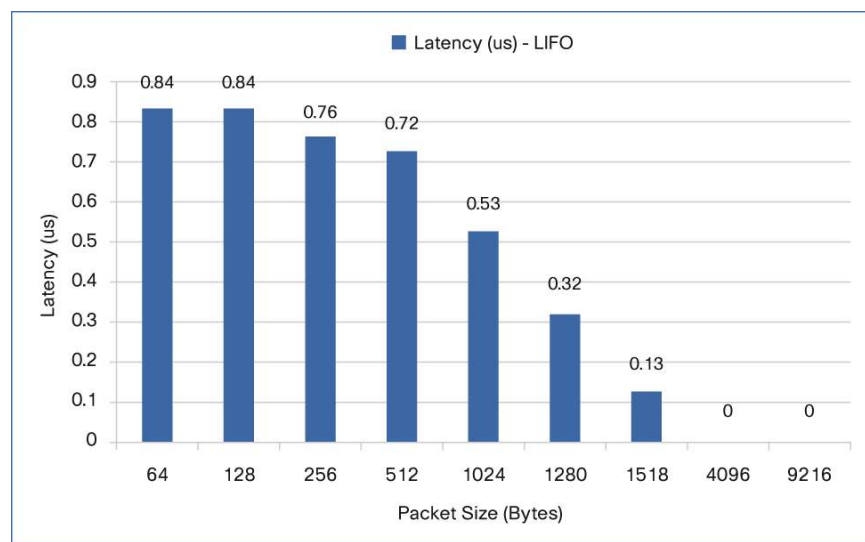
As noted earlier, the FIFO and LIFO latencies should be the same under normal circumstances; as shown in Figure 5, the difference between FIFO and LIFO latencies is only about ~10 nanoseconds, which is within the accuracy margin of the test tool. Spirent's timestamp resolution is ± 10 ns on its 10G CV cards, which is used in this paper.

LIFO Latency of a Cut-Through Switch

As noted earlier, a cut-through switch really should not be measured with LIFO; otherwise, measurement results will show negative or 0 latency. Some switch vendors claim they are racing to 0 latency, but they should use the correct test methodology to draw this conclusion. Figure 6 shows the actual results for LIFO measurement where latency becomes 0 for large packets. Those results are also matched with the formula mentioned previously:

$$\text{LIFO} = \text{FIFO} - (\text{Packet size in bits} / \text{Link speed})$$

Figure 6. LIFO Latency for Cut-Through Switch



Because the measurement tool does not report negative latency numbers, the latency results of the last two frame sizes (4096 and 9216) are shown as 0 latency. These results clearly show that LIFO should not be used to measure any cut-through switch.

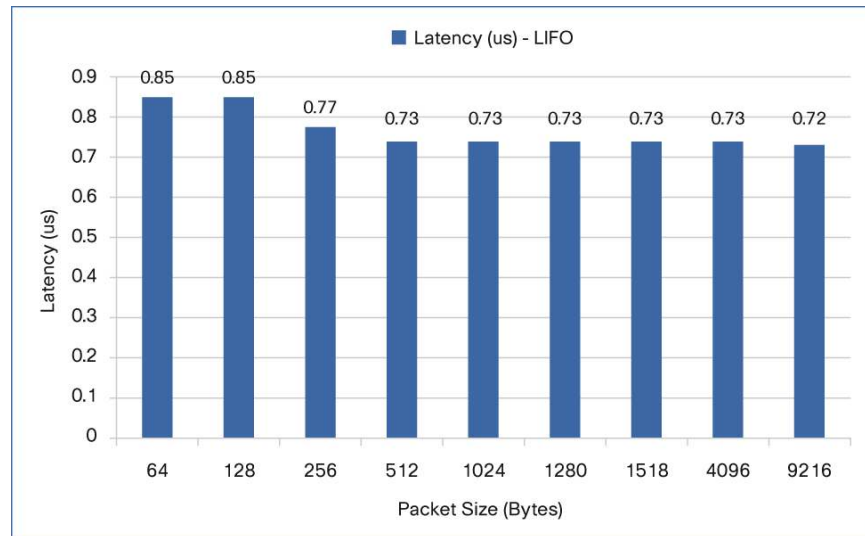
Measuring the Latency of a Store-and-Forward Switch

To demonstrate the latency measurement method of a Store-and-Forward switch, let's use the same setup as it was used in previous section, and measure the store-and-forward switch latency with different methods, as discussed in the following sections.

LIFO Latency of a Store-and-Forward Switch

The most common way to measure the latency of a store-and-forward switch is to use LIFO as defined in RFC 1242. The results are shown in Figure 7.

Figure 7. LIFO to Measure a Store-and-Forward Switch

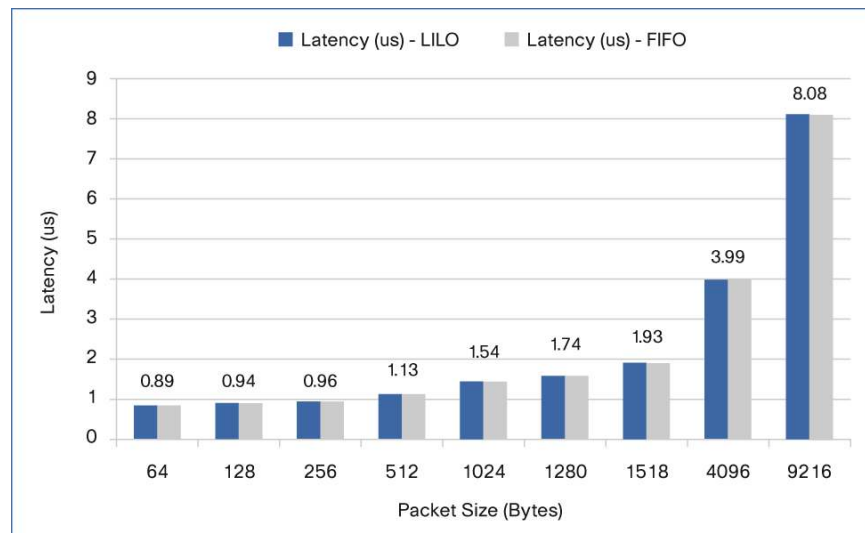


Comparing Figure 7 to Figure 5, FIFO and LIFO latencies of a cut-through switch, shows that the LIFO latency of a store-and-forward switch is less than the FIFO latency of a cut-through switch; does this result mean the store-and-forward switch is actually faster than the cut-through switch? Of course not; the reason is simply because we use LIFO to measure the store-and-forward switch and use FIFO to measure the cut-through switch, and $\text{LIFO latency} = \text{FIFO latency} - (\text{Packet size in bits} / \text{Link speed})$. To properly compare these results, we need to use the FIFO latency to compare the latency of the store-and-forward and cut-through switches.

FIFO Latencies of a Store-and-Forward Switch

In order to compare the apple to apple, we also measured the FIFO latency of this store-and-forward switch to compare the results of different measurement methods. Figure 8 illustrates the actual measurement results of FIFO latency. (As we did in previous section, we also measured the LIFO to show that FIFO and LIFO are identical in this configuration as well).

Figure 8. LIFO and FIFO Latencies of a Store-and-Forward Switch

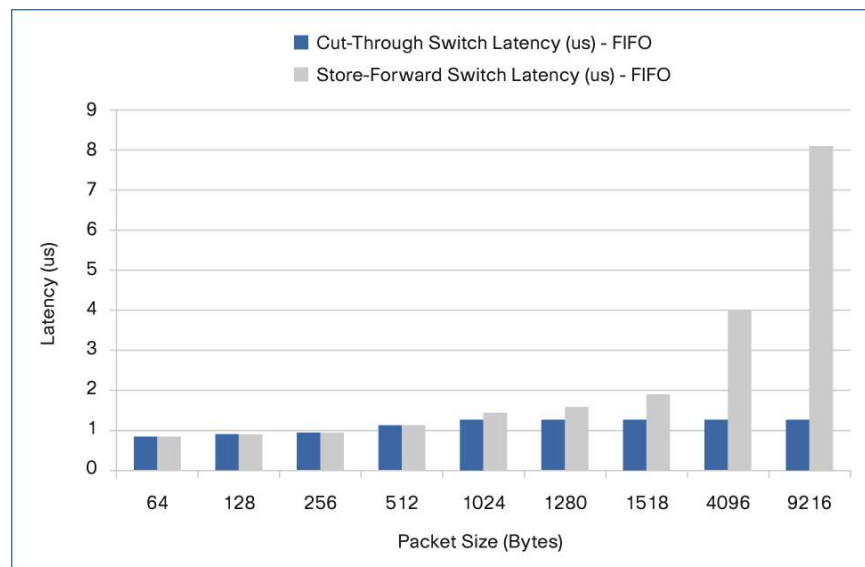


The above results clearly show the FIFO latency of a store-and-forward switch is much higher than the FIFO latency of a cut-through switch, meaning the switch needs more time to store the entire packet - especially for the large packets.

Which One Is Faster - Cut-Through or Store-and-Forward?

Comparison of the FIFO latency results of both cut-through and store-and-forward methods, shown in Figure 9, clearly reveals that the packets stay in a cut-through switch a much shorter time than they do in a store-and-forward switch - especially for large-size packets. Not only is the FIFO latency of a cut-through switch much smaller than that of a store-and-forward switch, but the “standard” LIFO latency of a store-and-forward switch is also lower than the FIFO latency of a cut-through switch. So it's very important to note, for like comparisons, it is recommend that you use FIFO to measure all switch latencies.

Figure 9. FIFO latency comparison between cut-through switching and store-and-forwarding switching



Other Factors to Consider

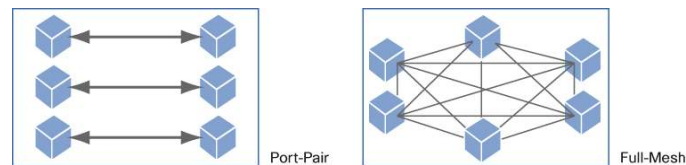
Traffic Pattern and Traffic Rate

All the previous measurements were taken at the line rate of fully meshed traffic with all 64 ports connected to test tools, this is the most stress test on the switch, and this should be the standard way to measure latency, i.e, latency need to measured at maximum throughput without congestion. The latency results will be lower if:

- Number of ports changed - change from fully loaded configuration to a only a few ports
- Traffic pattern changed - change from full-mesh to port-pair
- Traffic rate changed - change from line rate to significantly lower rate.

The diagram below illustrates the different traffic patterns:

Figure 10. Traffic pattern illustration



Depending on the switch architecture, the latency difference between most complicated configuration and most simple configuration could be huge (300-500% or even higher!). Due to the superior forwarding architecture, the latency delta on Nexus 3064PQ switch under different configuration is only ~100ns at most.

Testing equipment doesn't restrict what particular configuration should be used. Instead, it offers great flexibility on number of ports, traffic pattern and traffic rate configuration, it's important to remember that the less ports, the less traffic load and the less complicated traffic patterns, the better (lower) the latency you will get.

Intrinsic Latency on Measurement Tools and Cable Latency

Now consider what happens to the latency introduced by the transceivers, cables between measurement equipment and switches, and the measurement equipment itself? Normally, depending on the test equipment, its software version and cable length, etc, there is a ~110- to ~130-nanosecond Intrinsic latency with 2-meter fiber cable and SFP+ transceivers, that should be subtracted from the initial results if the goal is to get "absolute" switch latency. It's understandable that Intrinsic latency can also be kept in the testing results since this is what most users would see - because all 10G switch need to use transceivers and cables to connect to the network. In this case, the best practice to compare the latency of multiple switches is to move the exactly same cables and transceivers from one switch to another so all results will include the same intrinsic latency, though it will be a little bit higher than latency from the switch design specifications.

Choosing the Right Measurement Tools and Software

Measuring tools themselves could have bugs or inaccurate readings, so it is important to select the right measuring tools and their software versions. For example, due to a software bug, one test measurement tool reports ~180ns lower latency than other test tool during our lab testing, the bug was reported to test tool vendor and it was fixed.

If possible, get your results from two different measuring tools, or at least from two different software versions if this is your first latency measurement; then if the two results are quite different, something is likely wrong with one of the tools, need further investigation with the test tool vendor.

Conclusion

To get to the most accurate latency measurement of the Ethernet switch, consider the following factors carefully before evaluating the switch:

- Switching method: cut-through or store-and-forward
- Latency measurement method: FIFO or LIFO
- Traffic patterns and packet size
- Traffic rate
- Number of ports
- Testing tools and software versions
- Cable and transceiver latencies

RFCs defines different latency measurement method for cut-through and store-and-forward switching, this sometimes adds confusion to compare the latency between switches using different switching methods. When you compare two switches latency, the best practice is to use the exactly same setup during the test, practically this means only move cables from one switch to another and keep everything else the same.

In addition, as a rule of thumb for any measurement, test should be executed multiple times to get the average results.

About the Author

Yang Yang is a senior Technical Marketing Engineer currently focused on high-performance low-latency switch architecture, performance and end-to-end data center solutions. Ultra low-latency switch performance characterization is one of his recent projects.



Americas Headquarters
Cisco Systems, Inc.
San Jose, CA

Asia Pacific Headquarters
Cisco Systems (USA) Pte. Ltd.
Singapore

Europe Headquarters
Cisco Systems International BV Amsterdam,
The Netherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at www.cisco.com/go/offices.

Cisco and the Cisco logo are trademarks or registered trademarks of Cisco and/or its affiliates in the U.S. and other countries. To view a list of Cisco trademarks, go to this URL: www.cisco.com/go/trademarks. Third party trademarks mentioned are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. (1110R)