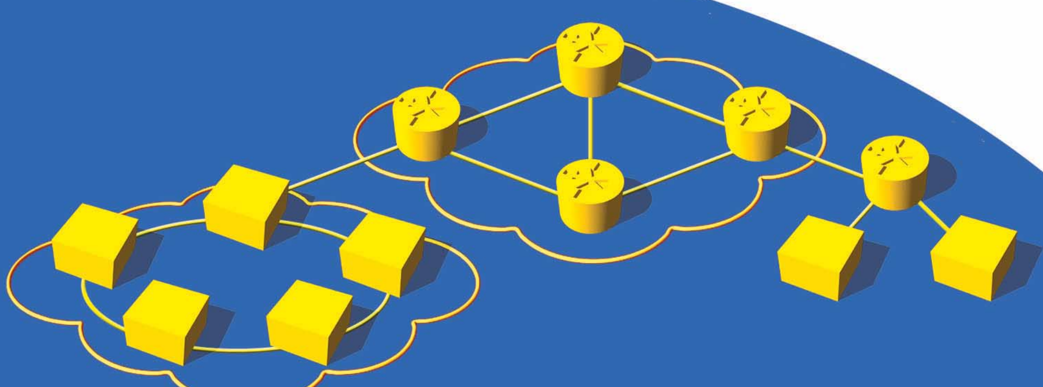
Analyze | Assure | Accelerate™

White Paper

How to Test 10 Gigabit Ethernet Performance

May 2005



P/N 71-000047 REV A

Spirent Communications, Inc.

26750 Agoura Road
Calabasas, CA
91302 USA

Email: productinfo@spirentcom.com
Web: http://www.spirentcom.com and http://www.spirentcom.com.cn (China)

North America

+1-800-927-2660

Europe, Middle East, Africa

+33-1-6137-2250

Asia Pacific

+852-2511-3822

All Other Regions

+1-818-676-2683

How to Test 10 Gigabit Performance

This White Paper explains how to test 10 Gigabit Ethernet performance, recognizing the factors and influences of 10GbE throughput and describing how to evaluate a system's reported 10GbE test results.

Contents

Overview

Spirent Communications has collaborated with the leading manufacturers of 10 Gigabit Ethernet (10GbE) network equipment around the world in the development of the most advanced systems available. The information presented here is a result of our experience in evaluating network devices for maximum performance. We will clarify here for the network community, the contributing factors and influences of 10GbE throughput and how to determine if your 10GbE test system is reporting accurate test results.

Theoretical Maximum Rate

There are two important concepts related to 10GbE performance: frame rate and throughput. The MAC bit rate of 10GbE, defined in the IEEE standard 802.3ae, is 10 billion bits per second. Frame rate is a simple arithmetic calculation based on the bit rate and frame format definitions. Throughput, defined in IETF RFC 1242, is the highest rate at which the system under test can forward the offered load, without loss. Manufacturers can claim line-rate throughput only if their switch forwards all the traffic offered at the 10Gb/s line rate for the entire duration of the test.

The bit rate at which 10GbE Media Access Layer (MAC) operates, 10 billion bits per second, is only one of the parameters in defining the transmission rate for this important new technology. The usual description of true network performance is frame rate, which indicates how many Ethernet frames are moving across the network. The maximum frame rate for 10GbE is determined by a formula that divides the 10 billion bits per second by the preamble, frame length, and inter-frame gap fields, expressed in bits.

The maximum frame rate is calculated using the minimum values of the following parameters, as described in the IEEE 802.3ae standard:

- Preamble - 8 bytes * 8 = 64 bits
- Frame length - 64 bytes (minimum) * 8 = 512 bits
- Inter-frame gap - 12 bytes (minimum) * 8 = 96 bits

Therefore,

$$\text{Maximum Frame Rate} = \frac{\text{MAC Transmit Bit Rate}}{(\text{Preamble} + \text{Frame Length} + \text{Inter-frame Gap})}$$

$$= 10{,}000{,}000{,}000 / (64 + 512 + 96)$$
$$= 10{,}000{,}000{,}000 / 672$$
$$= 14{,}880{,}952.38 \text{ frame per second (fps)}$$

Why are Other Rates Also Mentioned in 802.3ae?

In addition to the MAC layer rate at 10Gb/s, there are two bit rates defined in the IEEE 802.3ae standard: the 10GBASE-R which operates at 10.3125 Gb/s line rate, which is used in LAN applications, and the 10GBASE-W, which operates at 9.95328 Gb/s line rate and is used in SONET/SDH formats in WAN applications. In this paper, we will only discuss performance for 10GBASE-R.

Inter-Frame Gap and How it Affects 10GbE Transmission Rates

The 10GbE specification says that the Inter-Frame Gap (IFG), must fall on a byte boundary. In other words, the number of bits in the IFG must be a multiple of eight. To test the capacity of a system to handle different frame rates, it is common to keep the data field size at a constant value and vary the IFG.

Due to the 8-bit boundary rule for IFG, it is not possible to vary the offered frame rate in continuous steps; it can be incremented in discrete steps only. The test tool has a minimum resolution for the traffic frame rate it can transmit. For example, starting from the 100% frame rate, the next possible frame rate is 98.82%. It is impossible to transmit at any rate between 98.82% and 100%. Here is how to calculate transmit rates for values less than 100%.

As shown above, the shortest frame period is 672 bits, yielding a frame rate of 100%, or 14,880,952 fps.

The next shortest frame period is 680 bits, achieved by adding 8 bits to the IFG. Using the frame rate calculator shown above, it yields a line rate of 98.82%. The frame rate of 98.82% can be calculated by dividing 10Gb/s by 680, as follows:

$$10,000,000,000/680 = 14,705,882 \text{ fps}$$

Since the minimum discrete rate step is 8 bits, we can also calculate the minimum discrete step in frame rate:

$$14,880,952 - 14,705,882 = 175,070 \text{ fps}$$

So far, we have shown how to vary frame rates by keeping the data constant at 64 bytes while varying the IFG.

Another common way to test Ethernet devices is to use a variety of data frame lengths to assure performance consistency in real-world conditions. One recommendation frequently referenced is Internet RFC 2544, *Benchmarking Methodology for Network Interconnect Devices*. The RFC 2544 identifies seven frame sizes for this purpose: 64, 128, 256, 512, 1024, 1280, and 1518. When we apply these frame sizes with the minimum IFG of 96 bits, we can determine the nominal frame rate for each frame size. See *Table 1 on page 6* for the nominal frame rate for each frame size. Note that the simple addition of one or two bytes of IFG can reduce the frame rate by thousands of frames per second.

Table 1. Maximum Frame Rates for Each Frame Size

Frame Size	Frame Rate (FPS)		
	12 byte IFG	13 byte IFG	14 byte IFG
64	14,880,952	14,705,882	14,534,884
128	8,445,946	8,389,262	8,333,333
256	4,528,986	4,512,635	4,496,403
512	2,349,624	2,345,216	2,340,824
1024	1,197,318	1,196,172	1,195,029
1280	961,538	960,799	960,061
1518	812,744	812,216	811,688

The Effect of Deficit Idle Count

Many 10GbE implementations use the XAUI (pronounced “zowie”) interface as defined in Clause 47 of IEEE 802.3ae. XAUI provides a well-structured way to connect modular transceivers, such as those used in XENPAK, X2, and XPAK Multi-Source Agreements, to the XGMII interface via an extender. XAUI provides four "lanes" of electrical connection between the 10GbE MAC and the Physical Sublayer, called Lanes 0 through 3, each operating at 3.125 Gb/s, which includes overhead. *Figure 1 on page 7*, taken from the IEEE 802.3ae, Clause 46, provides a graphical representation of the sublayers that make up the 10GbE physical layer, including XAUI.

At the XGMII transmit interface, it is mandatory to align the Start of Frame (SOF) at Lane 0. This is an inefficient mechanism because the system may have to insert idle characters while waiting to place the SOF in Lane 0. Using this technique, it is possible that the transmission rate will be lower than the maximum 10Gb/s.

An alternative implementation is to add a Deficit Idle Count (DIC). DIC adds or subtracts up to 3 bytes to/from the nominal 12-byte inter-frame gap in order to maintain the 10GbE frame rate. Therefore, the minimum gap at the XGMII transmit interface can be 9 to 15 bytes[1], but averages 12 bytes. In order to achieve the maximum 10Gb/s throughput, test tools and devices under test must have DIC enabled.

[1]IEEE 802.3ae para. 46.3.1.4

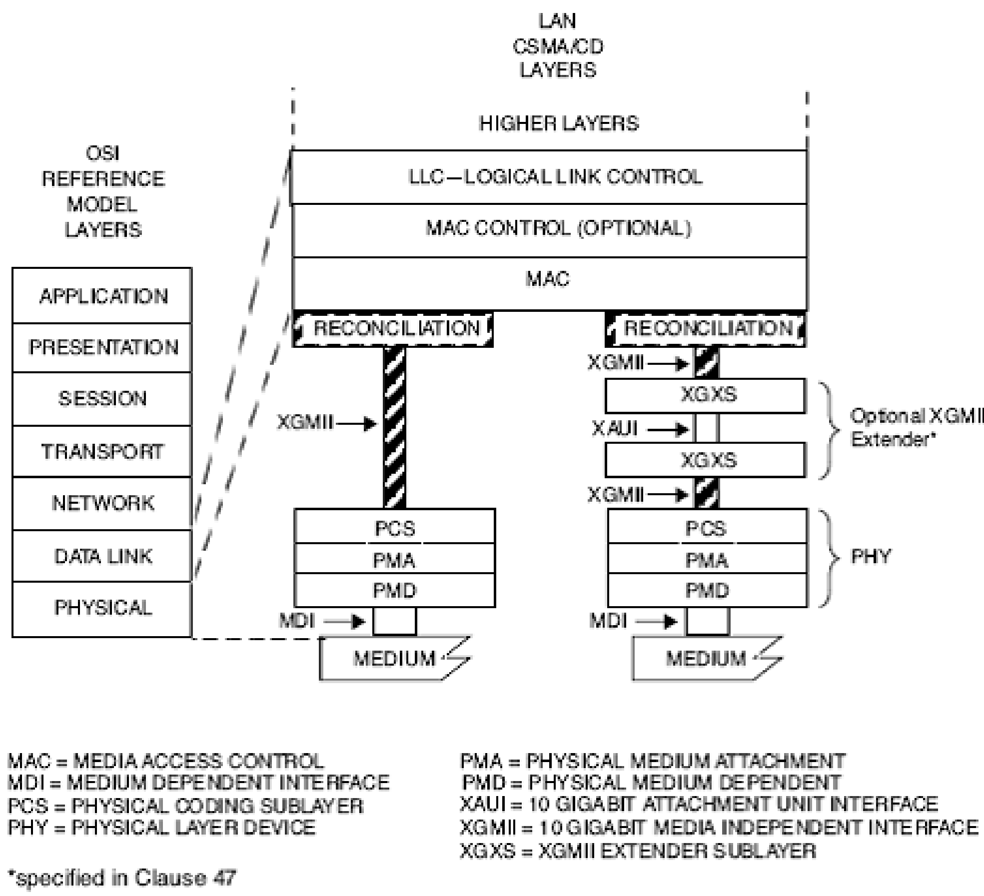


Figure 1. Sublayers that Compose the 10GbE Physical Layer

The data in Table 2 illustrates the possible 10GbE frame rates between 90% and 100%, when calculated using 64-byte frames and varying the IFG. Although it might be interesting to test transmission rates in .5% increments, or even smaller, the finest possible resolution in offered load is slightly greater than 1%. Therefore, the frame rates in *Table 2 on page 8* are the only possible rates when testing network device throughput. A test tool may allow the operator to configure a transmit rate such as 98.5%, but it will actually transmit at 98.82% or 100%, which are the nearest possible frame rates allowable due to the 8-bit IFG boundaries defined in the IEEE 802.3ae standard. This has implications for the test tool that are explained below.

Many suites of tests are configured to measure throughput using transmission frame rates varied in 5% or 10% load increments. Many networking devices operate well at throughputs up to 95% of the maximum frame rate. As the test approaches 100% frame rate, it is beneficial to transmit at each of the rates shown in *Table 2 on page 8* in order to accurately characterize system performance.

Table 2. Possible Frame Rates when Testing Network-Device Throughput

FPS	% of Max
14,880,952	100%
14,705,882	98.82
14,534,884	97.67
14,367,816	96.55
14,204,545	95.45
14,044,944	94.38
13,888,889	93.33
13,736,264	92.31
13,586,957	91.30
13,440,860	90.32

The Need to Test Throughput at Line Rate

The IEEE 802.3ae standard defines the bounds within which all standard compliant equipment must operate. 10GbE devices are compliant with the standard if they transmit at 10Gb/s ± 100 parts per million (ppm), which is a tolerance of ± .01%. Equipment manufacturers, in establishing the throughput for their equipment, may be tempted to exercise their equipment with test tools operating at less than the nominal 10Gb/s rate, but within ± 100 ppm. Unfortunately, this technique will fail throughput tests in the field where there is no control over the performance of the attached network devices that may be operating at rates that are above the nominal 10Gb/s rate up to + 100 ppm. There is no way to predict the exact performance of every attached device in a customer network. If a device operates within the nominal 10Gb/s ± 100 ppm rate band, it must be able to operate over a sustained period of time.

Table 3 on page 9 shows the frame rates for 64-byte frames when the bit rate of the transmitter is different from the nominal 10Gb/s. All the frame rates in this table are within the required ± 100 ppm, however, the frame rates vary by nearly 3,000 frames per second between the upper and lower limits. This variety of frame rates may be observed at any 10GbE network device in the field, even though the nominal 14,880,952 fps is expected. It is important that any switch intended to operate at maximum frame rate throughput be able to accept and forward all received traffic. For example, if the switch is tested only at a rate which is 100 ppm below the nominal frame rate, in the laboratory, it may face frame rates at 100 ppm above the nominal value and must still forward all frames without loss.

Table 3. Frame Rates for 64-byte Frames when the Transmitter Bit Rate is Not 10Gb/s

Clock Rate	Frames/Second
10G + 100 ppm	14,882,440
10G + 80 ppm	14,882,143
10G + 60 ppm	14,881,845
10G + 40 ppm	14,881,548
10G + 20 ppm	14,881,250
10G	14,880,952
10G - 20 ppm	14,880,655
10G - 40 ppm	14,880,357
10G - 60 ppm	14,880,060
10G - 80 ppm	14,879,762
10G - 100 ppm	14,879,464

Some test plans may try to prove maximum throughput for the network equipment under test by forcing the test tool to transmit at rates running slower than the nominal 10Gb/s. Tests run in this fashion create a false sense of security because a network device that cannot forward network traffic at the full 10Gb/s rate will drop frames in an IEEE 802.3ae compliant network. This type of test configuration does not represent a realistic network environment, because you cannot control the speed of client nodes and other attached network devices. For instance, in the network diagram shown in *Figure 2 on page 10*, if Nodes A and B had been tested only at 14,879,464 fps (10Gb/s minus 100 ppm) frame rate, they could interoperate flawlessly. However, if Node C, that transmits at 14,882,440 fps (10Gb/s plus 100 ppm), is added to the network, a network overload will occur and frames will be dropped if Nodes A and B cannot sustain throughput at the higher frame rates.

Allowing the test tool to transmit at a rate below the nominal 10Gb/s rate, even though it is within the allowable tolerance band of ± 100 ppm, may defeat the purpose of detecting a faulty network device that is intended to operate within the framework of the IEEE 802.3ae standard.

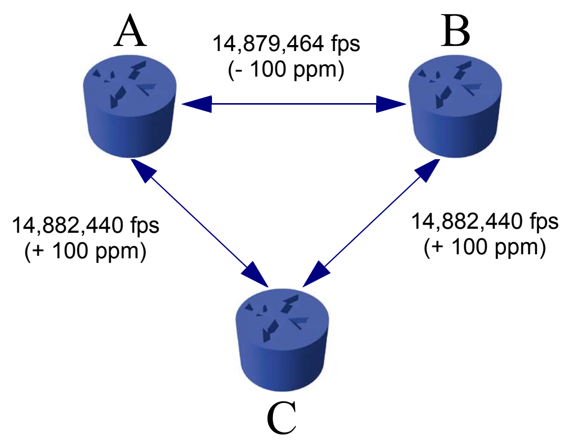


Figure 2. Network Diagram Showing 10GbE Node Interaction

Summary/Conclusion

Spirent Communications test equipment has been used to reliably evaluate the performance of hundreds of 10GbE network devices. We defined the MAC bit rate, the frame rate, and their correlation in 10GbE systems. We also discussed IFG constraints due to both byte boundary and XAUI lane alignment. It has been shown how the DIC is used to overcome the constraint of the XAUI lane alignment and how DIC is required when testing 10GbE system throughput. We have defined the possible frame rates at 10GbE and what can happen to equipment interoperability when a 10GbE network device is only tested below the nominal 10Gb/s rate. Network equipment operating at 10Gb/s must have test plans that consider the correlation between the MAC bit and frame rates.

The purpose of the test equipment is to subject the network device to the full 10Gb/s rate to find potential errors in frame reception and forwarding. Throughput tests that are run below the nominal 10Gb/s rate may provide results that are within the standard specification but will not assure interoperability in an IEEE802.3ae compliant network. Spirent Communications recommends that network equipment manufacturers of 10GbE devices test at the full 10Gb/s rate to assure interoperability in live networks.

Glossary/Acronyms/Abbreviations

- **10GbE**—10 Gigabit Ethernet
- **FPS**—Frames per second
- **Frame period**—The total time to transmit an Ethernet frame plus the interframe gap.
- **Gb/s**—Gigabits per second
- **IFG**—Inter-Frame Gap

- **MAC**—Media Access Control, IEEE 802.3ae
- **PHY**—Physical Layer, IEEE 802.3ae
- **ppm**—Parts per million
- **RFC**—Request for Comment, the internet "standards" administered by the Internet Engineering Task Force.
- **XAUI**—10 Gigabit Attachment Unit Interface, IEEE 802.3ae
- **XENPAK—**Inter-Frame GapModular 10GbE transceiver www.xenpak.org
- **XGMII**—10 Gigabit Media Independent Interface, IEEE 802.3ae
- **XGXS**—Inter-Frame GapXGMII Extender Sublayer - IEE 802.2ae
- **X2**—Modular 10GbE transceiver www.x2msa.org

References

- Bradner S. (1991). *Benchmarking Terminology for Network Interconnection Devices*. Retrieved from the The Internet Engineering Task Force web site on May 16, 2005. http://www.ietf.org/rfc/rfc1242.txt?number=1242
- The Institute of Electrical and Electronics Engineers, Inc. (2002, August). *Media Access Control (MAC) Parameters, Physical Layers, and Management Parameters for 10 Gb/s Operation*. Published by the IEEE computer society, LAN/MAN standard committee. New York, NY.
- The Internet Engineering Task Force (2005). A large open community dedicated to the evolution of the Internet. http://www.ietf.org/

About the Authors

Rick Rabinovich, Hardware Technology Director, Associate Fellow, Spirent Communications

Rick Rabinovich has more than 27 years of experience in board and system design. He joined Spirent Communications in 1997 and has been responsible for designing many critical hardware projects for the company, including a XAUI interface card, for which he holds a patent. Rick is a Spirent Engineering Associate Fellow and a Spirent Technical Ladder Associate. Rick represents Spirent at the IEEE802.3 Working Group as a voting member, and at the XFP MSA as a contributing member.

Rick was a speaker at a High-Speed Computer Design Conference for Hewlett-Packard during a South East Asia tour. He is the author of two EDN Main Design features articles, as well as other articles.

Prior to joining Spirent, Rick was a senior electronic engineer at Telematics International and a lead electronic engineer at Ascom Timeplex. Rick has a Bachelor of Science in Electrical Engineering from Buenos Aires University and has completed post-graduate courses in computer design at California State University, Los Angeles, and University California, Irvine.

John Reidy, Product Manager, Spirent Communications

John is responsible for the 10GbE business activities at Spirent Communications. He has been involved in data communications for 25 years, holding positions in product support, sales and marketing.

Charles Seifert, Director of Product Management, Spirent Communications

Charles joined Spirent Communications in 2000 and participated in the team that delivered the world's first 10GbE test tool in 2001. This product won the 2002 InfoWorld Technology of the Year award for 10 Gigabit Ethernet Products. Today, Charles manages Spirent's LAN Switching Performance Analysis-Broadband group for Ethernet test products.