# GUARANTEED END-TO-END LATENCY THROUGH ETHERNET

**Øyvind Holmeide, OnTime Networks AS, Oslo, Norway**
oeyvind@ontimenet.com
**Markus Schmitz, OnTime Networks LLC, Texas, USA**
markus@ontimenet.com

**Abstract:**

Latency sensitive data in a Flight Test Instrumentation (FTI) system represents a challenging network requirement. Data meant for the telemetry link sent through an on-board Ethernet network might be sensitive for high network latency. Worst case latency through the on-board Ethernet network for such data, might be as low as a few hundred microseconds. This challenge is solved by utilizing the Quality of Service (QoS) properties on Ethernet FTI switches.

This paper describes how to use Ethernet layer 1, layer 2 or layer 3 QoS principles of a modern Ethernet FTI network.

**Keywords**: QoS, IEEE802.1p, IP ToS/CoS, FTI.

## Introduction

Ethernet in a Flight Test Instrumentation (FTI) system is a fairly new approach. Even though Ethernet is widely accepted for FTI systems, the use of Ethernet for FTI remains fairly simple. Advanced Ethernet QoS techniques are used only to a small extent today. QoS can, if configured properly, guarantee a worst-case latency of less than hundred microseconds for latency sensitive data.

This paper targets the following FTI network challenge:

- How to guarantee worst-case latency for latency sensitive data?

**Abbreviations:**

| | |
|---|---|
| CoS | Class of Service |
| DSCP | Differentiated Services Code Point |
| FCS | Frame Check Sequence |
| IED | Intelligent Electronic Device |
| IP | Internet Protocol |
| IPG | Inter-Packet Gap |
| MAC | Medium Access Control |
| QoS | Quality of Service |
| RTOS | Real Time Operating System |
| UDP | User Datagram Protocol |
| TCP | Transmission Control Protocol |
| ToS | Type of Service |

**How to guarantee worst-case latency for latency sensitive data?**

Ethernet switches may have support for priority containing two or more output queues per port, where the high priority queue(s) are reserved for latency sensitive critical data offering best possible QoS for such data. Relevant packet scheduler schemes for an Ethernet switch with four priority queues can be:

1. Round-robin weighting; i.e. N-highest (Priority Level 3) packets are sent from the highest priority queue, before N-high (Priority Level 2) packets are sent from the high priority queue, before N-low (Priority Level 1) packets are sent from the low priority queue, before N-lowest (Priority Level 0) packets are sent from the lowest priority queue. The packet scheduler will move directly to the next priority queue in the chain if no packets are present in the given queue.
2. Strict priority scheduling. i.e. all available packets in the highest priority queue will be transmitted from the highest priority queue before any of the lower priority queues are served. Thus, packets from a queue will only be sent if all higher priority queues are empty.

Note that a high priority packet also will be delayed due to a low priority packet if the transmission of the low packet is started before the high priority packet enters the egress port. The high priority packet will then be delayed by the time it takes to flush the rest of the low priority packet. Worst case will be that the transmission of a low priority packet with maximum packet length (1518 bytes) is just started when a high priority packet arrives the given egress port. The extra switch queuing delay will then be 122µs in case of 100Mbps egress port speed, and 12µs in case of 1Gbps port speed.

A high priority packet may also be delayed through the switch due to other latency sensitive packets that are already queued for transmission in the same high priority queue for a given egress port. It is, however, often a straightforward job to calculate the worst-case switch latency such a packet may experience if the network load and traffic pattern of the latency sensitive applications using the high priority queues are known, and all other traffic in the network have lower priority. Typical worst-case switch latency for a high priority packet in such a system will be a few hundred µs through each network hop in case 100Mbps is used on the egress port and less than 50us case 1Gps is used on the egress port.

Example 1:

- 100 Mbps with full duplex connectivity is used on all drop links.

- The switch is a store-and-forward switch, with a minimum switch latency of 10µs.

- The switch uses strict priority scheduling.

- The latency sensitive packet has a length of 200 bytes including preamble, MAC, IP, UDP, payload, FCS and minimum IPG. The latency sensitive packets are treated as high priority packets, all other packets have less priority.

- Up to five other end nodes may generate similar latency sensitive packets of 200 bytes that may be in the same priority queue before the packet enters the queue, and causes extra switch delay.

- All latency sensitive packets are generated in a cyclic manner.

The worst case switch latency of a latency sensitive packet will then be:

1.) 16µs, store-and-forwards.

2.) 10µs, minimum switch latency.

3.) 122µs, worst case latency due to flushing of a packet with maximum packet length.

4.) 80µs, five latency sensitive packets already in the same priority queue.

5.) 228µs, total.

Example 2:

Same as above, but with 1Gps rate on the egress port. The worst-case switch latency of a latency sensitive packet will then be:

1.) 16µs, store-and-forwards.

2.) 10µs, minimum switch latency.

3.) 12µs, worst-case latency due to flushing of a packet with maximum packet length.

4.) 8µs, five latency sensitive packets already in the same priority queue.

5.) 46µs, total.

Example 3:

Same as above, but with 1Gps rate and rate shaping set to 256Mbps on the egress port. The worst-case switch latency of a latency sensitive packet will then be:

1.) 16µs, store-and-forwards.

2.) 10µs, minimum switch latency.

3.) 48µs, worst-case latency due to flushing of a packet with maximum packet length.

4.) 31µs, five latency sensitive packets already in the same priority queue.

5.) 105µs, total.

These three examples represent worst-case latency for the latency sensitive packets identified as high priority packets. These estimations are valid regardless of any other network load with less

priority in the network. Several priority implementations exist with respect to how a packet is identified as a high priority packet. The priority handling depends on the switch functionality.

**Layer 2 priority**

A layer 2 switch performs switching based on the Ethernet MAC destination addresses, see Figure 1.


Figure 1, MAC header (layer 2), no VLAN tag

A layer 2 switch may provide priority information based on:

MAC addresses. Both the MAC source- and destination address can be used for priority identification, see Figure 1. This is not a very flexible feature.

Ethernet port. One or multiple ports of the switch can be configured for high priority. This means that all packets received on these ports will be treated as high priority packets. The technique requires a static setup and all packets received on a given port will be treated with the same priority.

Priority tagging. The IEEE 802.1p (and IEEE 802.1Q) standard specifies an extra field for the Ethernet MAC header. This field is called Tag Control Info (TCI) field, and is inserted between the source MAC address and the MAC Type/Length field of an Ethernet packet (see Figure 2). This field contains a three bit priority field that is used for priority handling. These three priority bits map to the priority queues of the switch. The mapping depends on the number of queues the switch supports. For example: priority field = 111 will map to priority queue 7 on a switch with 8 priority queues, while priority field = 111 and 110 will both map to priority queue 3 on a switch with four priority queues. Both unmanaged and managed switches can support this feature. Thus, no switch configuration is needed. A disadvantage with this method is that most end nodes do not support VLAN tagging. Configuring the switch to remove the tag after switching can solve this, and should be done before the packets are sent on the output ports, where stations without support for this feature are connected. This requires managed switch operation. Another problem could be that other existing Ethernet switches in the network do not support priority tagging. The maximum Ethernet packet size will, due to the VLAN tag, increase by four bytes to 1522.

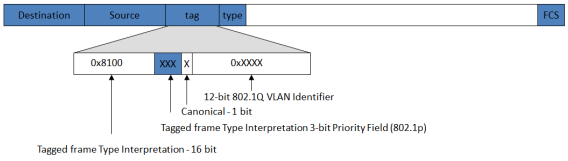


Figure 2, MAC header (layer 2) with VLAN tag

**Layer 3 priority**

A layer 3 switch can perform packet switching based on both the Ethernet MAC destination addresses and the layer 3. E.g. the header fields of IP packets.

A layer 3 switch may provide priority identification based on the same criteria's as a layer 2 switch. The following layer 3 field is also relevant:

IP ToS/Cos. Each IPv4 header contains a ToS/CoS field, see Figure 3. The RFC standards known as Differentiated Services see RFC 2474, partition the ToS/CoS field into two fields: DSCP (6 bit) and CU (2 bit). The DSCP field is used to determine the required priority. The 6 bit of the DSCP field represents 64 possible "code points" that is split in three pools:

-Pool 1 DSCP = [0 .. 31] reserved for standard actions (e.g. VOIP)

-Pool 2 DSCP = [32 .. 47] reserved for experimental or local use, but may be allocated for standard actions in the future.

-Pool 3 DSCP = [48 .. 63] reserved for experimental or local use.

Any subset of the 64 possible code points can be used as a high priority identification criterion in the switch. The high priority code points should preferably be user configurable. The code points from Pool 3 are the preferred alternative for a given nonstandard IP based real time application. F.ex. an FTI UDP stream.

High priority setting of the IP ToS field of real time critical packets must be set in the IP protocol of the sending station. This can be done on TCP/UDP socket level by a setsockopt( ) command both on the client and server socket side in most Operating Systems (OS).

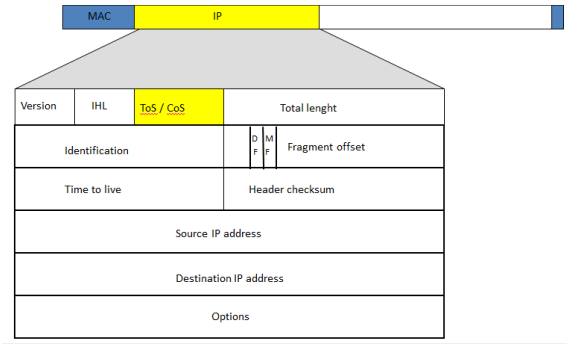


Figure 3, IPv4 header (layer 3)

An IPv6 header contains a corresponding field called Traffic Class. This field has the same function as the ToS/Cos field of IPv4. The Traffic Class octet has the same location in the IPv6 header as the ToS field has in the IPv4 header.

**Measurement results**

Measurement of the network latency for latency sensitive data configured as high priority packets that are sent through an FTI network consisting of three network hops is shown in Figure 4.

The transmit time of the latency sensitive data is measured at PHY level on PC 1 on the setup and the corresponding receive time of the same data is measured at PHY level on PC 2. Both PC1 and PC 2 are connected to the FTI network through 100BASE-T(x) links. A gigabit tester is used as a traffic generator in the system. Six data streams of low priority data are generated with infinite duration. Each stream is based on test packets of 1450 bytes with only 100ns inter-packet gap. The test packets are sent on 1000BASE-T(x) ports. This means that each stream is close to 1Gps. The data streams are sent to/from the tester according to the following setup:

- data stream 1 (red) is sent from port 1 to port 2 on the tester through all three FTI switches (downlink)
- data stream 2 (red) is sent from port 2 to port 1 on the tester through all three FTI switches (uplink)
- data stream 3 (green) is sent from port 3 to port 4 on the tester through the two top most FTI switches (downlink)
- data stream 4 (green) is sent from port 4 to port 3 on the tester through the two top most FTI switches (uplink)
- data stream 5 (red) is sent from port 5 to port 6 on the tester through the two lower FTI switches (downlink)
- data stream 6 (red) is sent from port 6 to port 7 on the tester through the two lower FTI switches (downlink)

This means that the amount of test data sent on each of the two links that connect the three FTI switches exceeds the bandwidth with almost 100% in both directions since the tester sends close to 2Gbps of data over the 1000BASE-T(x) links. The FTI switches will therefore drop almost 50% of all packets queued for these ports, and the switch output queues will continuously be full.

Latency sensitive data is sent from PC1 to PC2 through a client/server UDP socket application, where the socket is configured for high priority; i.e. the sending end node (PC1) configures the socket with the setsockopt( ) command with TOS field set to e.g. 0xF8.

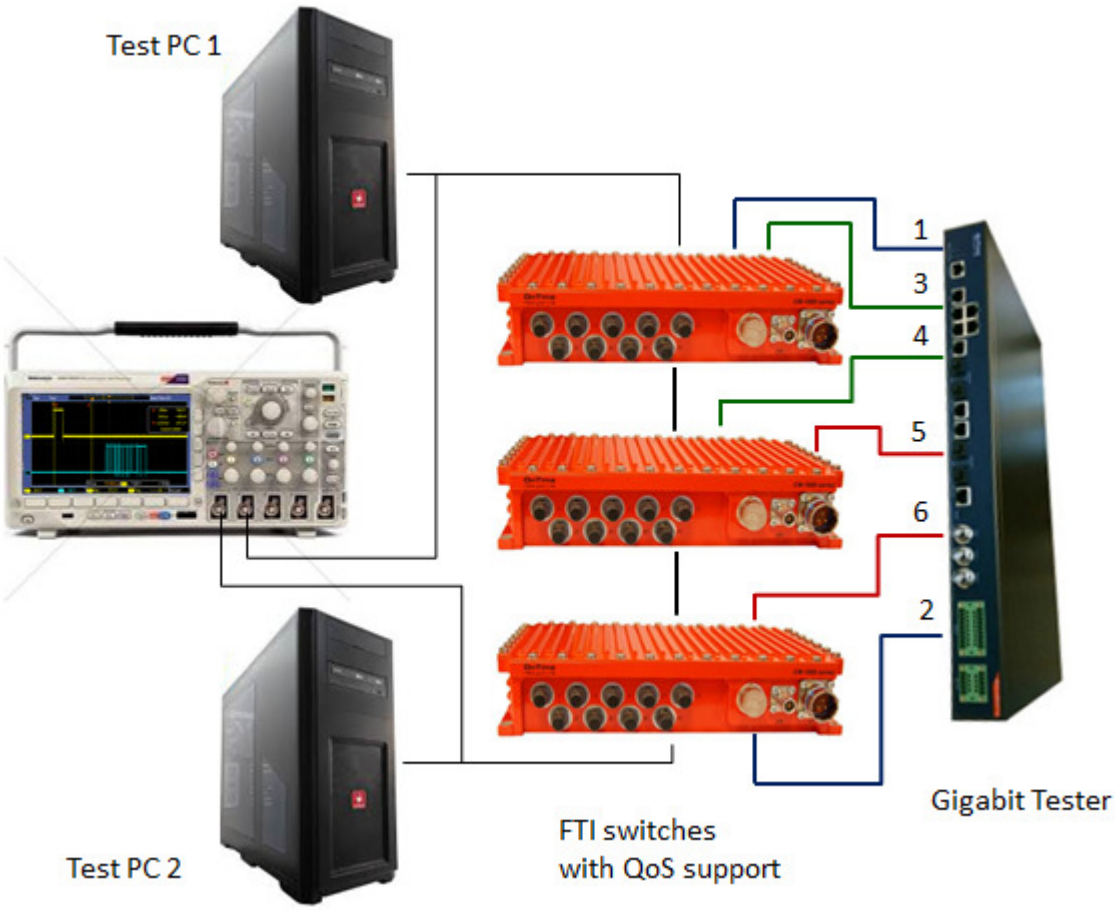


Figure 4, Latency test setup

The first test performed was done without any test load generated from the tester. Figure 5 shows the latency from PC1 to PC2 for latency sensitive data. Several packets were sent. The results show that the latency through the three FTI switches is 23.8µS.

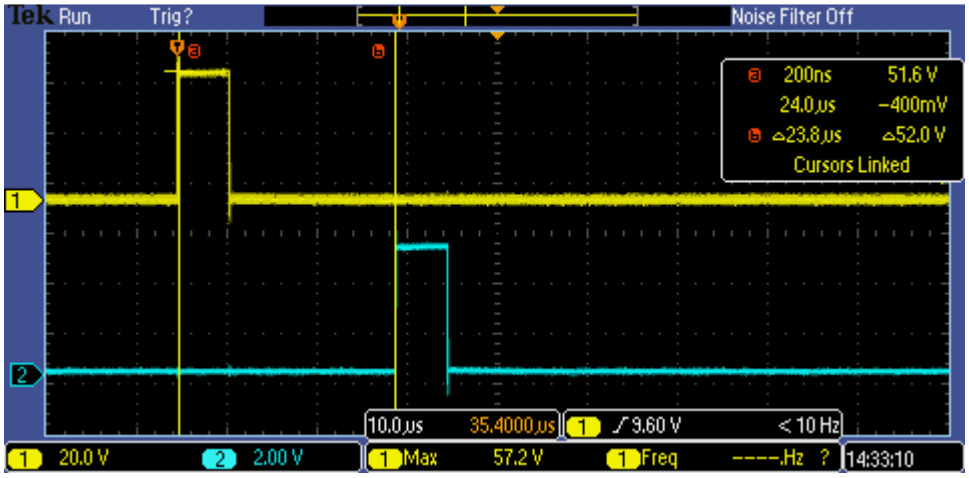


Figure 5, Network latency, three network hops – no load

The second test performed was done with full tester load, but without any QoS settings on the latency sensitive data. No latency sensitive packets were received at PC2 due to the network congestion caused by the tester.

The third test performed was done with full tester load and with QoS settings on the latency sensitive data. Figure 6 shows the measured results. The worst case latency is 52.2µs. Worst case latency means that a given high priority packet is queued for the duration of flushing a 1450 test packet on both the top most FTI switch and the middle FTI switch in the setup. That means approximately 2 x 12us.

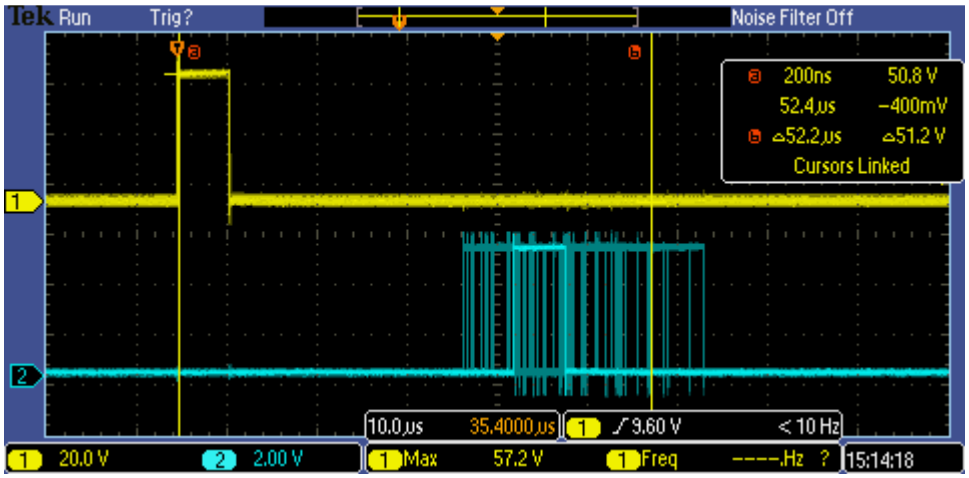


Figure 6, Network latency, three network hops – maximum load with high priority settings on latency sensitive data

**Conclusion**

This paper has demonstrated that worst case switch latency for latency sensitive data through an FTI network consisting of several network hops can be guaranteed. The latency for such data can be far less than 200µs if:

- The total amount of latency sensitive data only represents a small fraction of the total network load
- Latency sensitive data is identified by the FTI switches as high priority data and all other data has less priority
- Gigabit speed is used in the FTI back bone
- Strict priority scheduling is used on the FTI switches