QoS Validation

WHITE PAPER

Quality of Service Validation

Contents

Introduction

What is Quality of Service (QoS)

QoS is the capability of a network to provide better service to selected network traffic over various underlying technologies such as:

- Frame Relay
- ATM
- Ethernet and 802.1networks
- SONET
- IP-routed networks

In other words, QoS is to differentiate network traffic and prioritize different traffic flows accordingly in order to provide a certain degree of predictability to different networks services. A QoS enabled network should provide service guarantees for various application types while making efficient use of network resources. Primary goals of QoS include dedicated bandwidth, controlled jitter and latency (required by some real-time and interactive traffic), and improved loss characteristics.

QoS Parameters

Different applications have different requirements regarding how to handle their traffic in the network. Applications generally require that the network be able to carry the traffic at the rate at which they generate. Some applications are more tolerant on traffic delays and jitters and can tolerate some degree of traffic loss, while others are more stringent. These requirements are expressed in the following QoS-related parameters:

- **Throughput/Bandwidth:** The rate of successful frame/packet delivery over a communication channel
- **Latency:** The time taken by a data message to travel the system from one end to another.
- **Jitter:** The variation of the delay encountered by similar frames/packets following the same path through the system
- **Loss Rate:** The rate at which data is dropped, lost or corrupted while going through the system

Why is QoS Important?

It is very uncommon that a network only serves a specific application. On the contrary, today's networks need to support multiple kinds of application over the same infrastructure and needs to be able to provide preferential treatment to different traffic. There are two approaches to satisfy the requirements of services, over-provisioning, i.e. provision enough network resources in the network to meet the traffic requirements, and QoS mechanisms, i.e. enabling QoS to treat services resource efficiently.

By over-provisioning, the network is allocated with resources that can handle the peak aggregate offered load. However, over-provisioning is not viable in practice at the edge of the network, where demands for bandwidth from applications combined has more than kept pace with the availability of bandwidth. Thus, QoS is often required even in a well-designed and well-provisioned network in order to support services and applications with stringent SLA requirements. The benefits of enabling QoS mechanisms are as follows:

- Reduces costs by using existing resources efficiently, thereby delaying or reducing the need for expansion or upgrades.
- Gives administrators control over network resources and allows them to manage the network from a business, rather than a technical, perspective.
- Ensures that time-sensitive and mission-critical applications have the resources they require, while allowing other applications access to the network.
- Improves user experience.

Implementation Methods

QoS can be implemented at Layer 2 and/or Layer 3. There are various methods to implement QoS. Typical methods are as follows:

- Classification: A function of the edge routers that group traffic into classes based on specific criteria.
- Congestion management: It allows the control of congestion by determine the order in which packets are sent out based on priorities assigned to those packets. Congestion management include the creation of queues, assignment of packets to those queues based on the classification of the packet, and scheduling of the packets in a queue for transmission.
- Congestion avoidance: It monitors network traffic loads in an effort to anticipate and avoid congestion at common network bottlenecks.
- Policing and shaping: It regulates network traffic based on committed access rate (CAR) and peak rate. Traffic policing usually mark down the packets that violate its contract for dropping. Traffic shaping, on the contrary, regulates network traffic by delaying excess traffic, smoothing bursts, and preventing unnecessary drops.
- Signaling: It provides a way for an end station or network node to signal its neighbors to request special handling of certain traffic. QoS signaling coordinates traffic handling techniques provided by other QoS features. It is essential in configuring successful overall end-to-end QoS service across a network.

QoS Mechanisms for Different Domains

Based on different networking technologies and resources, different domains and layers can employ different QoS mechanisms and technologies for traffic handling, such as:

- Layer 2 Class of Service – 802.1p
  On Layer 2, QoS provides Best Effort (BE) or Class of Service (CoS) from 802.1p without requiring reservation. 802.1p is an aggregate traffic handling mechanism suitable for use in many local area networks (LANs). It defines a 3-bit field in the media access (MAC) header of Ethernet packets shown in Figure 1, offering 8 possible priority values, shown in Table 1.

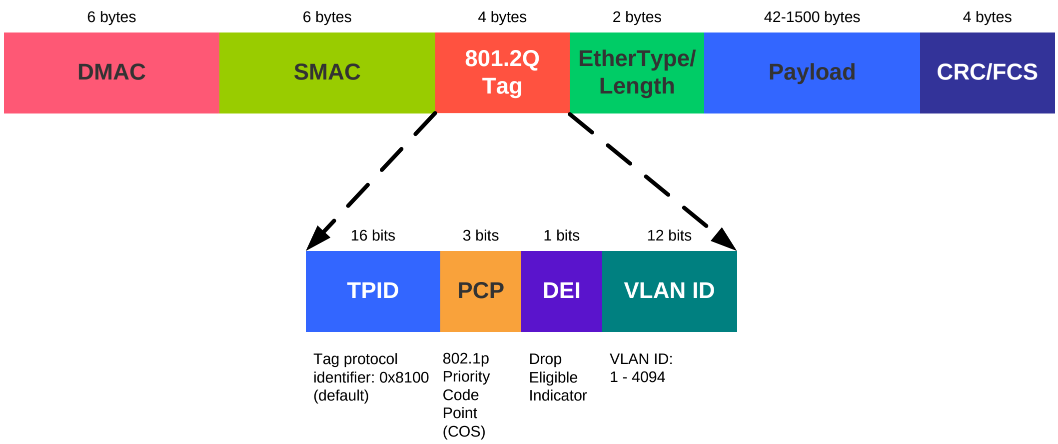


Figure 1. VLAN header

Before sending traffic into a LAN, hosts and routers mark each transmitted packet with the appropriate priority value. LAN devices, such as switches, bridges, and hubs are expected to treat the packets accordingly.

PCP	Priority	Acronym	Traffic Types
1	0 (lowest)	BK	Background
0	1	BE	Best Effort
2	2	EE	Excellent Effort
3	3	CA	Critical Applications
4	4	VI	Video, < 100 ms latency and jitter
5	5	VO	Voice, < 10 ms latency and jitter
6	6	IC	Internetwork Control
7	7 (highest)	NC	Network Control

Layer 2 CoS is an important component when implementing end-to-end QoS across networks, since Ethernet has become a ubiquitous Layer 2 technology. It is important to map the Layer 3 QoS requirements to the correct CoS value in order to guarantee the QoS across two layers.

- Layer 3 QoS – Differentiated Service (DiffServ)

  DiffServ is an aggregate traffic handling mechanism suitable for use in large IP-routed networks. These networks may carry many thousands of conversations, making it impractical to handle traffic on a per-conversation basis. DiffServ operates on the principle of traffic classification, where each data packet is placed into a limited number of traffic classes, rather than differentiating network traffic based on the requirements of an individual flow.

  DiffServ defines a 6-bit DiffServ Code Point (DSCP) field in the IP header as shown in Figure 2. The value is used by routers to classify packets and apply specific queuing behavior accordingly. Traffic from many flows with similar QoS requirements is marked with the same DSCP, and aggregated to a common queue or scheduling behavior.

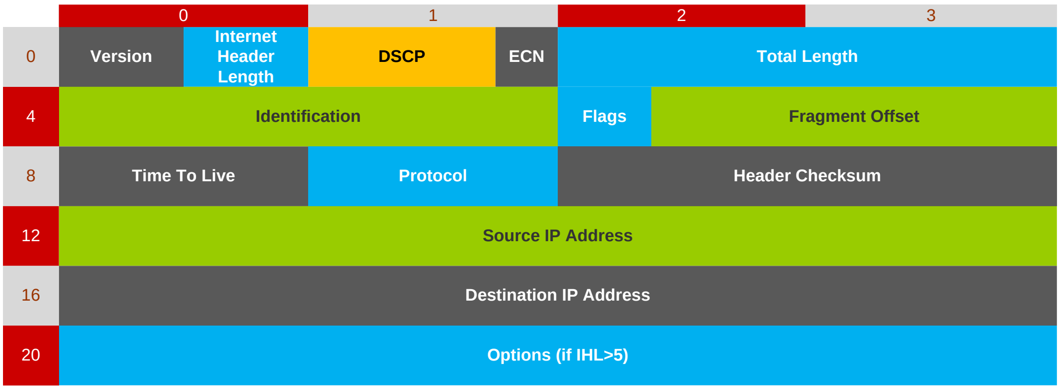


Figure 2. DSCP in IPv4 header

DiffServ-aware routers implement per-hop behavior (PHB), which define how packets associated with a class of traffic are forwarded, including scheduling, queuing, policing and shaping. Different PHBs can be define to offer low-latency, low-loss, for instance. PHB is determined by the DSCP value, which in theory can have up to 64 (i.e. 26) different traffic classes. Based on different DSCP values, operators can define corresponding PHBs. The following PHBs are widely used:

- Default Forwarding: The Default Forwarding (DF) PHB has best-effort forwarding characteristics. The DSCP value for the DF PHB is 000000.

- Expedited Forwarding: Defined in RFC 3246, the Expedited Forwarding (EF) PHB has the characteristics of low latency, low jitter and low loss. These are suitable for voice, video and other real-time services. EF traffic is often given strict priority queuing above all other traffic classes. The DSCP value for expedited forwarding is 101110.

- Voice Admit: Voice Admit (VA) is defined in RFC 5865, and has identical characteristics to the EF PHB. The DSCP value for expedited forwarding is 101100.
- Assured Forwarding: Assured Forwarding (AF) allows the operator to provide assurance of delivery as long as the traffic does not exceed some subscribed rate. Different priority and drop precedence values are assigned, which allows for more granular QoS control of packets.
- Class Selector: Class Selector (CS) PHB is used to ensure compatibility with Type of Service (TOS)/IP Precedence. The Class Selector code points are of the form “xxx000”. The first three bits are the IP precedence bits.

DiffServ is a coarse-grained, class-based mechanism for traffic management. In contrast, Integrated Service (IntServ) is a fine-grained, flow-based mechanism.

- Layer 3 QoS – Integrated Service (IntServ)

  IntServ is a framework for defining services. With a set of underlying traffic handling mechanisms, IntServ is typically applied on a per-conversation basis. Typically, IntServ uses Resource Reservation Protocol (RSVP) protocol to provision QoS along the path of the data flow as shown in Figure 3. In order for IntServ to work, all routers along the traffic path must support RSVP. Furthermore, many states must be stored on each router. As a result, IntServ works on a small-scale, but as the system scales up, it is difficult to keep track of all of the reservations.

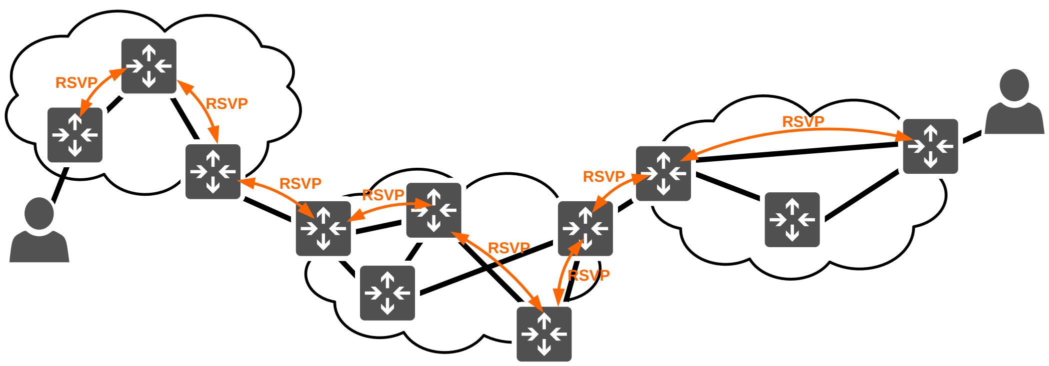


Figure 3. RSVP along the data path in the IntServ architecture

Enterprises, where Layer 2 technologies such as Ethernet is mainly used, can control their networks and systems using IEEE 802.1p to mark Ethernet frames to different priorities. 802.1p marking allows switches to provide prioritization of different classes of traffic, e.g. voice, video, data, etc., offering preferential treatment to certain flows across VLANs.

Network Service Providers (NSP) aggregate and forward traffic within their own networks or to another NSP. Technologies such as DiffServ or IntServ are used by NSPs to classify and manage network traffic on IP networks. NSP SLAs enforcement for transit traffic must be implemented by QoS policies.

QoS Validation and Testing Challenges

Validation and testing are essential for QoS deployment and QoS-related products. However, several challenges exist when validating the QoS of networks:

- **Network Convergence**

Different services such as internet access, voice, video, and data storage are using the same network to transport data with shared bandwidth. These types of traffic have very different requirements. For example, voice applications have stringent requirement on delay, jitter and loss, but video applications can tolerate some degree of delay, jitter and loss but require higher throughput/bandwidth. Thus, different services stress the network uniquely and need to be treated accordingly. Thus, the test solution must be able to generate configurable and flexible test models that are multi-port, multi-stream and multi-service. The high complexity offered by the test solution will help the provider and enterprise to keep their service quality.

- **High Availability**

In a converged network, where services are running on a single network infrastructure, service availability is critical. As mission-critical applications are converged to the network, resiliency and fault-tolerance become increasingly important. Even a short period of network outage can cause a very costly result to the service provider and its customers. Network availability and robustness have therefore become a prerequisite for providers offering reliable carrier-grade services. QoS test solutions should be to validate the provisioned QoS in connection with resiliency and availability, such as one-way latency, recovering time, etc.

- **Complexity and Scalability**

In a network where triple play services (voice, video, and data) and various types of traffic are provided simultaneously, it is no longer adequate to test services separately. Differentiated services should be tested and verified by generating traffic streams in the same format as used in real networks, i.e. VLAN, MPLS, and IP. At the receiving side, traffic flows must be separated and have individual per-flow QoS measurement. Tests should also be able to provide scalable traffic and stress the network or the device under test (DUT) with high traffic load in order to test the performance impact.

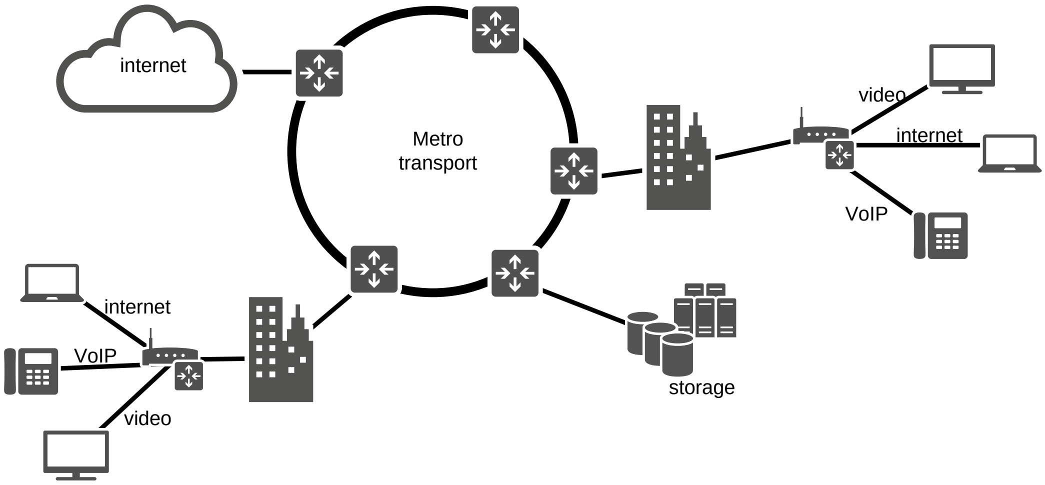


Figure 4: Differentiated services should be tested and verified by generating traffic streams in the same format as used in real networks.

Xena QoS Validation Solution

Xena offers a Layer 2-3 Gigabit Ethernet test platform that delivers a breakthrough price-performance benchmark for QoS validation, load stress and functional testing of Ethernet equipment and network infrastructure. With a robust and versatile Layer 2-3 test platform, Xena is able generate wire-speed Gigabit Ethernet traffic up and analyze the results via user-friendly software. Various types of traffic (Ethernet, Ethernet II, VLAN, ARP, IPv4, IPv6, UDP, TCP, LLC, SNAP, GTP, ICMP, RTP, RTCP, STP, MPLS, PBB, or fully specified by users) can be generated into the device under test (DUT) so that the impact of complex traffic on DUT can be tested.

Service providers, network equipment manufacturers (NEMs), component manufacturers, and operators can verify service attributes of entire networks end-to-end, as well as isolating problems down to individual networking devices and subsystems. With Xena, test traffic can easily be scaled up from a single stream to multiple streams with different traffic patterns, offering great scalability and flexibility. Xena receiving engine is capable of providing highly sensitive latency measurement for network resiliency, stringent QoS, etc.

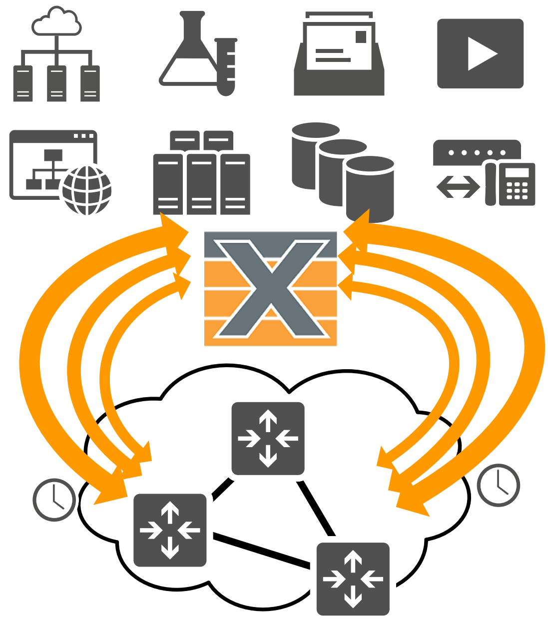


Figure 5. Xena is capable of generating scalable multi-port, multi-stream and multi-service traffic with high latency sensitivity for QoS validation

Xena Layer 2 QoS

As a ubiquitous Layer 2 technology, Ethernet is widely used in data centers, enterprise networks, metro and access networks, mobile backhaul, etc. Services such as voice, video, internet data, and mobile backhaul must comply with strict Service Level Agreements (SLAs). SLA is a binding contract between a service provider and a customer. It guarantees the minimum performance assured for the provided services.

Router and switch manufactures must validate that their equipment can deliver QoS guarantee by features such as classification, queuing and scheduling, and buffer management under different traffic loads. Various queuing and scheduling mechanisms have been used by the vendors to provide QoS, such as weighted round robin (WRR), deficit weighted round robin (DWRR), priority-based deficit weighted round robin (PB-DWRR), and weighted random early detection (WRED). These mechanisms must be correctly implemented in order to prevent high-priority traffic from being dropped. Network administrators need to validate their network design and configuration in order to guarantee the SLAs.

Both ValkyrieBay and ValkyrieCompact testers provide Layer 2 QoS testing methods RFC 2544 (Valkyrie2544) and ITU-T Y.1564 (Valkyrie1564), and offer full supports to 802.1q (VLAN) and 802.1ad (Q-in-Q) traffic, with high-performance per stream tracking and analysis. An example of configuring VLAN traffic with ValkyrieManager (Layer 2-3 software) is shown below.

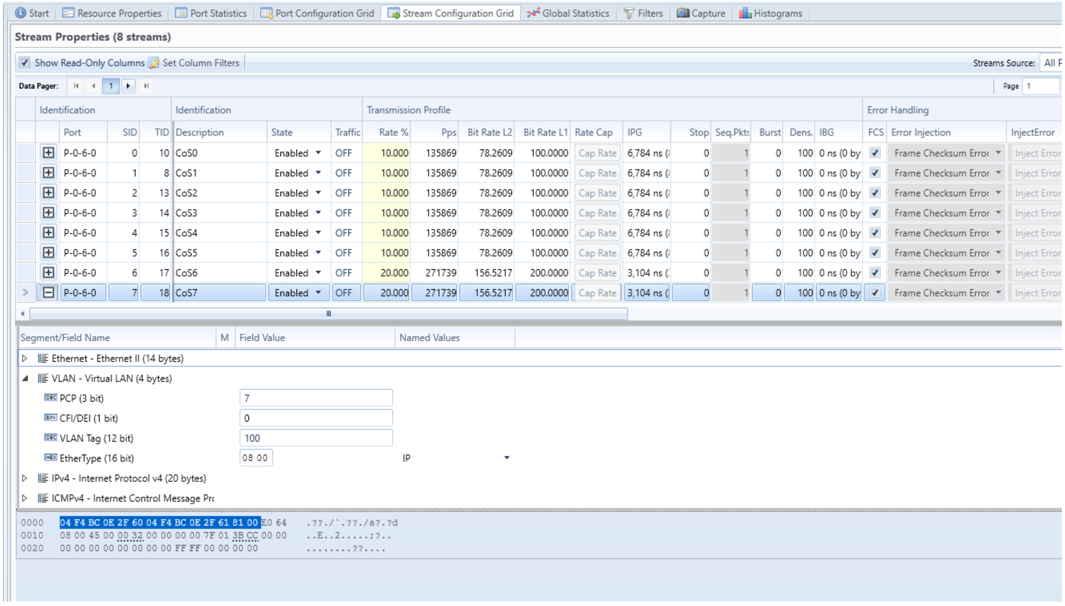


Figure 6. Example of configuring VLAN traffic using ValkyrieManager

VALKYRIE2544

Valkyrie2544 is an application provided by Xena, making it easy to create, edit and execute test configurations in accordance with RFC 2544. Valkyrie2544 offers full support for the four test types specified in RFC 2544, and lets customers partially enable one or more test types. Valkyrie2544 supports different network topologies and traffic flow directions on both Layer 2 and Layer 3, as well as both IPv4 and IPv6. Key features of Valkyrie2544 are as follows:

- Perform all RFC 2544 tests over VLANs and IPv4 & IPv6 protocols
- Test with jumbo frames and verify low-latency and wire-rate
- Large port count and full mesh tests, uni-directional or bi-directional testing
- Easy point-to-point, point-to-multipoint, and multipoint-to-point testing
- Supports both single stream and multi-stream testing
- Reduce time-to-test through easy configuration and fast execution
- Summary and comprehensive detailed results
- Support for both standard and fully user defined test configuration

Typically, the Layer 2 QoS is implemented through the CoS field in the VLAN tag of Ethernet frames. Packets, or frames to be precise, are switched instead of routed in the access and metro Ethernet networks where processing time is reduced by inspecting Layer 2 headers without viewing the Layer 3 content. The following key measurement metrics should be collected for Layer 2 QoS as they are the key performance indicators (KPIs):

- Frame loss
- Latency
- Jitter
- Throughput/Bandwidth

VALKYRIE1564

Valkyrie1564 is an application provided by Xena that lets customers perform the Y.1564 methodology using one or more Xena test chassis. Y.1564 is the ITU-T standard for turning up, installing and trouble-shooting Ethernet-based services. Formulated as an improvement over RFC 2544, Y.1564 is the only standard test methodology that allows for complete validation of Ethernet service-level agreement's (SLAs) in a single test. Valkyrie1564 provides full support for both the configuration and performance test types described in Y.1564. The simple intuitive graphic user interface (GUI) makes it easy to connect one or more ValkyrieBay and/or ValkyrieCompact chassis for testing Layer 2 and Layer 3.

For services provided by Ethernet, bandwidth profile defines the following four traffic profiles:

- Committed Information Rate (CIR)
  The CIR is the maximum sustained information rate that the network is committed to carry while meeting the performance level guaranteed in the SLA.

- Committed Burst Size (CBS)
  The CBS is the number of allocated bytes available for bursts of ingress service frames transmitted at temporary rates above the CIR while meeting the SLA guarantees provided at the CIR.

- Excess Information Rate (EIR)
  The EIR is the maximum sustained information rate by which a user can exceed its CIR with some expectation that the excess traffic might be carried through the network.

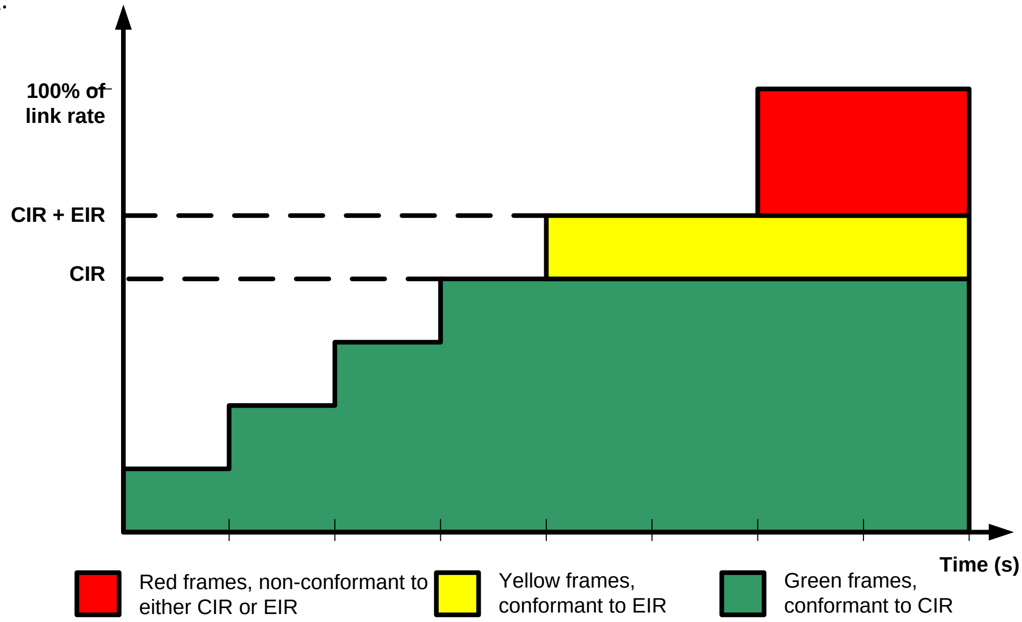
- Excess Burst Size (EBS)
  The EBS is the number of allocated bytes available for bursts of ingress service frames sent at temporary rate above the CIR+EIR while remaining EIR-conformant.

CIR and CBS are related in such a way that CBS must be defined when CIR is set at a value that is greater than zero. EIR and EBS are related in the same way as CIR and CBS. Customer traffic is processed based on their conformance to CIR/CBS/EIR/EBS, and classified into three class. Figure 6 illustrate the relation between CIR, EIR and the different color frames.

- Green frames: Green frames are conformant to CIR.

- Yellow frames: Yellow frames are conformant to EIR. They have higher discard precedence than the green frames and are expected to be dropped first when congestion is encountered at the service layer.

- Red frames: Red frames are non-conformant to either CIR or EIR, and are dropped at the interface by traffic policing.



Xena Layer 3 QoS

Traffic at Layer 3 is classified into different types. Different types of traffic must be handled differently in order to achieve optimal network performance. Layer 3 QoS mechanisms should apply correct queue size configuration to provide guaranteed delay and jitter for both latency- and bandwidth-sensitive application. IntServ and DiffServ are the most common methods of Layer 3 QoS implementation.

IntServ uses RSVP protocol to reserve network resources along the path of the stream to provide a granular QoS solution. It requires that the network elements keep track of each stream and maintain a soft state. DiffServ offers a coarser and simpler QoS method by categorizing traffic into different classes and providing per hop behavior (PHB). Key measurement metrics for Layer QoS are:

- Packet loss: The loss of many consecutive packets in a VoIP scenario can dramatically degrade the speech quality. In a data transfer scenario, excessive packet loss will cause unnecessary network resource allocation for data retransmission.

- Latency: One-way latency (OWL) and latency variation (jitter) measurements are important for testing the QoS of real-time applications such as VoIP, telepresence and transactional services. High precision is required for both wireline (Ethernet access) and wireless backhaul networks for 4G (WiMAX/LTE), and typical SLA specifications call for unidirectional jitter below 1-5 ms and latency in the 3-10 ms range.

- Jitter: Many equipment has jitter buffers that can absorb a certain degree of jitter effect. However, if the jitter buffer is too small, a large number of packets can be dropped by the receiver, which lead to service degradation.

Xena solutions let customers test DUT Layer 3 QoS performance under different load conditions with different traffic patterns. Different DSCPs, such as DF, AF 11 to AF 43, CS 1 to CS7, and EF, can be allocated to streams of different IP address. This flexibility lets network administrators create required traffic profile to test their QoS policies and rules and validate their Layer 3 QoS mechanisms under various loads.

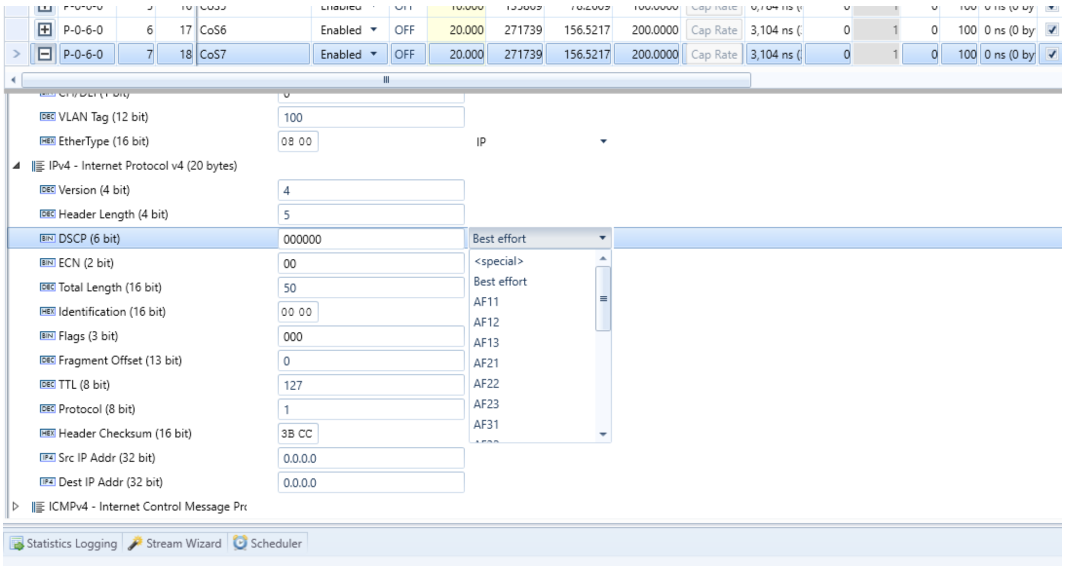


Figure 8. Example of configuring Layer 3 QoS using ValkyrieManager

ValkyrieTimeSynch is a hardware option that enables OWL measurements, synchronized traffic start and accurate time-stamping when using ValkyrieManager. ValkyrieTimeSynch enables multiple Xena testers to synchronize their local time to each other. This can be used for OWL measurements between two test chassis, synchronized traffic start between multiple chassis and accurate timestamping of captured packets in exported PCAP files. ValkyrieTimeSynch ensures the test modules can measure with microsecond precision. Both the ubiquitous Network Time Protocol (NTP) and the newer, more precise IEEE 1588 Precision Time Protocol (PTP) are supported and can be mixed as appropriate for the network.

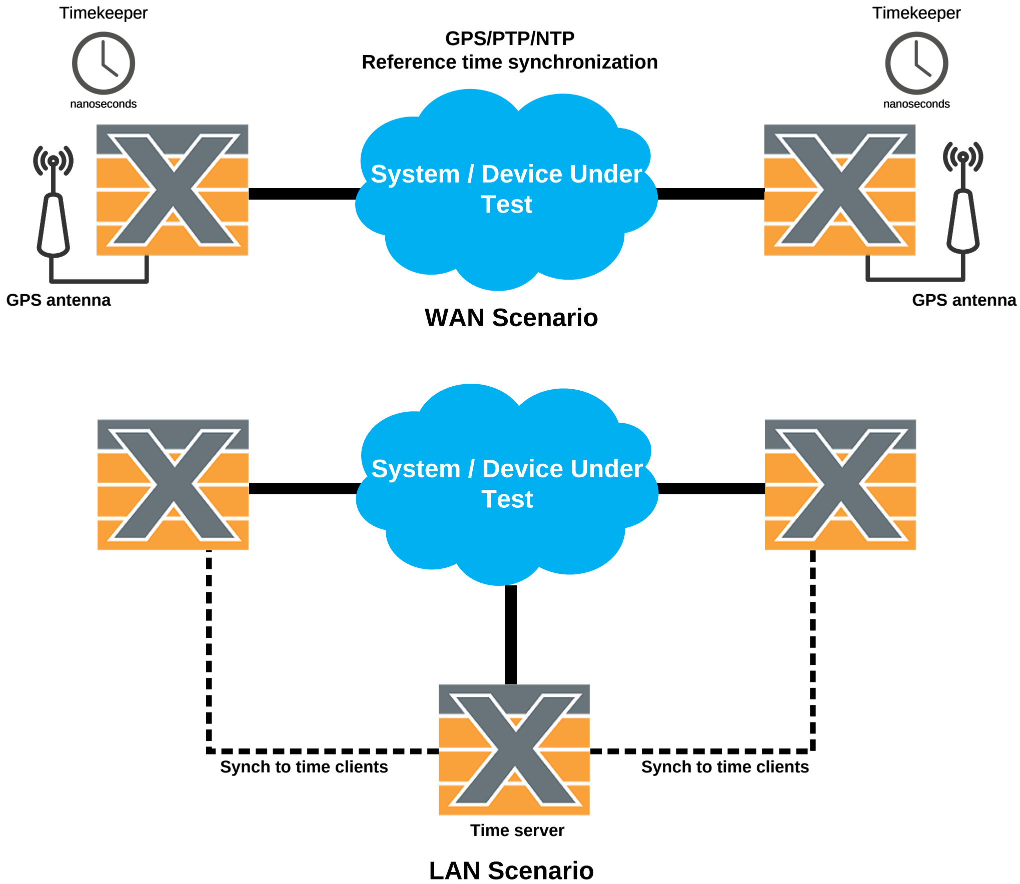


Figure 9. Xena is capable of measuring one-way latency of high precision for QoS validation in both WAN and LAN scenario setups.

Xena Advanced Statistics

For both Layer 2 and Layer 3 QoS validation, Xena provides powerful traffic statistics functions in the ValkyrieManager such as statistic charting, logging, and histogram.

Xena statistics charting function as shown in Figure 10 allows users to check the QoS latency performance over extended period, such as over a 24h interval. Xena statistics logging function as shown in Figure 11 allows users to select desired statistics and export results to standard formats. The histogram function can show the precise latency distribution for all or specific traffic. For example shown in Figure 12, with two histograms, users can then compare latency of a traffic with different packet sizes, or different priorities.

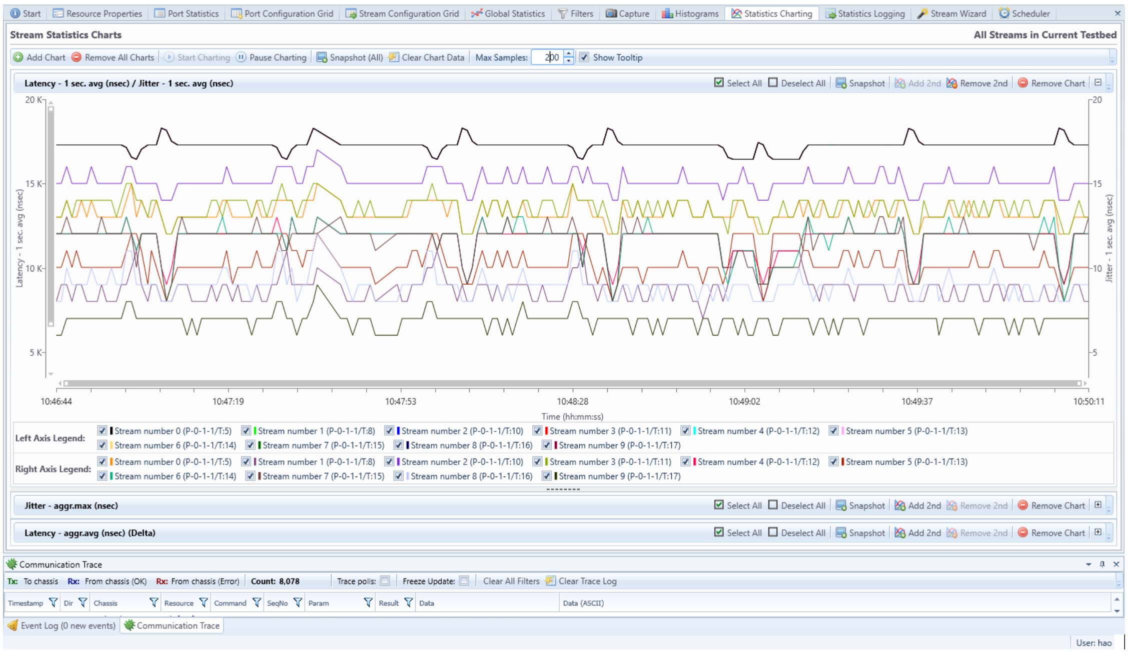


Figure 10. Xena statistics chart where 10 streams are displayed in real time with latency and jitter.

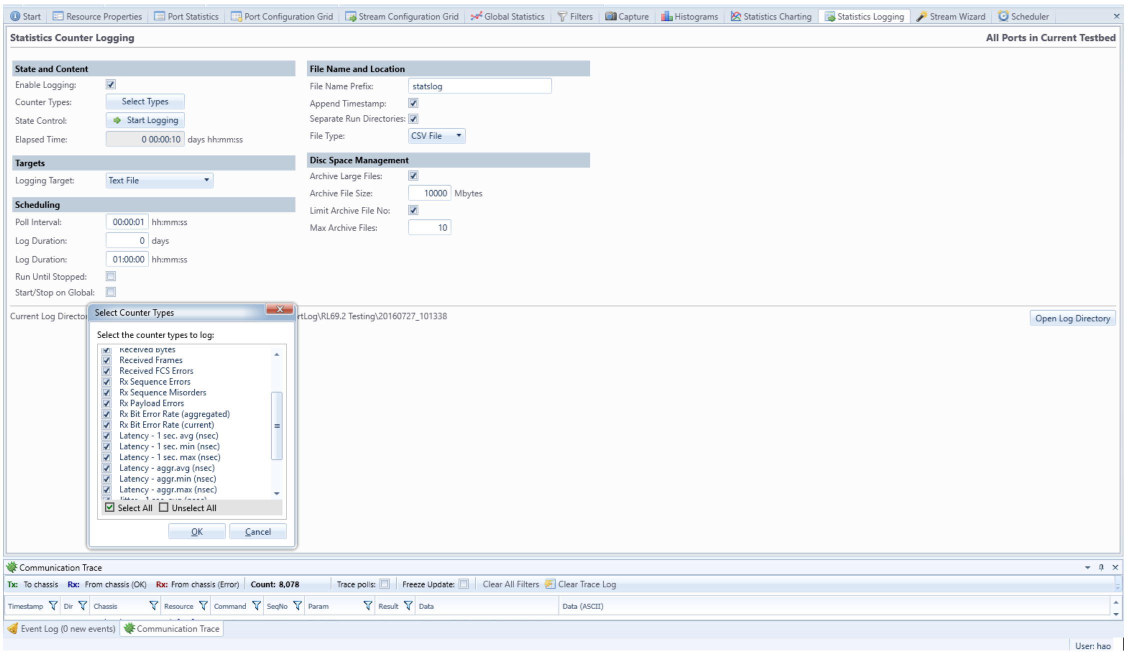
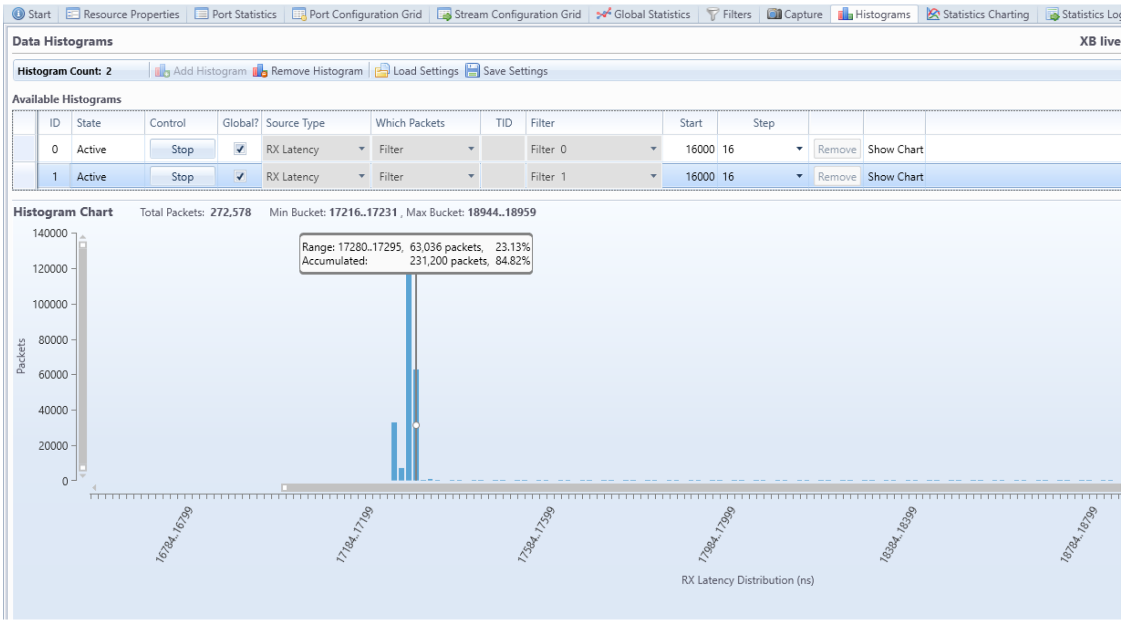
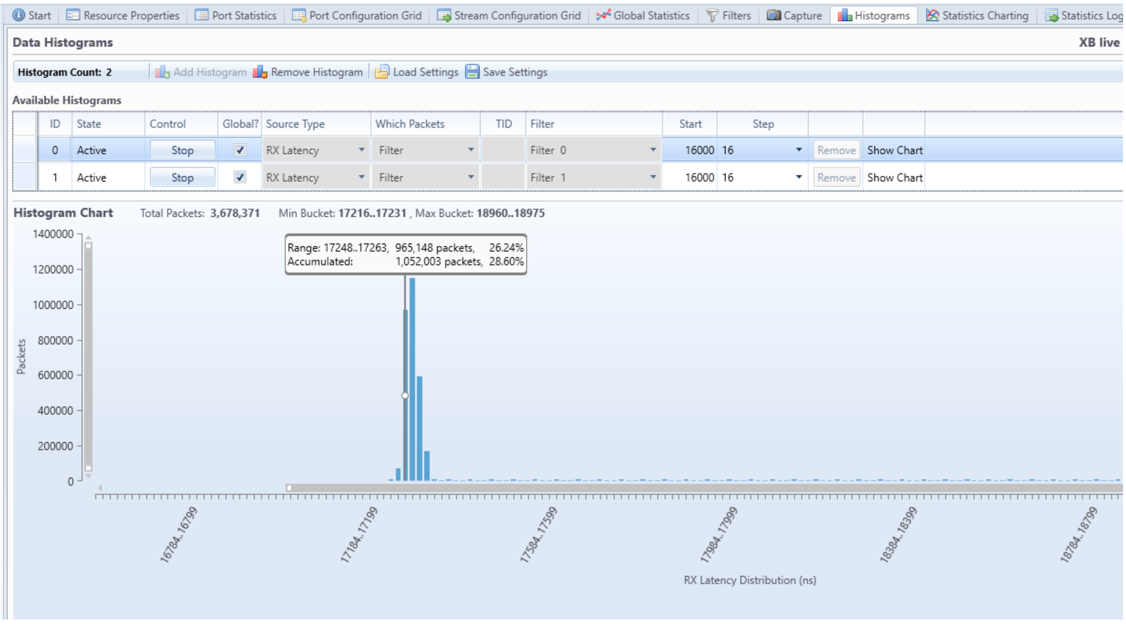


Figure 11. Xena statistics logging allows users to select desired statistics and export results.



(a)



(b)

Figure 12. Xena statistics histogram shows the distribution of statistics over traffic. (a) shows latency distribution over packet sizes of at least 100 bytes (b)) shows latency distribution over packet sizes of at most 100 bytes

CONCLUSION

QoS differentiates network traffic and prioritizes different traffic flows accordingly in order to provide a certain degree of guarantee to different networks services, such as voice, video and data. A QoS enabled network should provide service guarantees for various application types while making efficient use of network resources. Validation and testing are essential for QoS deployment and QoS-related products.

With high-performance Layer 2-3 network test and measurement equipment ValkyrieBay and ValkyrieCompact, Xena is capable of delivering powerful QoS validation solutions in accordance with RFC 2544 and Y.1564 to enterprises, data centers, network administrators, operators, device vendors, etc. to test and evaluate the QoS mechanisms, which are either deployed or developed. Xena offers a wide selection of port speeds, ranging from 1G to 100G. With the advance statistics functions, users can track, analyze and troubleshoot QoS, in order to maintain a high service quality guarantee.

>> Request for FREE Valkyrie demo

>> Book a FREE consultation with a Xena expert