VMware SD-WAN QoS Overview

DMPO and integration with MPLS QoS

velocloud™
Now part of VMware

Introduction

Quality of Service over the network is critical for application performance. As organizations access more applications over the Internet, there is a need for QoS on the Internet. SD-WAN has made this possible.

With the rapid adoption of SD-WAN by enterprises, hybrid deployments with MPLS VPN as a WAN transport complemented with Internet links, have become a standard SD-WAN branch design. For MPLS networks that offer QoS in order to guarantee end-to-end service for voice, video, and data, it is critical to understand how QoS works within VMware SD-WAN® by VeloCloud™ and how to integrate VMware SD-WAN with an existing MPLS VPN network.

This document explains the QoS architecture within VMware SD-WAN and its respective network and link schedulers, the challenges that come with integrating MPLS QoS with VMware SD-WAN, and recommendations and best practices that would allow for an efficient migration.

QoS for MPLS VPN without SD-WAN

Before the adoption of SD-WAN technologies, enterprise private WAN networks typically used MPLS VPNs where MPLS service providers worked with their customers to implement QoS policies to ensure end-to-end service levels for voice, video and data. In this model, QoS policies that enforce queueing, marking, policing, and shaping are provisioned on the provider edge (PE) and customer edge (CE) routers, while the service provider MPLS core supports differentiated services (DiffServ) through the application of RFC 3270—which defines various ways to map DiffServ information into MPLS labels.

Figure 1 provides a high-level overview of how QoS is implemented in an MPLS VPN to guarantee end-to-end QoS for enterprise traffic flows from left to right. On the CE router, an outbound QoS policy is implemented on the interface from CE to PE for queueing, shaping, and remarking. The CE to PE link is where a common queueing mechanism such as low latency queueing (LLQ) is configured and applied. On the PE router ingress interface, service providers will likely implement ingress policies to identify whether traffic flows are in compliance with the contract and optionally remark the original differentiated services code point (DSCP) values. When the traffic arrives on the other end of the MPLS VPN network, the remote PE router will apply PE to CE policies such as LLQ and class-based weighted fair queueing (CBWFQ) based on the original packet markings.

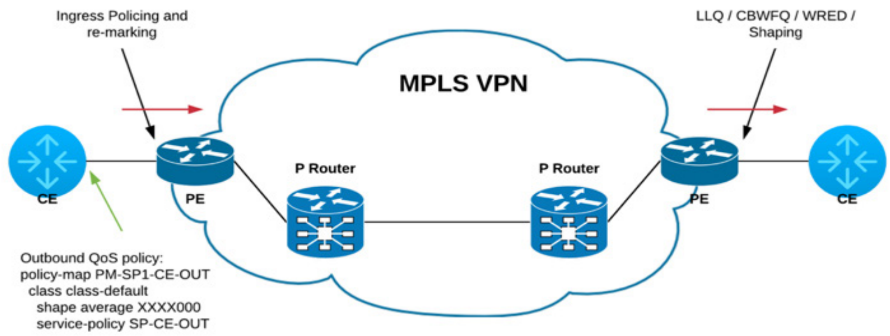

FIGURE 1: QoS for MPLS VPN

With MPLS VPN, QoS is needed to prioritize certain critical or network-sensitive traffic during congestion to deliver the desired SLA.

VMware SD-WAN DMPO and QoS

VMware Dynamic Multipath Optimization (DMPO) is a collection of techniques utilized by the VMware SD-WAN overlay to provide real-time link monitoring, per-packet load balancing and steering, and on-demand link remediation. It is used between all of the VMware SD-WAN components that process and forward data traffic. The overlay tunnels encapsulate the original payload in user datagram protocol (UDP) and add proprietary headers to keep track of the original flow and impose their own sequence numbers in the outer header. The default DSCP value applied to the overlay header is CS0/DF. The VMware SD-WAN data plane utilizes the overlay headers to apply the configured QoS policies, as well as the appropriate on-demand remediation techniques such as forward error correction (FEC) and jitter de-buffering for real-time traffic. VMware SD-WAN's DMPO prevents congestion natively and therefore does not depend on MPLS QoS for traffic classification and prioritization.

Figure 2 shows an example where there are two WAN transport types available at each branch Edge: Internet and MPLS. The Edges are configured with overlay tunnels across both WAN circuits and through DMPO, the Edge is able to monitor both overlay paths and use them simultaneously with per-packet load balancing and steering. By default, the original user traffic is encapsulated in overlay headers by the Edge with an outer DSCP value of CS0 and the original packet's DSCP unchanged. The Edges will classify the user traffic, apply QoS to the traffic based on the application types and their traffic class, and direct the packets to be processed through the network scheduler and link scheduler. Based on the overlay tunnels' queue depth, available bandwidth and end-to-end latency, the best available path is chosen on a per-packet basis. Therefore, when there are multiple WAN overlay paths between VMware SD-WAN components, traffic may be prioritized within the overlays based on their traffic class (i.e. real-time vs bulk) and sent across both overlays as shown in the diagram.

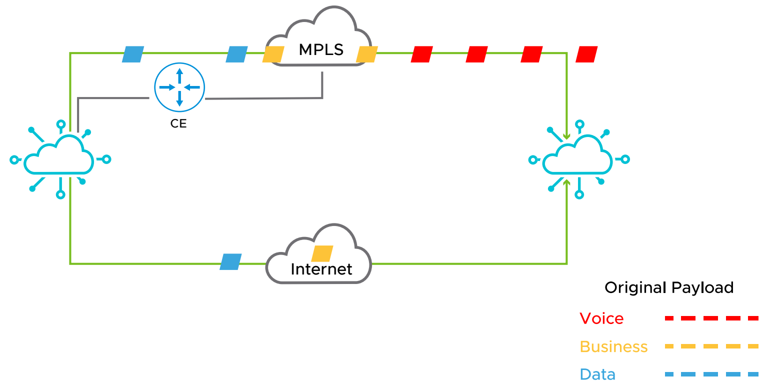

FIGURE 2: VMware SD-WAN DMPO Example

In this example, If the CE router depicted in the diagram does not apply QoS and only sees the outer overlay headers, the Edges will handle end-to-end packet queueing, prioritization and transmission for traffic between the devices. The Edges will provide steering and remediation as needed—where the sending Edge may reorder packets based on the configured QoS policy, and the receiving Edge may buffer traffic and reconstruct flows utilizing DMPO and the outer headers. This is the default behavior when auto-steering is configured in the business policies. Note that by default the original packet's DSCP value is not copied to the overlay header unless the business policies explicitly instruct the Edge to do so. This implies that if an overlay is established across MPLS, the Edges will rely on the provider's treatment of CS0 traffic to determine path quality of the MPLS WAN overlay.

Data plane schedulers

The VMware SD-WAN data plane implements two sets of schedulers, shown in Figure 3, which influence the QoS of outbound traffic from the VMware SD-WAN Edge. The first and most important scheduler, from a QoS perspective, is the Network Scheduler on the left which implements the QoS hierarchy. The second scheduler is a WAN Link Scheduler which prevents oversubscription of local and remote network connections.

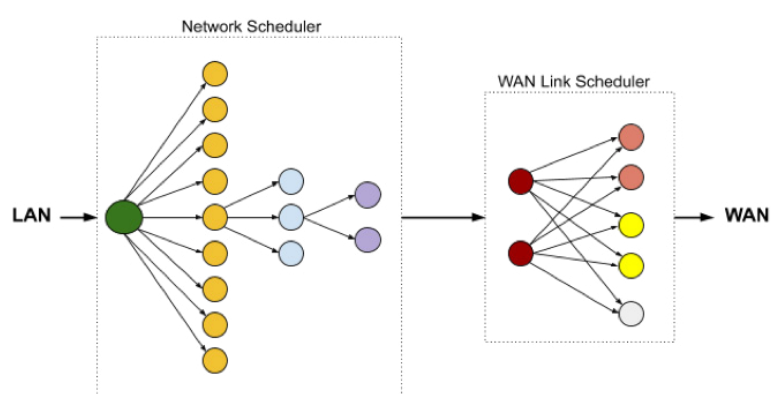

FIGURE 3: VMware SD-WAN QoS schedulers

The entire QoS processing flow with respect to the hierarchical queueing construct is outlined in Figure 4, which adds color to the high-level diagram shown in Figure 3.

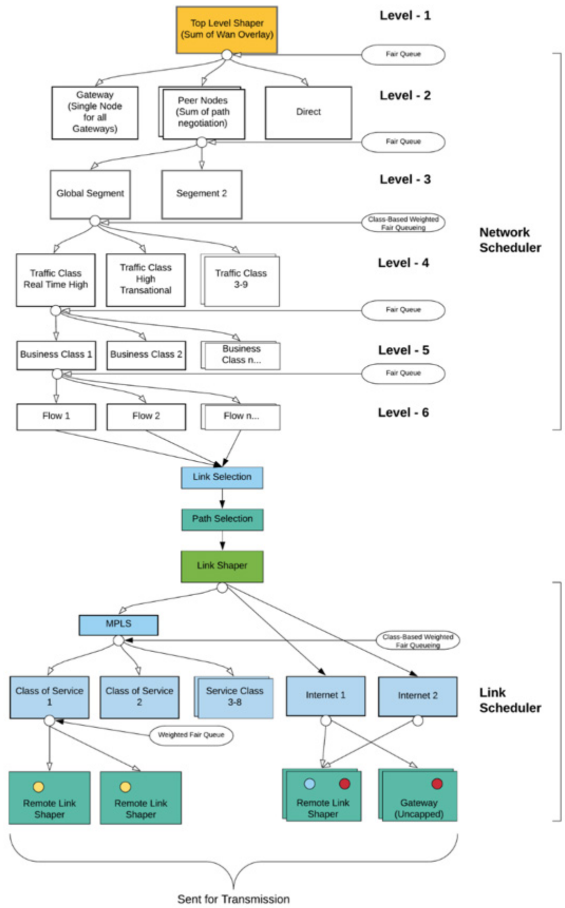

FIGURE 4: VMware SD-WAN QoS schedulers and hierarchical queues

Challenges with integrating DMPO and MPLS QoS

In this section we will discuss the various challenges when integrating MPLS VPN with VMware SD-WAN's native QoS operations and DMPO. The first and most common issue we see is when a customer has outbound QoS policies on the CE and configures business policies on the Edge to copy the original DSCP value to the outer header so the overlay traffic can be subjected to MPLS QoS. This causes reordering of overlay packets due to queueing on the CE and may also inject unintended loss and jitter in the overlay when some of the overlay packets are de-prioritized or dropped. We will go over in detail why this occurs.

Reordering of overlay packets by MPLS QoS

In order to understand the potential packet reordering issue caused by MPLS QoS, we first must understand VMware SD-WAN overlay technology and the operations of DMPO.

For the data plane to achieve per-packet load balancing, the VMware SD-WAN components—the Edge and Gateway—attach a sequence number in the overlay header to each packet that is sent on an overlay path to ensure in-order delivery on the remote end. This path sequence number partly contributes to the DMPO mechanism which monitors end-to-end loss, latency, and jitter of an overlay path. There is a second set of sequence numbers in the overlay header applied to the traffic flow that allows the Edge or Gateway to load-balance or duplicate the flow across multiple overlay paths on a per-packet basis. Given the overlay sequence numbers, packets for a single flow can traverse two links as shown in Figure 5. The numbering of the packets represents the flow sequence number which allows the receiving Edge to reconstruct the flow. Note that for other flow-based SD-WAN solutions, the path sequence number does not apply since packets for a flow are always sent out the same link and they do not provide on-demand remediation when overlay performance deviates.

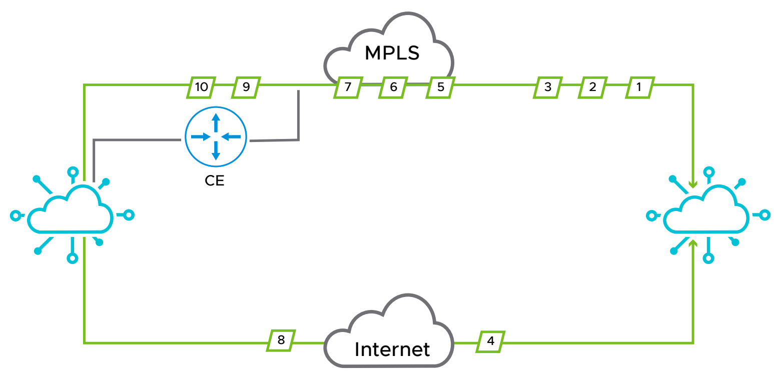

FIGURE 5: Single flow traversing multiple paths across overlay

Let's assume MPLS is the superior link to the Internet and the sending Edge forwards all packets for a flow to the MPLS overlay tunnel. If the receiving Edge receives packets "1, 2, 3, 5" before packet 4, at this point it doesn't know if packet 4 was lost or the sending Edge decided to send it over the Internet and packet 4 just hasn't arrived yet, or perhaps it was sent over the Internet overlay path and lost. This is one of many cases where the path sequence number comes into play. It creates another identifier associated to paths above flow level, which allows the Edges to monitor overlays end-to-end and notify the transmitting Edge of any packets that were lost on individual paths. This function of DMPO enables the sending Edge to employ various remediation techniques, such as overlay loss measurement or dynamically enabling/disabling error correction on a specific overlay path due to loss and retransmitting lost packets.

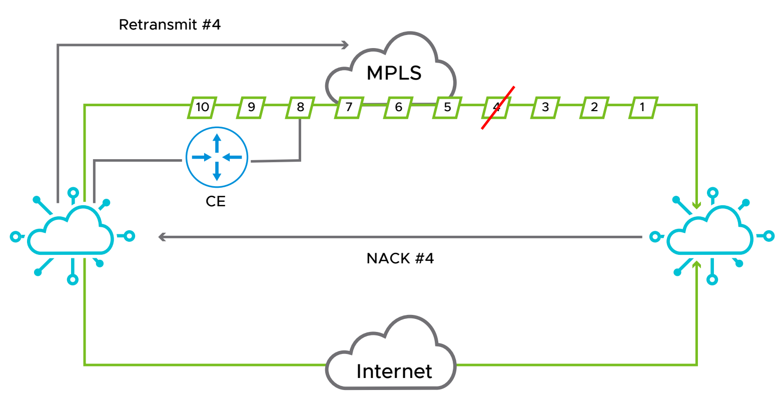

FIGURE 6: Overlay packet retransmission

The issue arises when class of service (COS) is enabled in the private WAN overlay configuration and/or business policies are configured to mark the outer overlay IP header with differentiating DSCP values. The queueing and scheduling mechanisms introduced by the CE router or the DiffServ policies facilitated by the provider's MPLS network may reorder the overlay packets the transmitting Edge sends.

If there are three flows and the 10 packets in the example consist of three classes: EF (6 packets in red), AF11 (1 packet in yellow), and AF31 (3 packets in blue), MPLS QoS may reorder the packets which the transmitting Edge sent based on VMware SD-WAN QoS configuration. The undesired result is that the receiving Edge may get the packets out of order in terms of overlay path sequence numbers, or even lose some packets due to shaping and/or policing by the MPLS VPN network. This problem is depicted in the diagram below.

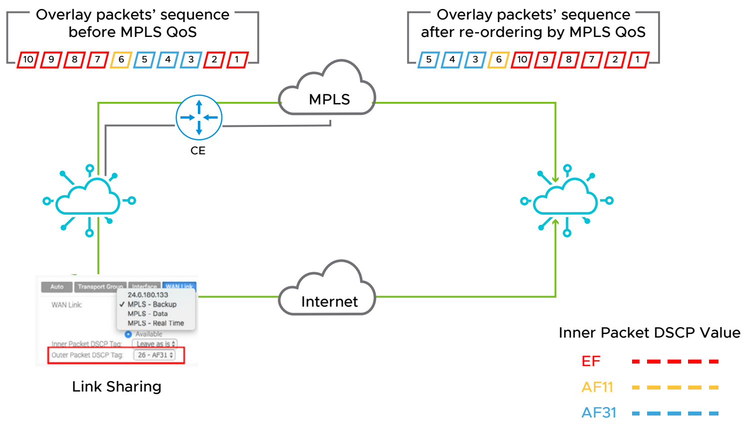

FIGURE 7: Overlay packets being reordered due to MPLS QoS

In the session load balancing world where DMPO techniques are not implemented, link aggregation and real-time remediation do not apply and therefore, reordering of the packets does not pose an issue. The EF packets may simply be prioritized and scheduled ahead of other packets with lower DSCP values and other classes of traffic such as AF11 would be treated according to the customer and provider's QoS policies. This is the expected behavior with QoS in MPLS VPN, as it is trying to react to congestion and prioritize the most important or network sensitive data under congestion—which the DMPO techniques inherently prevent and alleviate.

However, the reordering of the packets due to DSCP markings of the outer overlay header creates several potential problems as the Edge's DMPO will misinterpret the reordering as loss and jitter across the MPLS path. The unintended behavior may result in the following occurrences on the Edge:

- The transmitting Edge avoids the MPLS link.
- DMPO applies unnecessary error correction and consumes more bandwidth.
- The receiving Edge discards packets that arrive "late".

VMware SD-WAN DMPO assures congestion avoidance and traffic prioritization with real-time monitoring, per-packet steering and QoS operations. Here is how the Edge provides further mitigation:

- In the network scheduler, the Edge ensures that traffic is prioritized properly based on the total available bandwidth for transmitting packets to the peer.
- In the WAN link scheduler, the Edge ensures that it does not exceed the capacity of any overlay path by scheduling against both the upstream rate of the local WAN link and the downstream rate of the remote WAN link.
- In the event of congestion, the Edge backs off sending traffic on an overlay path gracefully until the congestion state has recovered.

Impact of link steering with business policy on QoS

The second challenge in integrating these technologies encompasses QoS limitations with link steering when customers configure multiple business policies that prefer or enforce traffic over a specific WAN link. Static link steering is usually configured with the intention of forwarding the most critical traffic over the best WAN link (i.e. put voice and streaming video on an MPLS link). However, forcing traffic over a particular path using business policies may oversubscribe the WAN link because the link scheduler skips auto-steering and the DMPO engine when the traffic is mapped to a specific link.

Let's consider a scenario where a hybrid SD-WAN branch has two WAN links and business policies are configured on the transmitting Edge to steer RTP and Zoom application traffic toward MPLS and both the Internet and MPLS links have throughput of 10 Mbps up and down. Given these conditions, when 20 Mbps of RTP and Zoom traffic is forwarded through the Edge matching the configured business policies, the network scheduler will process the packets through the six levels of QoS operations and shape the traffic to 20 Mbps. QoS on the VMware SD-WAN Edge is as expected at this point, with each traffic class fairly scheduled at 10Mbps. Link and path selection follows the network scheduler as the next steps.

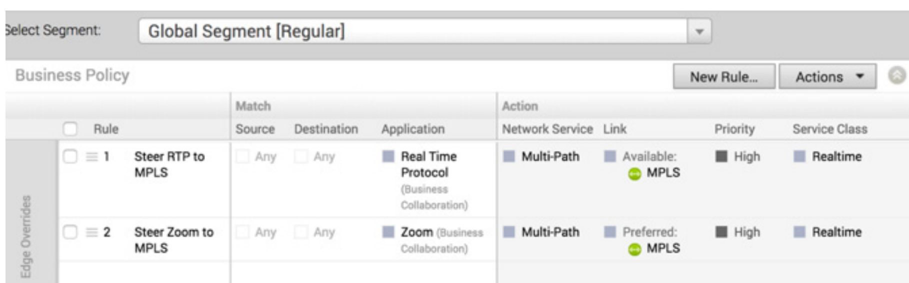

FIGURE 8: Static link steering of applications

This is where the limitation comes in and due to the manual configuration of link steering, link and path selection determined by the DMPO engine is skipped and the packets are directly placed on the MPLS WAN link's outbound shaper. Given that the MPLS link only has a bandwidth of 10 Mbps, placing 20 Mbps of traffic in the MPLS link's shaper queue and not using the available Internet path leads to high queue build up and congestion—which can cause high delay and jitter to voice traffic and negatively impact the user experience. For example, if link steering had been set to "Auto", auto steering plus the DMPO engine would have load balanced the 20 Mbps of traffic on both links to avoid congestion. The link steering limitation and the business policies' impact on the link scheduler processing is highlighted in Figure 9.

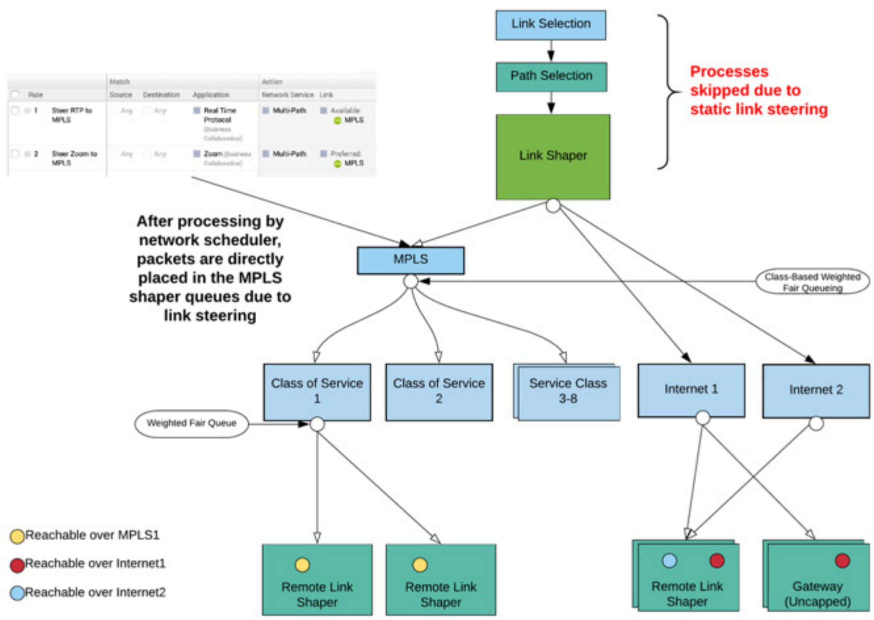

FIGURE 9: Static link steering impact on link scheduler

Given these challenges, it's recommended to configure the auto link steering policy whenever possible.

Best practices and recommendations

Use single class of service for overlay traffic through MPLS VPN and do not mark overlay outer header

MPLS may potentially reorder the overlay packets when the overlay headers are re-marked with DSCP markings and QoS policies are enforced on the CE router or the provider's MPLS VPN network. Reordering of overlay packets by MPLS QoS may cause the packets to become out of sequence when they arrive at the receiving Edge. This can occur when low priority traffic sent through the MPLS path is dropped. Since VMware SD-WAN DMPO performs per-packet steering and automatically reacts to congestion and prioritizes critical applications based on QoS configuration, our recommendation is to use a single MPLS class of service for all overlay traffic and do not mark the overlay's outer DSCP value. The suggestion is to use CS0 or another class that is not policed so overlay traffic can burst up to 100% of available bandwidth.

Use auto link steering

In the previous section, we discussed the limitation of link steering on QoS where manual configuration may lead to oversubscription of a WAN link. This is due to link steering causing link and path selection and the DMPO engine to be skipped when packets are processed by the link scheduler. Instead packets matching the business policies with static link steering—which may sum up to the aggregate bandwidth of all the available WAN links—are placed directly on the selected WAN link's outbound shaper and may overwhelm the link. Manual link steering could also cause the Edge to continue using a slightly degraded path when a better path exists.

With VMware SD-WAN, DMPO monitors end-to-end loss/latency/jitter and inherently steers application packets to the best path, applying QoS in the process. We recommend using Auto mode for link steering whenever possible.

Set static bandwidth measurement on private WAN overlays

Given that most private WAN circuits, such as MPLS, are fixed bandwidth, it's highly recommended to statically configure the bandwidth when bringing up a new private WAN overlay. This eliminates bandwidth measurement across private WAN and allows for expedited tunnel establishment. Manual bandwidth configuration also greatly reduces the load on the hub Edges when there are a large number of branch Edges.

SD-WAN enables end-to-end QoS

QoS over the network is essential for application performance. As organizations access applications in the cloud there is a need for QoS on the Internet. With the migration of applications from the private network to the cloud there is a need to combine QoS on private networks with QoS over the Internet. Before the advent of SD-WAN this was not possible. The integration of VMware SD-WAN DMPO with MPLS QoS has made it possible to configure end-to-end quality of service. By deploying VMware SD-WAN organizations can ensure the user experience across both private networks and the Internet.

For more information see, *www.velocloud.com*.

vmware®

VMware, Inc. 3401 Hillview Avenue Palo Alto CA 94304 USA Tel 877-486-9273 Fax 650-427-5001 vmware.com