TECHNICAL WHITE PAPER:
March 2024

VMware VeloCloud SD-WAN QoS Overview

Dynamic Multipath Optimization and Integration with MPLS QoS

Table of contents

Introduction

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 which offer QoS to guarantee end-to-end service for voice, video, and data, it is critical to understand how QoS works within VMware VeloCloud SD-WAN™ and how to integrate VeloCloud SD-WAN with an existing MPLS VPN network.

This document aims to explain QoS architecture within VMware VeloCloud SD-WAN and its respective network and link schedulers, the challenges that come with integrating MPLS QoS with VeloCloud SD-WAN, and recommendations and best practices for efficient migration.

QoS for MPLS VPN without SD-WAN

Before the adoption of SD-WAN technologies, enterprise private WAN networks typically involved MPLS VPN where MPLS service providers worked with their customers to implement QoS policies to ensure end-to-end service levels for voice, video and data. 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 label.

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 policers to identify whether traffic flows comply with the contract and optionally remark the original 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.

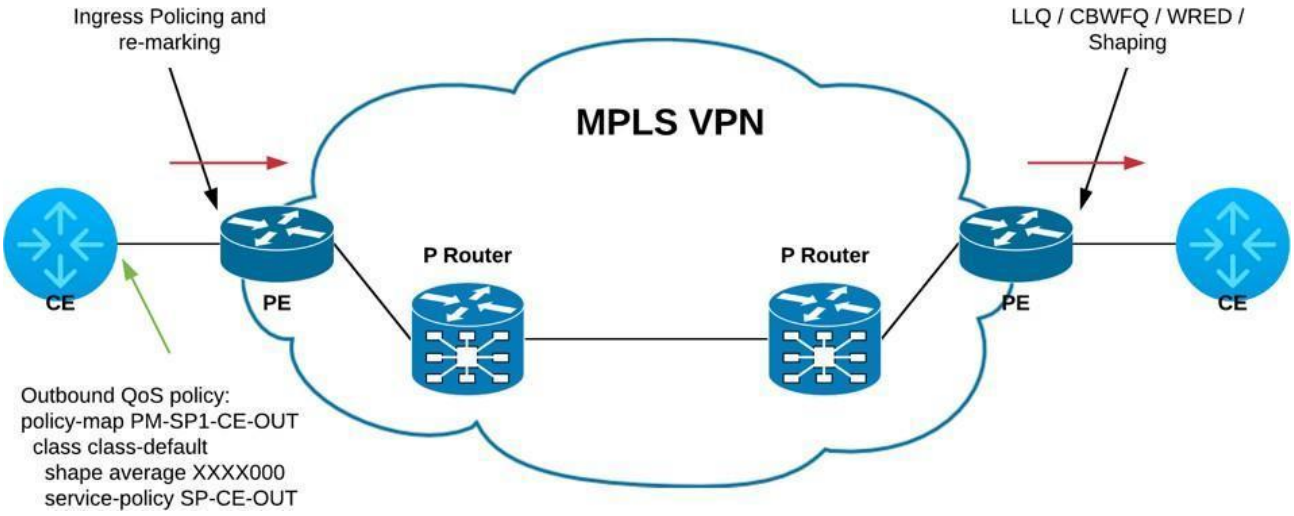


Figure 1: QoS for MPLS VPN

With MPLS VPN, QoS is needed to prioritize certain critical or network-sensitive traffic during congestion to deliver the desired SLA. The VMware VeloCloud SD-WAN feature Dynamic Multipath Optimization™ (DMPO) prevents congestion natively, and therefore does not depend on MPLS QoS for traffic classification and prioritization. In the subsequent sections we will focus mostly on the originating CE and PE routers, and how their QoS policies may cause challenges when integrating VMware VeloCloud SD-WAN.

VMware VeloCloud SD-WAN DMPO and QoS

Depending on the routing and business policy configurations, traffic processed by VeloCloud SD-WAN can be:

1. Direct (aka the underlay): forwarded directly on different interfaces without tunneling.
2. Tunneled within VeloCloud Multipath Protocol (VCMP) header (aka the overlay). This traffic is subject to Dynamic Multipath Optimization (DMPO) processing as discussed in the next paragraph.
3. Tunneled within IPsec when connecting to Cloud Security Service providers (CSS) or to Non-SD-WAN Destinations (NSD).

Except for the features related to DMPO (e.g. FEC or jitter buffering), the QoS mechanisms stated in this white paper shall apply to all of the above forwarding options (Direct/CSS/NSD).

DMPO is a collection of techniques used by the VMware VeloCloud 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 VeloCloud components that process and forward data traffic. The overlay tunnels encapsulate the original payload in 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 VeloCloud data plane utilizes the overlay headers to apply the configured QoS policies as well as the appropriate on-demand remediation techniques such as FEC and jitter de-buffering for real-time traffic. The details on how DMPO handles link steering decisions for real-time traffic vs TCP, link aggregation for TCP traffic, and the remediation techniques can be found in the white paper Application Performance with Dynamic Multipath Optimization™ (DMPO).

Figure 2 shows an example where there are two WAN transport types available at each branch VMware VeloCloud SD-WAN 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 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 VeloCloud 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.

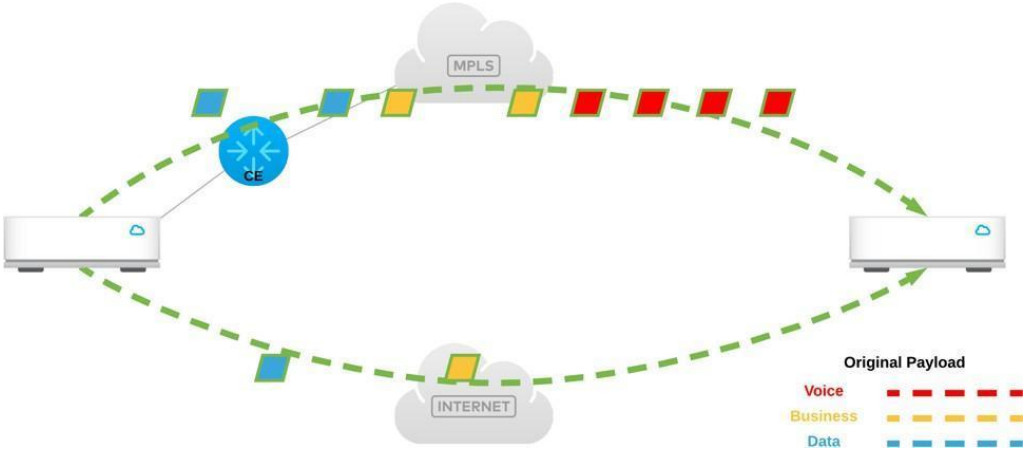


Figure 2: VeloCloud DMPO example

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 re-order 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. The challenges of integrating DMPO with existing MPLS QoS policies are discussed in subsequent sections.

Schedulers

The VeloCloud data plane implements two sets of schedulers shown in Figure 3, which influence the QoS of outbound traffic from the VeloCloud 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.

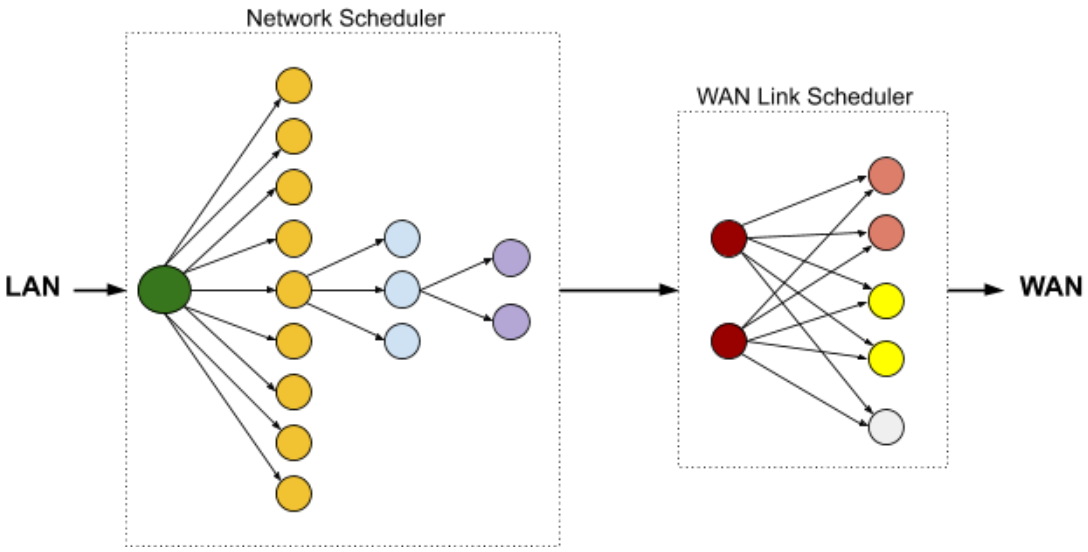


Figure 3: VMware VeloCloud SD-WAN QoS Scheduler

The entire QoS processing flow with respect to the hierarchical queuing construct is outlined in Figure 4, which adds color to the high-level diagram shown in Figure 3. In the ensuing sections, we will divide the flow into two main components—Network Scheduler and Link Scheduler—and deep dive into the scheduler and its different levels of queues as shown in Figure 4.

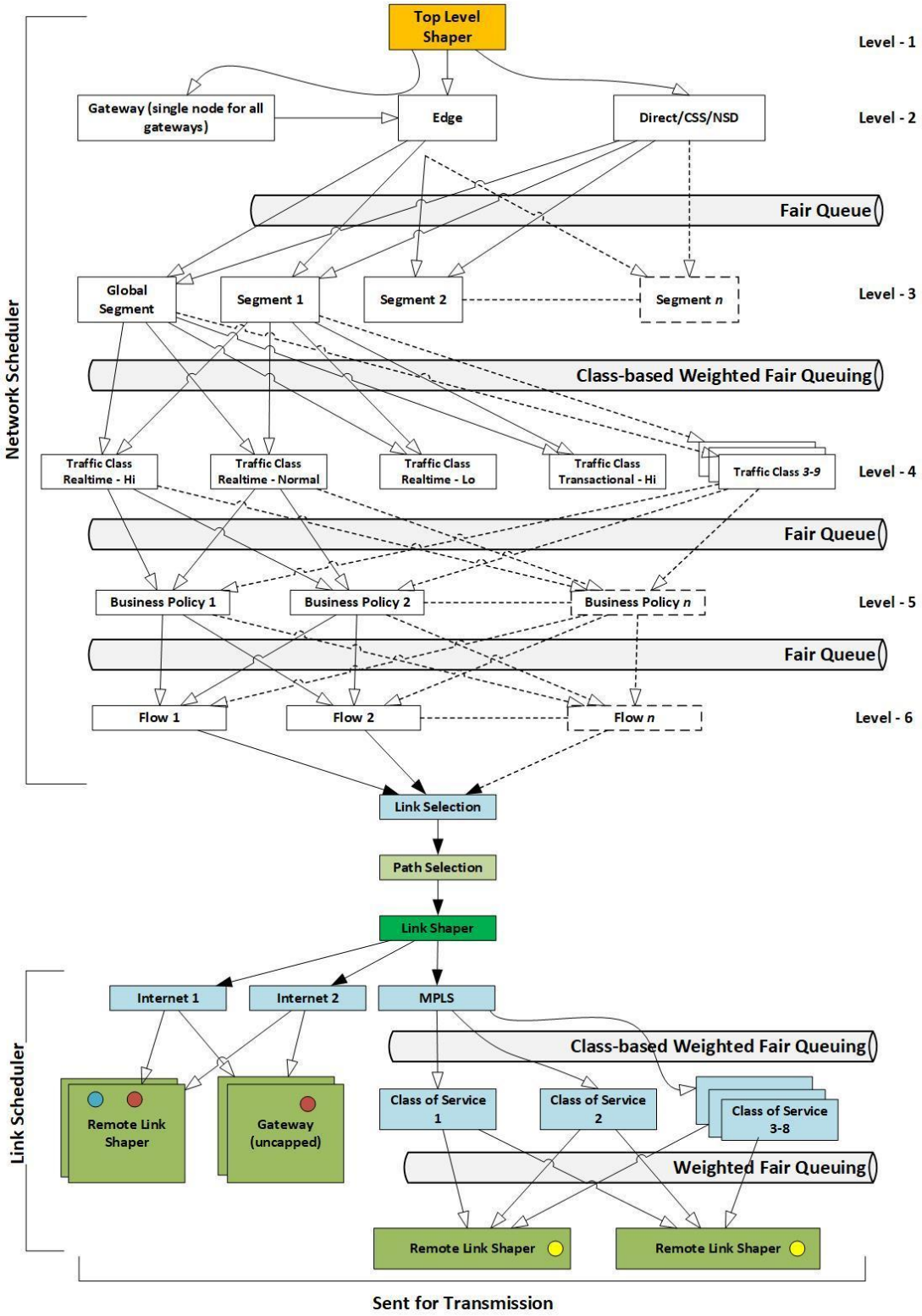


Figure 4: VeloCloud QoS, its schedulers and hierarchical queues

Network Scheduler Details

The network scheduler is the primary place where QoS is enforced in the VeloCloud data plane. This is where the top-level shaper limiting outbound traffic to the sum of WAN overlays is implemented and applied. Referring to the "Network Scheduler" portion of Figure 4, rate limiting is applied at the QoS level for each peer, ensuring an even distribution of traffic, QoS and aggregation performance guarantee per peer. Level 3 nodes in the hierarchy shown in Figure 4 contain the segments of which an enterprise can have multiple, with Fair Queueing applicable to the nodes. The shaper for the nine Traffic Classes—3x3 matrix of the Service Classes (Real Time, Transactional, Bulk) and Priority (High, Normal, Low)—is then applied per segment to each peer node; these are the Level 4 nodes. The queueing mechanism for the nine traffic classes employs Class-Based Weighted Fair Queueing where the weights of the classes are user-configurable. The next level in the hierarchy consists of the business policies implemented with Weighted Fair Queueing where each Business Policy maps to one of the nine traffic classes based on the configured priority and service class. The last level of leaf nodes in the Network Scheduler consists of the flows going outbound from the Edge. Every network flow going through the Edge must match a business policy and therefore be placed in the appropriate class of service.

The aforementioned six hierarchical levels construct the basis of the network scheduler, with each level described in more details below. The queues within each level of the Network Scheduler are also referred to as "nodes" in the following subsections.

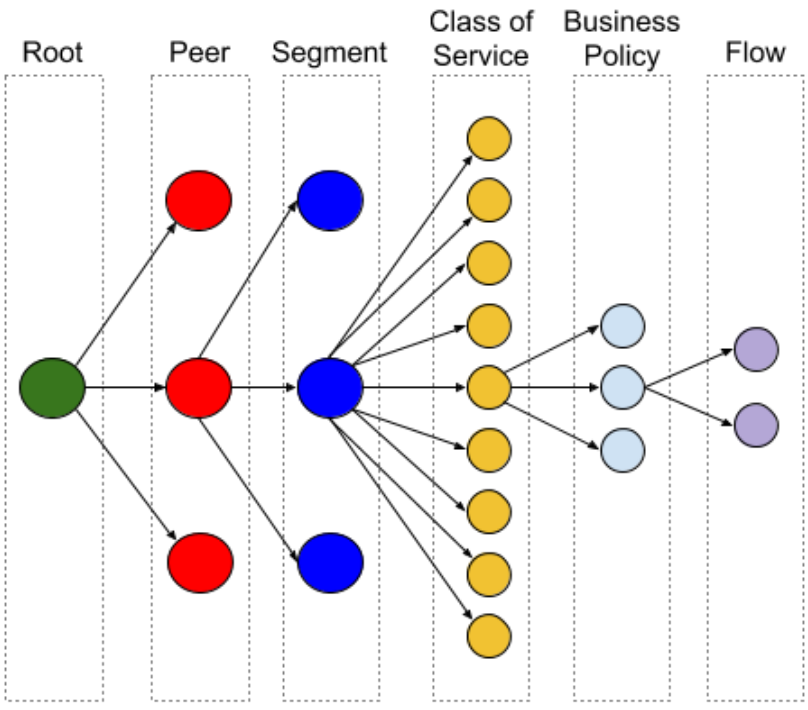


Figure 5: VeloCloud network scheduler queues

Root

The root level of the scheduler is rate limited to the sum of the upstream bandwidth of all WAN links at the site or the max capacity of the license, whichever is lower.

Peer

The second level of the Network Scheduler consists of peer nodes that represent the destination of the traffic. Peer nodes can be other Edges, Direct Internet or Gateway. All VeloCloud Gateways are combined into a single peer node in order to enforce

"SD-WAN Service Rate Limit". Peer nodes have equal fair share and are rate limited to the capacity of the individual peer's available aggregate bandwidth.

SD-WAN Overlay Rate Limit

SD-WAN Overlay Rate Limit is an optional feature under the Business Policy framework that can be utilized to rate limit overlay traffic sent to the VeloCloud Gateways; it does not affect direct traffic or overlay traffic sent multi-path to other Edges. This feature enforces a limit on how much traffic may be sent to and from all the connected VeloCloud Gateways.

When this setting is enabled:

- On the Edge:
  - The Gateway peer node in the Network Scheduler is rate limited to the configured value.
  - This ensures that the sum of traffic through all Gateways will not exceed the allowed rate.
- On the Gateway:
  - The Edge peer node in the Network Scheduler is rate limited to the configured value.
  - This ensures that each individual Gateway cannot exceed the allowed rate. However, multiple Gateways may exceed the permitted rate unless rate limiting is done at the Provider Edge.

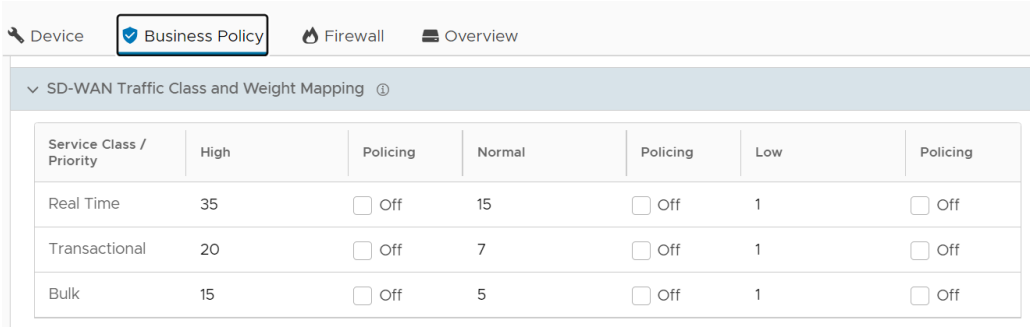
Segment

The Level 3 queues in the Network Scheduler are made up of segments. The segment nodes represent the network segments on the Edge. When there is more than one segment configured on the Edge, each segment will get a fair share when bandwidth is under contention.

Traffic Classes

Each segment node contains the Level 4 traffic classes. The nine traffic classes are defined on the Business Policy page and are derived from application service class and priority. A traffic class's fair share is set to the configured weight, and rate limited if policing is enabled on the user interface.

The weights of the service classes are configurable on a per-profile basis and as an Edge override. The weights do not have to add up to 100 and each class can burst up to 95% of the peer node bandwidth.



Service Class / Priority	High	Policing	Normal	Policing	Low	Policing
Real Time	35	Off	15	Off	1	Off
Transactional	20	Off	7	Off	1	Off
Bulk	15	Off	5	Off	1	Off

Figure 6: VMware VeloCloud SD-WAN traffic classes

Business Policy

Branching from each of the traffic classes, the business policies form the Level 5 nodes. A single node represents a business policy which a given flow matched. If the business policy contains a rate limit, it is enforced here to ensure that all matching flows are hierarchically rate limited to the specified rate.

Flow

The last level of the Network Scheduler contains nodes representing the traffic flows going through the Edge. Consequently, each flow matches a business policy that rolls up to a traffic class. Dequeuing is always done from the flow node.

Network Scheduler Summary

Putting all the above concepts together, Figure 7 summarizes the hierarchical QoS structure and how the queueing mechanisms are applied for each level of the network scheduler. The left-hand side of the diagram also shows the corresponding configuration that generates the queues/nodes. After the packet is processed by the aggregate network scheduler, path selection is performed to select the best path based on the real-time tunnel performance and path parameters before the packet is processed by the link shaper.

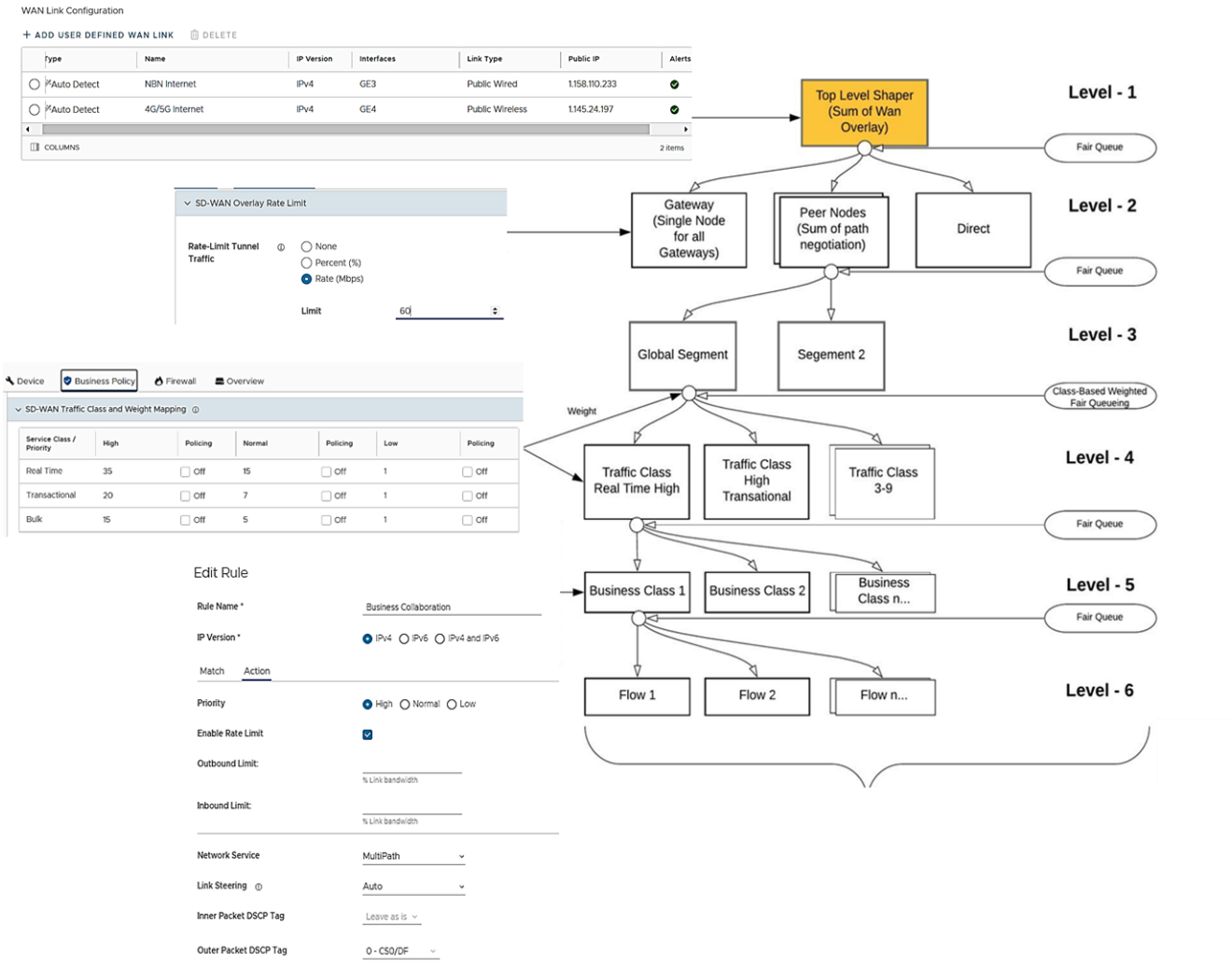


Figure 7: Network Scheduler details

Path Selection

This section describes the two cases for path selection:

- Overlay Path Selection
- Direct/CSS/NSD Link Selection

Overlay Path Selection

Overlay between two SD-WAN endpoints is referred to as path. A path in QoS perspective has a Local Link and a Remote Link context attached to it:

- Local Link: Link local to the device
- Remote Link: Link in the remote device where the overlay path terminates

Overlay path quality is monitored. The stability of the paths are tracked in different stability states:

- Peer node bandwidth cap depends on the actual bandwidth available to the peer. Any change in path state is reflected in the peer node bandwidth cap.
    - Peer Node Bandwidth Cap = Min (Sum(Local Link Tx), Sum(Remote Link Rx))
    - Ensures remote link bandwidth is not overrun by the sending edge

Once a path to a SD-WAN endpoint becomes unstable, the peer node in the NetQoS hierarchy shall be updated to reflect the current actual bandwidth available to that SD-WAN endpoint.

Path selection

- Pathscore = (Bytes-Queued Local-Link / Local-Link-Tx-Bandwidth) + (Bytes-Queued Remote-Link / Remote-Link-Rx-Bandwidth) + Path Delay
- Stable paths having the lowest path score are chosen for packet transmission
- This decision is made on a per packet basis—Achieves per packet aggregation (single flow can be aggregated on two links)

Direct/CSS/NSD Link Selection

1. Direct/CSS Link Selection mechanism:

1. Iterate through all WAN links to find stable link without any Dynamic Error Correction (DEC) flag per category (voice/video/transactional)
    a. If a single clean link is found – return link and Exit.
    b. If multiple clean links are found, then return the link with the best bandwidth. Else go to step 2.
2. Iterate through all WAN links to find link without loss flag set
    a. If a single link without a loss flag is found – return link and Exit.
    b. If multiple links are found, return link with best bandwidth. Else go to step 3.
3. Iterate through all WAN links to find link without latency flag set
    a. If a single link is found – return link and Exit.
    b. If multiple links are found, return link with best bandwidth. Else go to step 4.
4. Iterate through all WAN links to find stable link (ignore DEC/loss/latency flags)
    a. If a single link is found – return link and Exit.
    b. If multiple links are found, return link with best bandwidth. Otherwise return error.
5. Iterate through all WAN links to find SD-WAN Service Reachable link (ignore DEC/loss/latency flags)

a. If a single link is found – return link and Exit.

b. If multiple links are found, return link with best bandwidth. Otherwise return error.

2. NSD Link Selection mechanism:

1. Iterate through all Primary IPsec tunnels
    a. If a single Primary tunnel is found – return link and Exit.
    b. If multiple Primary tunnels are found, the Direct/CSS Link Selection applies and return link. Else to step 2.
2. Iterate through all Secondary IPsec tunnels
    a. If a single Secondary tunnel is found – return link and Exit.
    b. If multiple Secondary tunnels are found, the Direct/CSS Link Selection applies and return link. Else to step 3.
3. No Primary or Secondary tunnels found, then drop the packet

Link Scheduler Details

The link scheduler provides two levels of hierarchy, representing the transmit rate of the local WAN link(s) and the receive rate of remote WAN link(s). The local link shaper applies traffic shaping to outbound packets while the remote link shaper ensures the sending Edge does not overwhelm the remote link bandwidth. A third and fourth level of hierarchy for MPLS Class of Service is possible and is described in the section below.

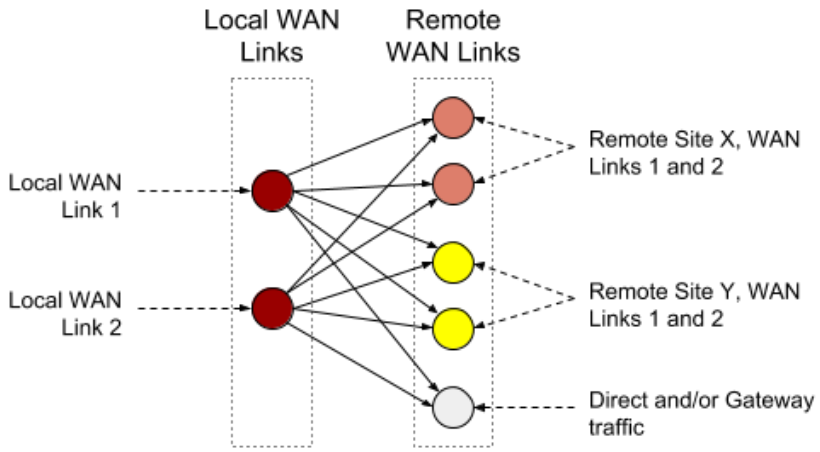


Figure 8: Link Scheduler summary

MPLS Class of Service

In some cases, customers choose to continue to leverage the Classes of Service provided by the Service Provider on the private WAN link. The Class of Service definitions under the Private Link Configuration of the WAN Overlay create the classes as available "WAN links" in Business Policy as shown in Figure 9.

Consider the below configuration, which maps the Service Provider CoS into three Classes of Service: Real-Time, Data, and Backup and their respective DSCP tags. This configuration creates an additional level in the link scheduler with the defined classes for the private WAN link. This allows traffic to be steered to MPLS with the proper rate limit and the appropriate Outer tunnel DSCP tag that maps to the Service Provider CoS. This approach, however, does not fully alleviate the issues integrating VeloCloud DMPO and MPLS CoS. The challenges in detail are described in the latter sections.

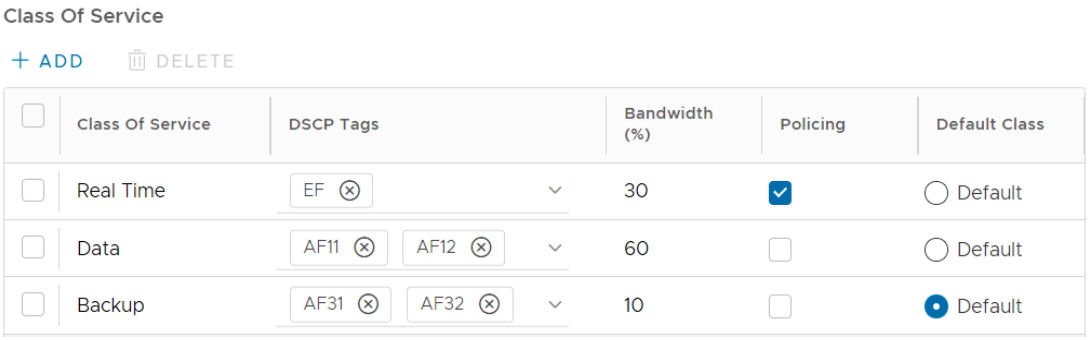


Figure 9: Private WAN Overlay Class of Service (CoS) Configuration

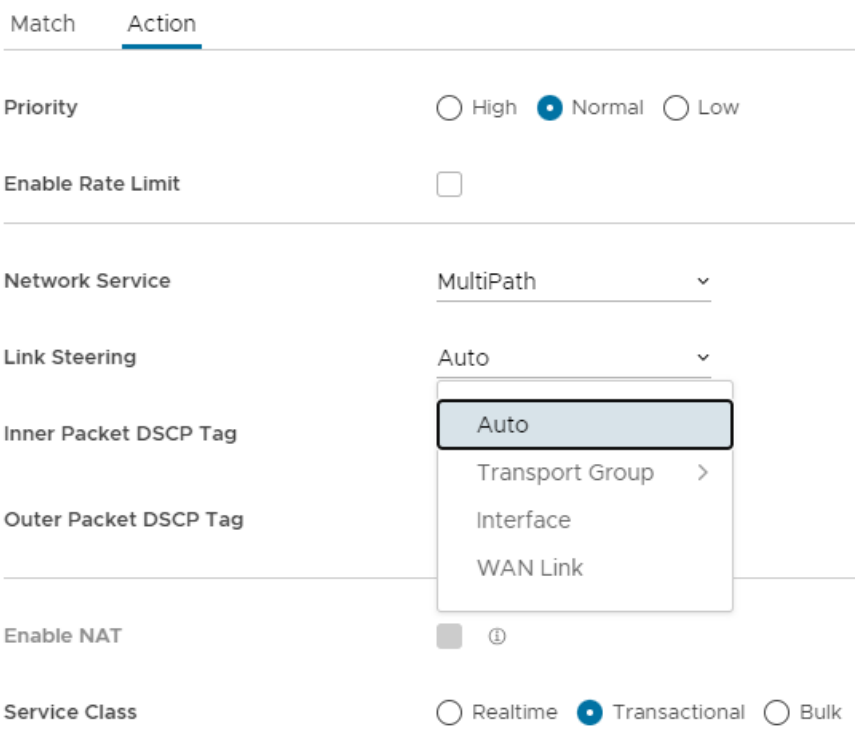


Figure 10: Link steering via business policy

Link Scheduler Summary

Operations of the link scheduler are summarized in the Figure 11. First, link selection and path selection are performed before the packet is processed by the link scheduler. In VeloCloud terms, each overlay tunnel between Edges or VCGs is referred to as a path and it is not a one-to-one mapping with the WAN Overlay interface. Multiple WAN overlay tunnels can be created on the same WAN interface with different local and remote links. As previously described in the functions and benefits of DMPO, overlay path quality is continuously monitored and stability of the paths is tracked in real time. The stable path having the lowest path score is chosen for packet transmission, and the decision is made on a per-packet basis to achieve link aggregation and other remediation techniques.

After path selection is complete, the packet is handed to the link scheduler for transmission. In this example, if the path selection selected Internet1 as the best available interface, the packet would get enqueued to the Internet1 link before it is then handed to the remote link shaper. Upon passing through the remote link shaper, the packet is forwarded for transmission.

If an MPLS link was chosen during link selection and MPLS "Class of Service (CoS)" was defined under the Private WAN link configuration, the CoS classes are created as Class-Based Weighted Fair Queues making up an additional level of hierarchy after the Local Link level to shape outbound traffic appropriately toward the MPLS link. Similar to traffic steered toward the public WAN links, after the packets pass through the CoS queues, they are handed to the remote link shaper and then sent for transmission.

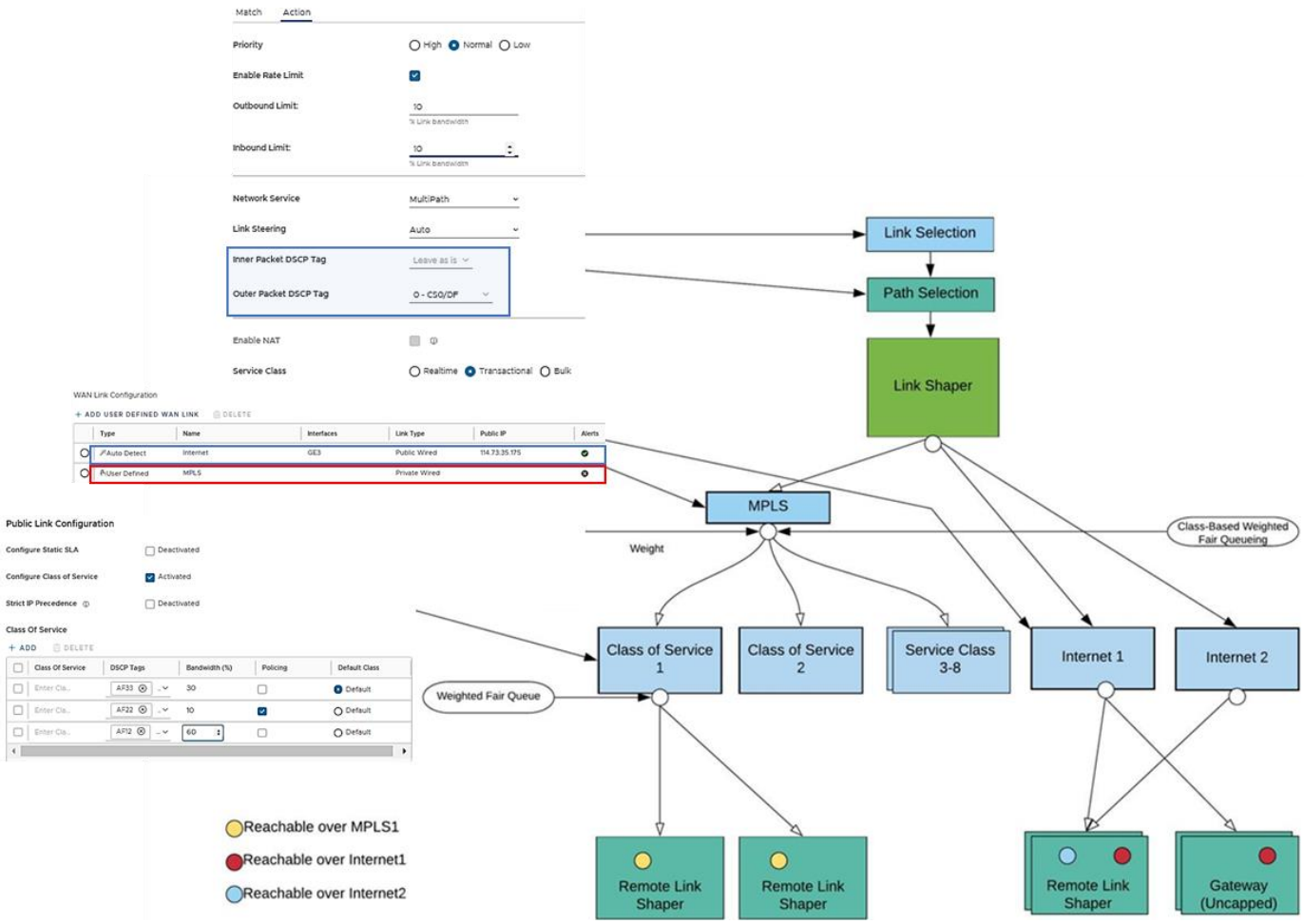


Figure 11: Link Scheduler details

Challenges in 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 re-ordering 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.

Re-Ordering of Overlay Packets by MPLS QoS

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

As described previously under section "VeloCloud DMPO and QoS", for the data plane to achieve per-packet load balancing, the VMware VeloCloud 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 12. 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.

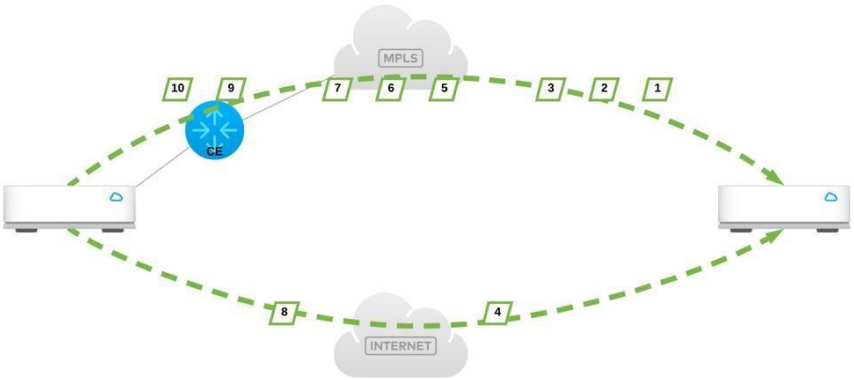


Figure 12: Single flow traversing multiple paths across overlay

Let's assume MPLS is the superior link to 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 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 with various remediation techniques such as overlay loss measurement, dynamically enable/disable error correction on a specific overlay path due to loss and, retransmit lost packets.

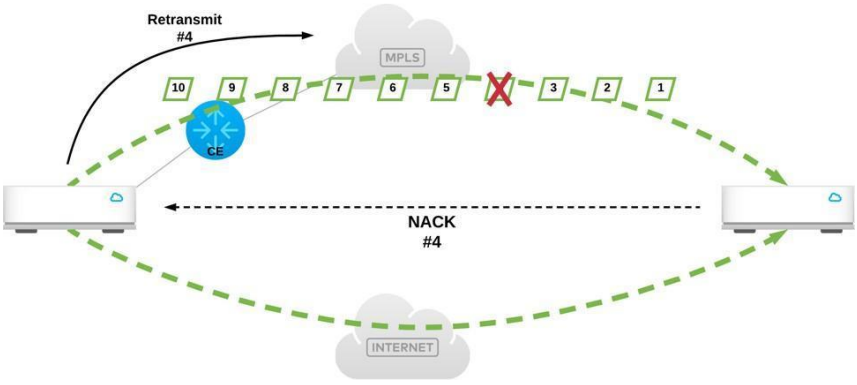


Figure 13: Overlay packet re-transmission

The issue comes in when Class of Service is enabled in the private WAN Overlay configuration as shown in the previous section 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 the three classes EF (6 packets in red), AF11 (1 packet in yellow), and AF31 (3 packets in blue), MPLS QoS may re-order the packets which the transmitting Edge sent based on VeloCloud 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. The problem is depicted in the diagram below.

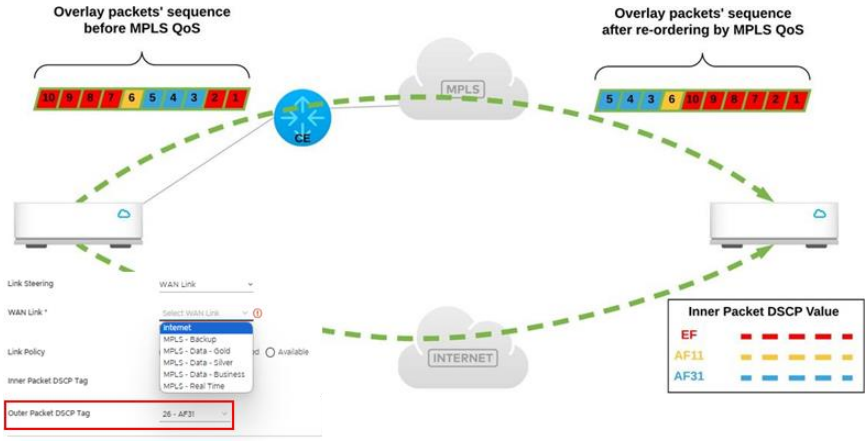


Figure 14: Overlay packets being re-ordered due to MPLS QoS

In the session load balancing world where DMPO techniques are not implemented, link aggregation and real-time remediation do not apply and therefore re-ordering 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 the 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 re-ordering of the packets due to DSCP markings of the outer overlay header creates several potential problems as the Edge's DMPO will misinterpret the re-ordering as loss and jitter across the MPLS path. The unintended behavior may result in the following consequences on the Edge:

- It may cause the transmitting Edge to avoid the MPLS link
- VeloCloud DMPO may apply unnecessary error correction and consume more bandwidth
- It may cause the receiving Edge to discard packets that arrive "late"

As we have pointed out in the previous sections, VeloCloud DMPO already assures congestion avoidance and traffic prioritization with real-time monitoring, per-packet steering and QoS operations:

- 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.

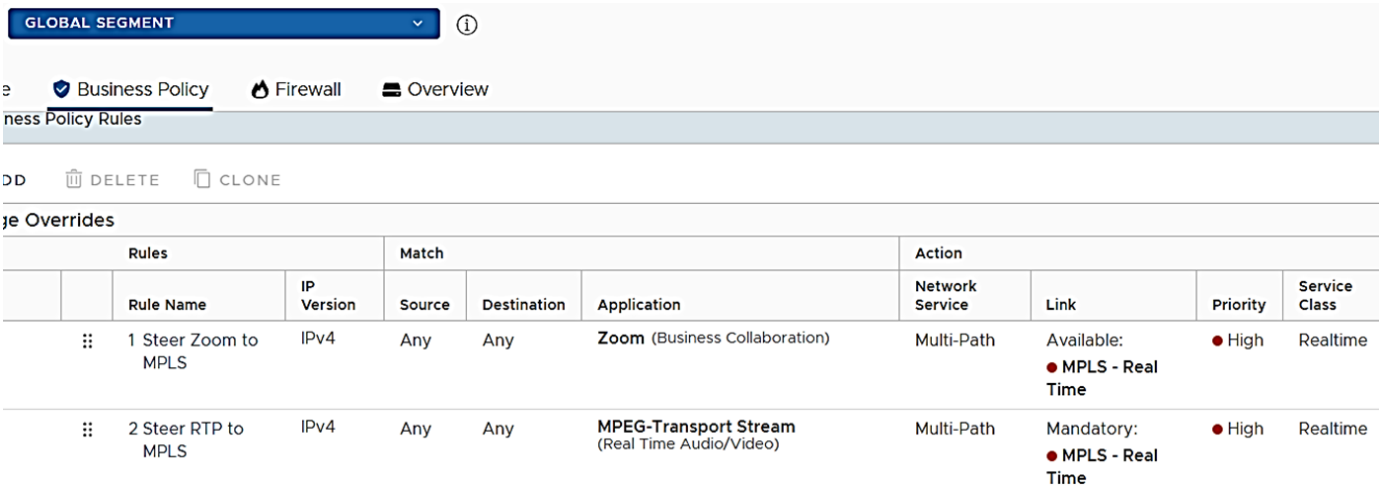
- In the event of congestion, the Edge backs off sending on an overlay path gracefully until the congestion state has recovered.

Please note this behavior applies to all releases prior to 3.4.0. If you are using any release earlier than 3.4.0, it is optimal to use a single MPLS class of service for all overlay traffic. Starting from release 3.4.0, "Per CoS Path Quality Measurements and Optimization" has been implemented. This feature provides solution to the re-ordering of packets by MPLS QoS issue. The feature is discussed in the "Recent QoS Enhancements" section.

Impact of Link Steering with Business Policy on QoS

The second challenge we see encompasses the 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 to forward the most critical traffic on the best WAN link—i.e. put voice and streaming video on MPLS link. However, forcing traffic over a particular path with 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.

Assume a hybrid SD-WAN branch has two WAN links and business policies as shown in the diagram below are configured on the transmitting Edge to steer RTP and Zoom application traffic toward MPLS. In this case let's assume 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 aforementioned network scheduler will process the packets through the six levels of QoS operations and shape the traffic to 20 Mbps as discussed in "Network Scheduler Details''. QoS on the VMware VeloCloud SD-WAN Edge is as expected at this point with each traffic class fairly scheduled at 10bMbps. Link selection and path selection follows the network scheduler as the next steps.



		Rules		Match			Action			
		Rule Name	IP Version	Source	Destination	Application	Network Service	Link	Priority	Service Class
	⠿	1 Steer Zoom to MPLS	IPv4	Any	Any	**Zoom** (Business Collaboration)	Multi-Path	Available: ● **MPLS - Real Time**	● High	Realtime
	⠿	2 Steer RTP to MPLS	IPv4	Any	Any	**MPEG-Transport Stream** (Real Time Audio/Video)	Multi-Path	Mandatory: ● **MPLS - Real Time**	● High	Realtime

Figure 15: 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 impact 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 16.

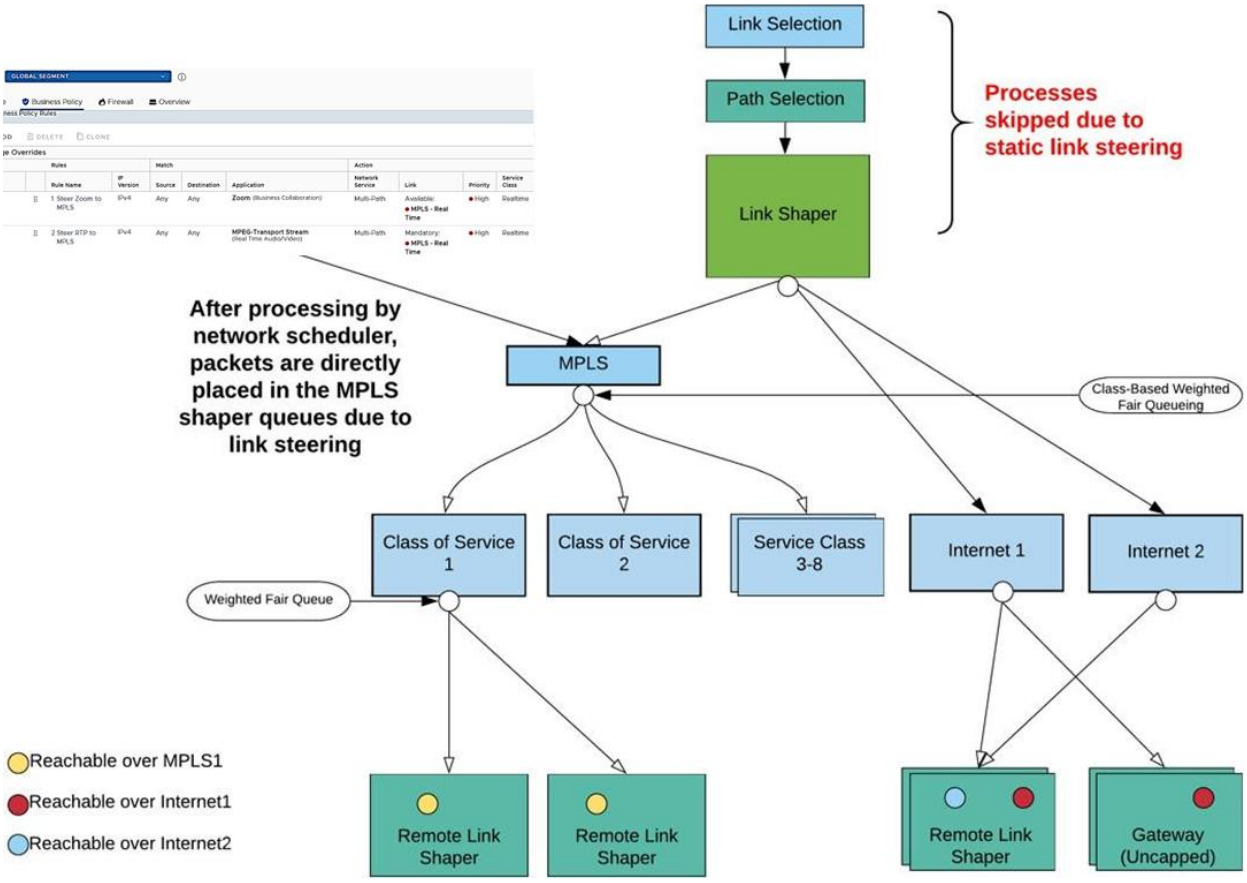


Figure 16: Static link steering impact on link scheduler

Given the challenge discussed, it's recommended to configure Auto link steering policy whenever possible. To mitigate the issues from manual link steering, the supplementary feature "Priority Classes in Link Scheduler" has been added to release 3.3.0. More details are provided in the "Recent QoS Enhancements" section.

Recent QoS Enhancements

As the previous sections have been focused on VeloCloud QoS and the existing challenges, in this section we will discuss some of the improvements in the recent releases.

DSCP Marking for Both Underlay and Overlay Traffic

From software release 3.3.0 and onward, VMware VeloCloud SD-WAN supports DSCP re-marking of packets forwarded by the Edge to the underlay. In prior releases, DSCP re-marking was only supported for traffic sent through the SD-WAN overlay tunnels, and selective customers faced challenges when the Edge was deployed as a CE router with direct connectivity to the PE router. DSCP for traffic sent underlay to the MPLS prefixes learned via BGP was unable to be re-marked by the Edge before. In release 3.3.0, the VMware SD-WAN Edge can re-mark underlay traffic forwarded on a WAN link as long as "Underlay Accounting" is enabled on the interface. DSCP re-marking is enabled in the Business Policy configuration as shown in Figure 17. In this example, assuming the Edge is connected to MPLS with both underlay and overlay traffic forwarded MPLS, if traffic matches the network prefix 172.16.0.0/12, the Edge will re-mark the underlay packets with DSCP value of 16 or CS2 and ignore the "Outer Packet DSCP Tag" field. For overlay traffic sent toward MPLS matching the same business policy, the DSCP value for the outer header will be set to "Outer Packet DSCP tag".

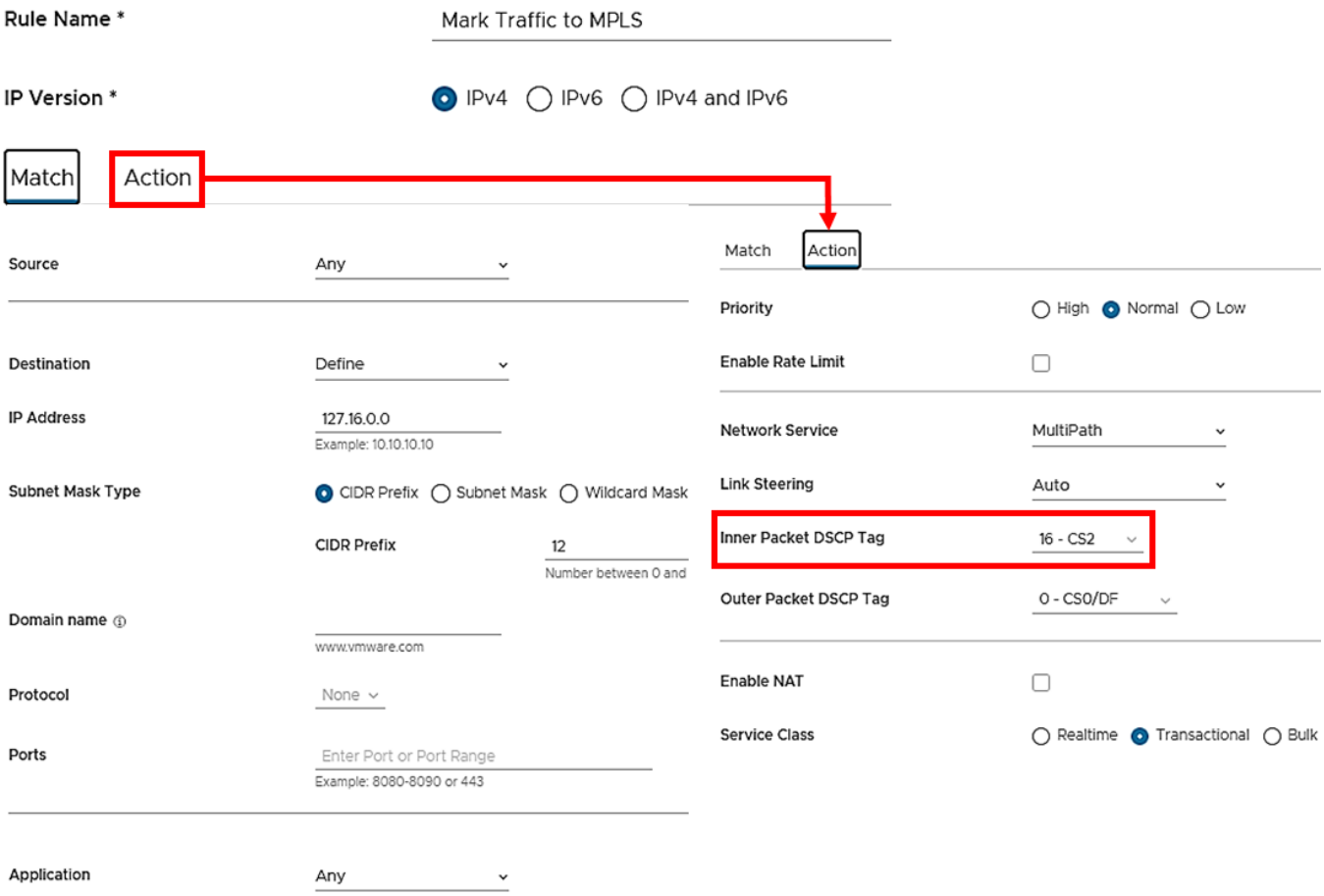


Figure 17: DSCP re-marking of underlay traffic

Priority Classes in Network Scheduler

Another QoS improvement added to release 3.3.0 encompasses implementation of priority classes in the network scheduler. The support for prioritization in the network scheduler (High > Normal > Low) mitigates static link steering oversubscribing a specific WAN link—which causes queue build up and congestion. In the same example discussed in the section "Challenges in Integrating DMPO and MPLS QoS" where a hybrid SD-WAN branch has two WAN links and business policies are configured on the transmitting Edge to manually steer RTP and Zoom traffic toward MPLS, in prior releases these real time application flows may experience high latency and jitter if the MPLS link is already saturated. With this feature improvement, the network scheduler will prioritize the higher QoS priority traffic in the queue and therefore give preferred treatment for business-critical applications such as VoIP and RTP.

This feature enhancement does not fully resolve the challenges from static link steering as manual steering can still over-saturate the WAN link if the total throughput of high priority traffic exceeds the bandwidth of the WAN link. That is not uncommon and Auto steering is highly recommended for this type of traffic profile.

Per CoS Path Quality Measurements and Optimization

From release 3.4, QoS improvements in the VMware VeloCloud SD-WAN solution involve implementing per-CoS (Class of Service) monitoring and scheduling to better integrate with MPLS VPN and existing service providers' QoS architecture. The Edge will measure path performance metrics such as loss, latency, and jitter for each CoS and treat each service class as different overlays from a QoS perspective. This will be a departure from existing functionality where all packets sent through

the overlay are combined to calculate the path quality measurements—which causes potential issue where overlay packets with CoS of CS0 may impact the performance of EF traffic sent through the same overlay tunnel. Per CoS monitoring, queueing and scheduling will also resolve the adverse effects from re-ordering of overlay packets by MPLS QoS as described previously.

Class of Service Mapping under Public Wireless Links

Prior to release 4.3.0, class of service mapping allows an Edge to create sub-paths under a private MPLS link by tagging a DSCP value to a given application within an MPLS class of service hierarchy. This creates a sub-path within DMPO across the private link which allows for individual SLA metric measurement and link steering and remediation treatment of each sub-path or differentiated service on the MPLS interface. From release 4.3.0, this feature has been extended to support public wireless links, so that differentiated services can be defined for a set of applications using DSCP values, just like what can be done under private MPLS links. Typical use case is a private wireless network, where DSCP to QCI mapping is done at the LTE network core. Please note that in order to support Class of Service Mapping, all components (Edge, Gateway, Orchestrator) must be running release 4.3 or newer. Below is an example where voice and video traffic are marked with DSCP EF and data traffic is marked with DSCP AF31 under the same wireless link.

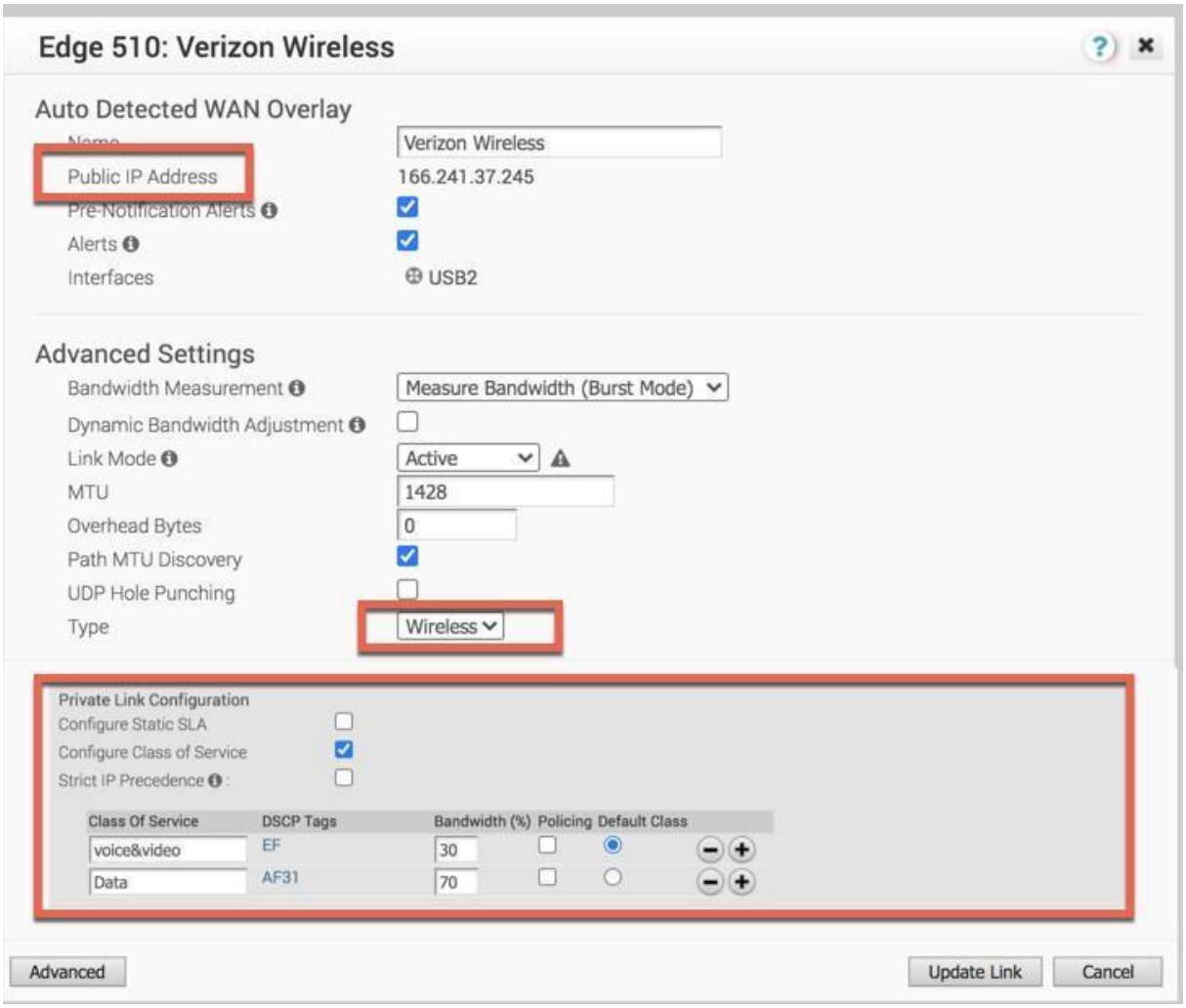


Figure 18: DSCP marking for voice and video traffic on a wireless link

Best Practices and Recommendations

This document covered the VMware VeloCloud SD-WAN QoS architecture, its operations and schedulers, and described in detail the challenges in integrating DMPO, VeloCloud QoS and traditional MPLS QoS. Some best practices and recommendations are listed here to optimize integration.

Use Single Class of Service for Overlay Traffic Through MPLS VPN

If you are using a release prior to 3.4.0, our recommendation is to use a single MPLS class of service for all overlay traffic. When integrating VMware SD-WAN with MPLS QoS, the "Challenges in Integrating DMPO and MPLS QoS" section discussed the issue where MPLS may potentially re-order 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. Due to the functionalities of DMPO and its remediation techniques, re-ordering of overlay packets by MPLS QoS may cause the packets received by the receiving Edge to be out of sequence and lead the Edges to misleadingly detect loss and jitter across the MPLS overlay path. The symptoms are often seen when low-priority traffic sent through the MPLS path is dropped, causing the Edges to detect degradation on the path and start applying remediation—which in turn causes the forwarding of other high-priority traffic to be impacted. Because VeloCloud DMPO performs per-packet steering and innately 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.

If you are using release 3.4.0 or newer and your MPLS service provider treats traffic differently based on class of service settings, our recommendation is to use different MPLS class of service for overlay traffic accordingly. The reason is that since release 3.4.0, each MPLS class is given its own unique set of sequence numbers. This facilitates measuring the loss, latency and jitter of each class of service independently and prevents out-of-ordering from happening due to intermediate queueing.

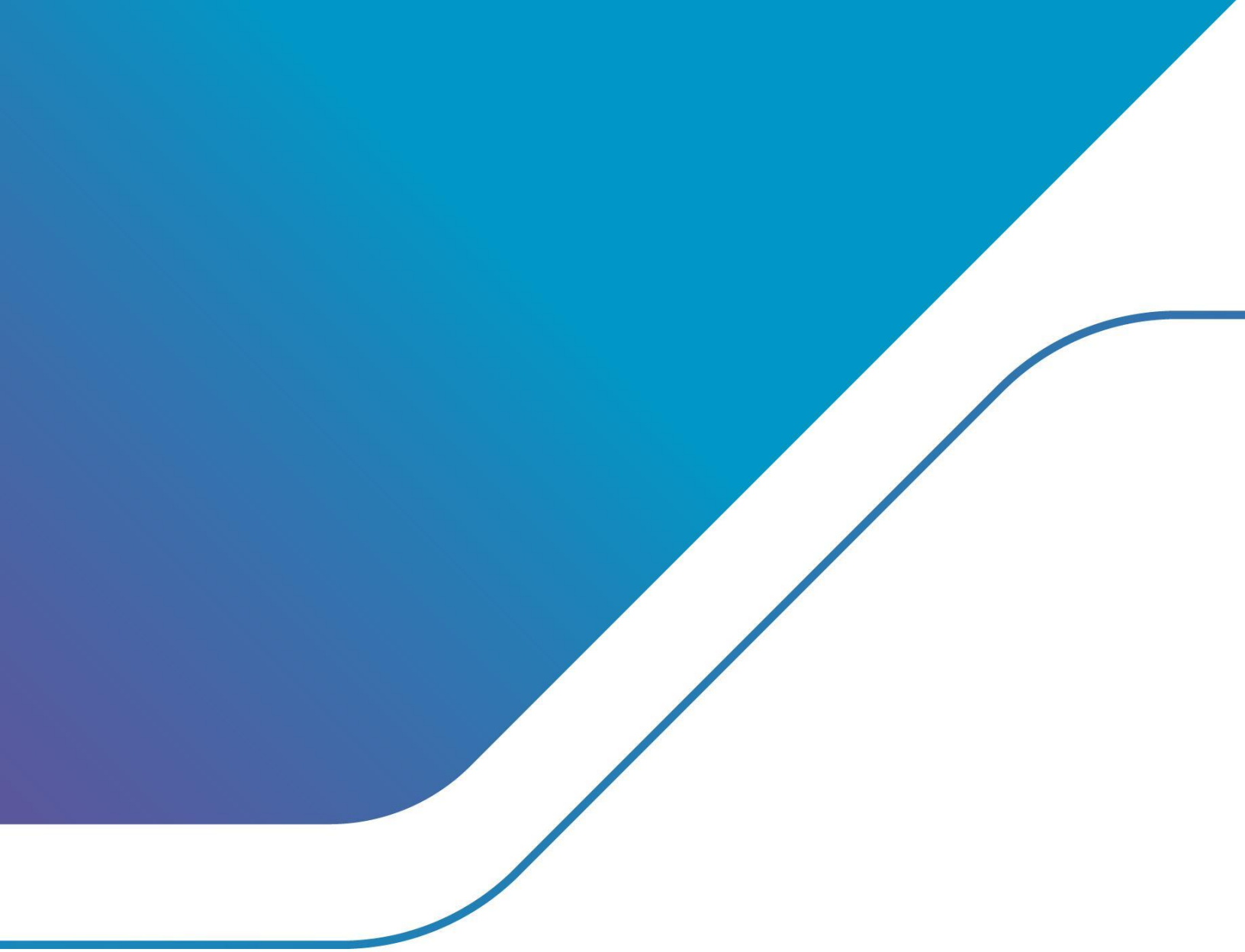
Use Auto Link Steering

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, the 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 therefore 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 VeloCloud SD-WAN, because DMPO monitors end-to-end loss/latency/jitter and inherently steer application packets to the best path and applies 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.



Item No: sdwan-702-qos-overview-so-0719