PAN-OS SD-WAN

Path Selection Primer

Palo Alto Networks, Inc.
www.paloaltonetworks.com

Revision Date: January 15, 2022

Table of Contents

The Importance of Fast and Accurate Path Selection

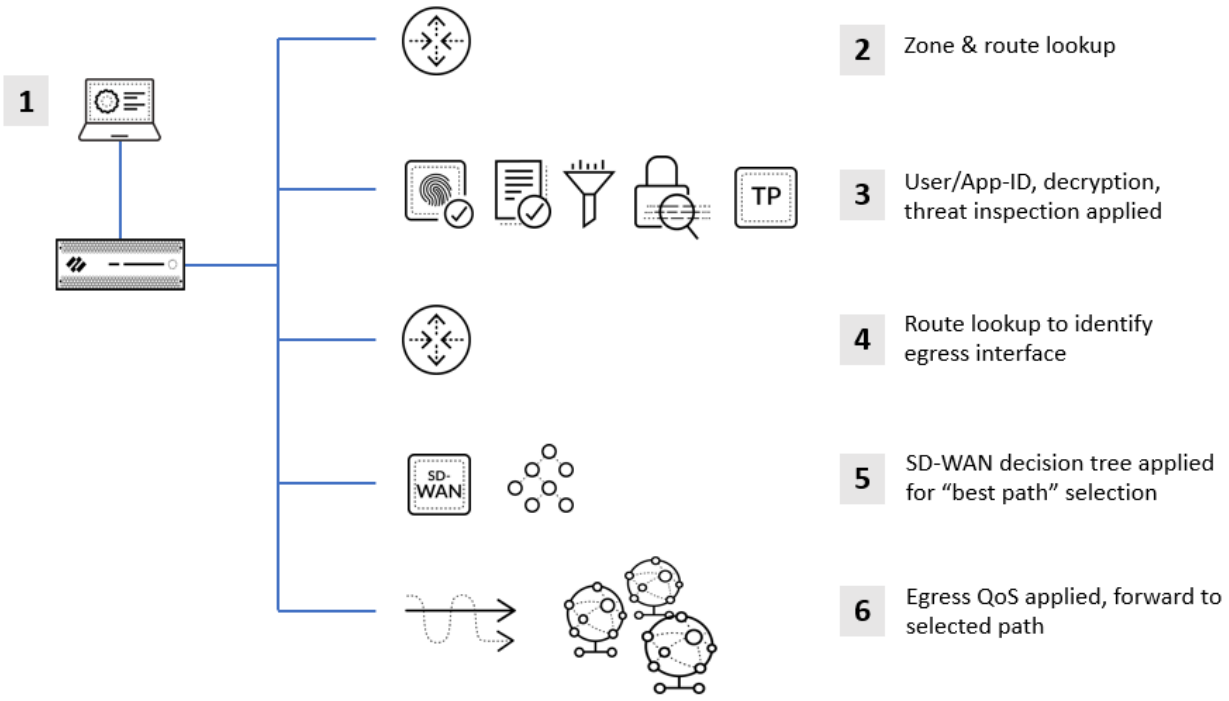
The ability of an SD-WAN solution to properly analyze and select the best path to ensure the highest level of application user experience is the hallmark of SD-WAN's basic functionality. PAN-OS SD-WAN uses multiple technologies to provide rapid path analysis, accurate path selection, and provides administrators with granular controls to tune the application failover experience.

As applications are not created equal, one application's response to path degradation can differ greatly from another and it is important to have features that allow granular adjustments as to how the path is measured. How frequently health probes are sent, how to weight health check parameters, and how to combine them with other health checks and failover mechanisms are all required. PAN-OS SD-WAN provides all of these capabilities for advanced policy creation.

In this paper, core concepts around path health checks and path selection are discussed to give a deep understanding of the PAN-OS SD-WAN building blocks which allows you to create advanced profiles and policies to control the user experience for critical business applications.

SD-WAN Processing Overview

When a new session arrives at the firewall, it will go through multiple stages for security processing and path selection by the SD-WAN engine. The following illustration shows how a new session is processed by the PAN-OS SD-WAN firewall and the decision trees used. The sections that follow will expand on the SD-WAN functions to provide greater detail.



SD-WAN Packet Flow

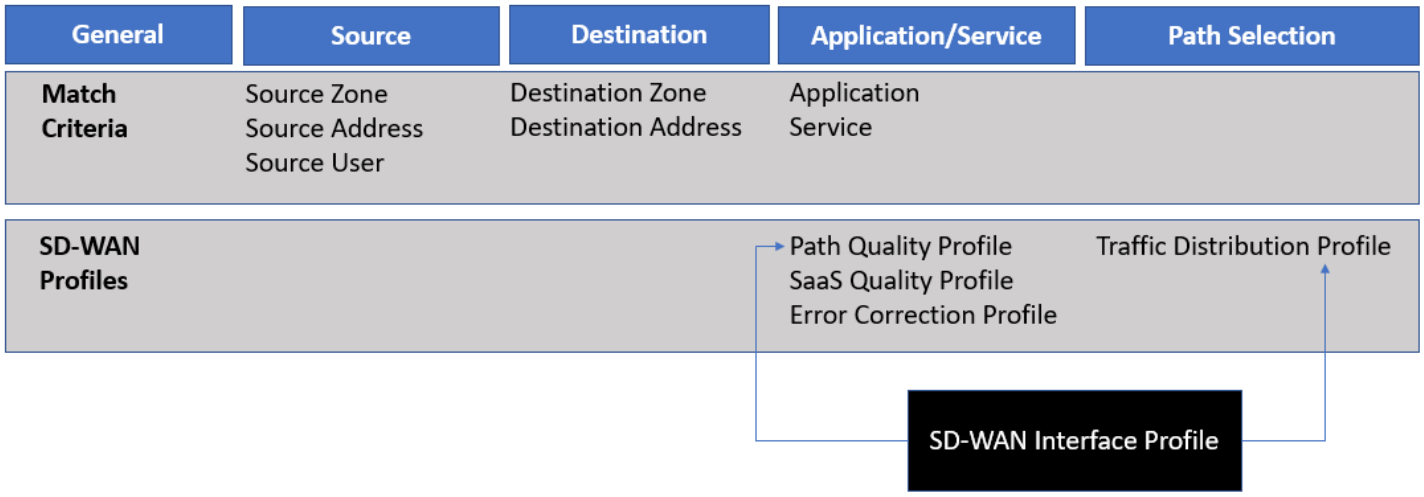
1. A new session is created by the client host and sent to the firewall for security and SD-WAN processing.
2. The firewall checks its session table for an existing session and creates a new entry in the session table if needed.
    a. The header is inspected to identify match criteria for security and SD-WAN policy processing.
        i. Source and destination IP addresses
        ii. Source and destination zones
        iii. Protocol and application ports
    b. L3 route lookup is performed to identify the egress interface.
    c. New sessions are processed in slow path. Packets matching existing sessions are offloaded and processed in fast path.
3. Processing for application and user identity begins.
    a. The App Cache is checked to see if the application already exists. If found, the App-ID from cache is used.
    b. If not, the App-ID engine uses one or multiple packets to decode and identify the application. Identified applications are stored in the App Cache to speed up future lookups. Service ports can be used for single packet application identification if needed.
    c. User-ID identifies the application owner.
    d. Using PAN-OS's single pass architecture, security policies are applied to scan for malware and vulnerabilities. URL filtering and decryption is performed if configured.
    e. The traffic is permitted or denied. If traffic is blocked/dropped by the security policies, an end log is created and the traffic flow stops.
4. For permitted traffic, the egress interface from step 2 is referenced and a determination is made if the traffic is destined for an SD-WAN virtual interface (VIF).
    a. Direct Internet Access (DIA) traffic will go out of a DIA SD-WAN VIF servicing the default route.
    b. If DIA AnyPath is configured, DIA traffic can failover to an SD-WAN tunnel VIF if needed.
    c. Internal traffic will go out of an SD-WAN tunnel VIF based on the route lookup.
5. The application is matched against the SD-WAN policy list. For matched application traffic:
    a. The path quality profile thresholds are compared against the egress VIF's health measurements and all qualifying paths (physical and virtual) are identified.
    b. The traffic distribution profile is used to select the egress path based on priority defined by the traffic distribution method.
    c. VPN tunnel selection for DIA AnyPath failover consults the interface's VPN failover metric for further path selection priority.
    d. Lost packet error correction is applied if FEC or Packet Duplication profiles are configured.
    e. Applications that are not matched will be enforced by the default implied SD-WAN policy at the end of the policy list that will select the best available path from the egress VIF with no path failover.
6. Post SD-WAN processing is applied and traffic is forwarded through the selected egress interface.
    a. If a QoS policy is configured on the egress interface, traffic shaping/policing is applied based on the bandwidth and priority assigned to each QoS class for the matched application.
    b. Traffic sent between PAN-OS SD-WAN enabled firewalls have symmetric return automatically enabled on SD-WAN interfaces.

TIP: If packets from the same flow are sent over multiple paths with varying degrees of available bandwidth, latency, jitter, packet loss, hop count, etc, the risk of increasing the number of out-of-order packets increases and this can affect overall application performance and consume firewall resources unnecessarily. PAN-OS prevents this type of behavior by automatically enabling symmetric return on SD-WAN interfaces. Symmetric return ensures that S2C traffic is sent through the receiving interface and helps reduce the risk of out-of-order packets.

SD-WAN Policy Overview

The SD-WAN policy and profiles are used to control how applications are matched and processed through the SD-WAN engine to select the best egress path. Like other PAN-OS policies, SD-WAN allows context to be matched on source, destination, and application/service.

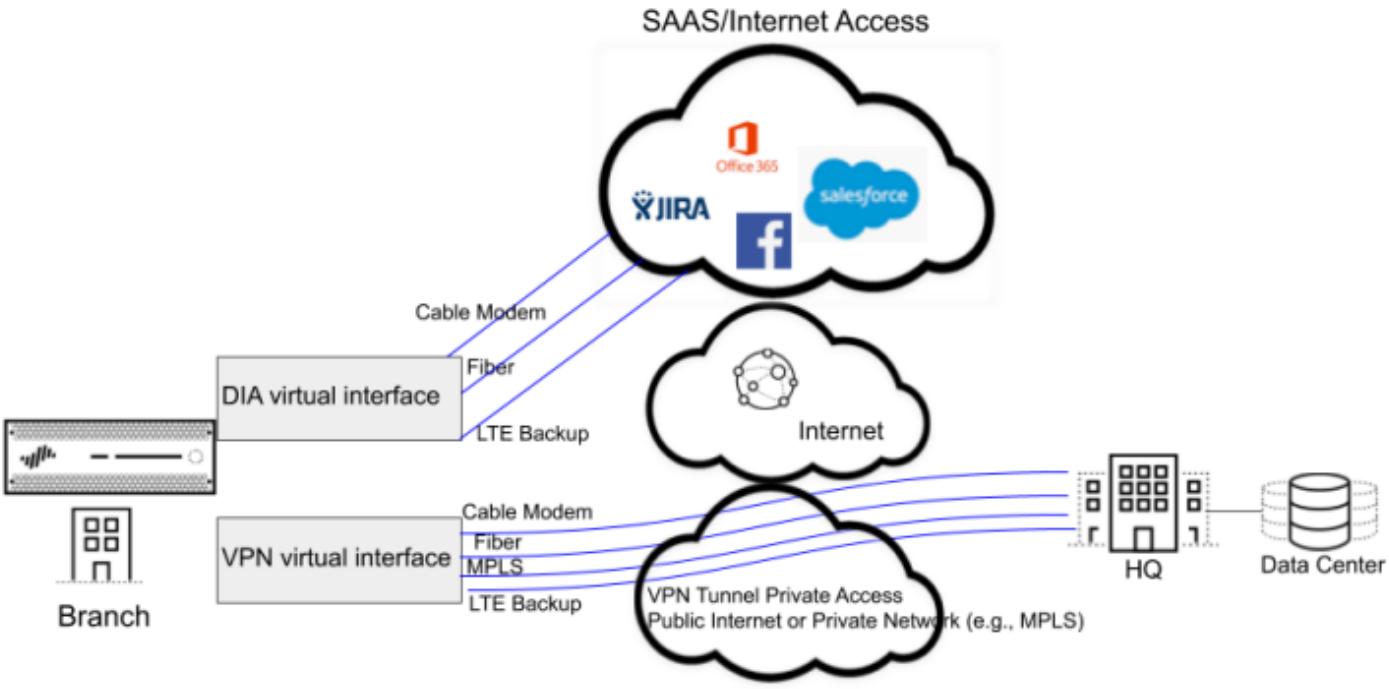
The relationship and workflow order of how SD-WAN profiles are applied by the SD-WAN policy is shown in the following illustration. Although the SD-WAN Interface Profile is not directly applied in the SD-WAN policy, it does provide information to the SD-WAN decision tree for path selection. Reference these concepts when going through the remaining sections to fully understand the workflow of how packets are processed through each component.



SD-WAN Virtual Interface (VIF) Overview

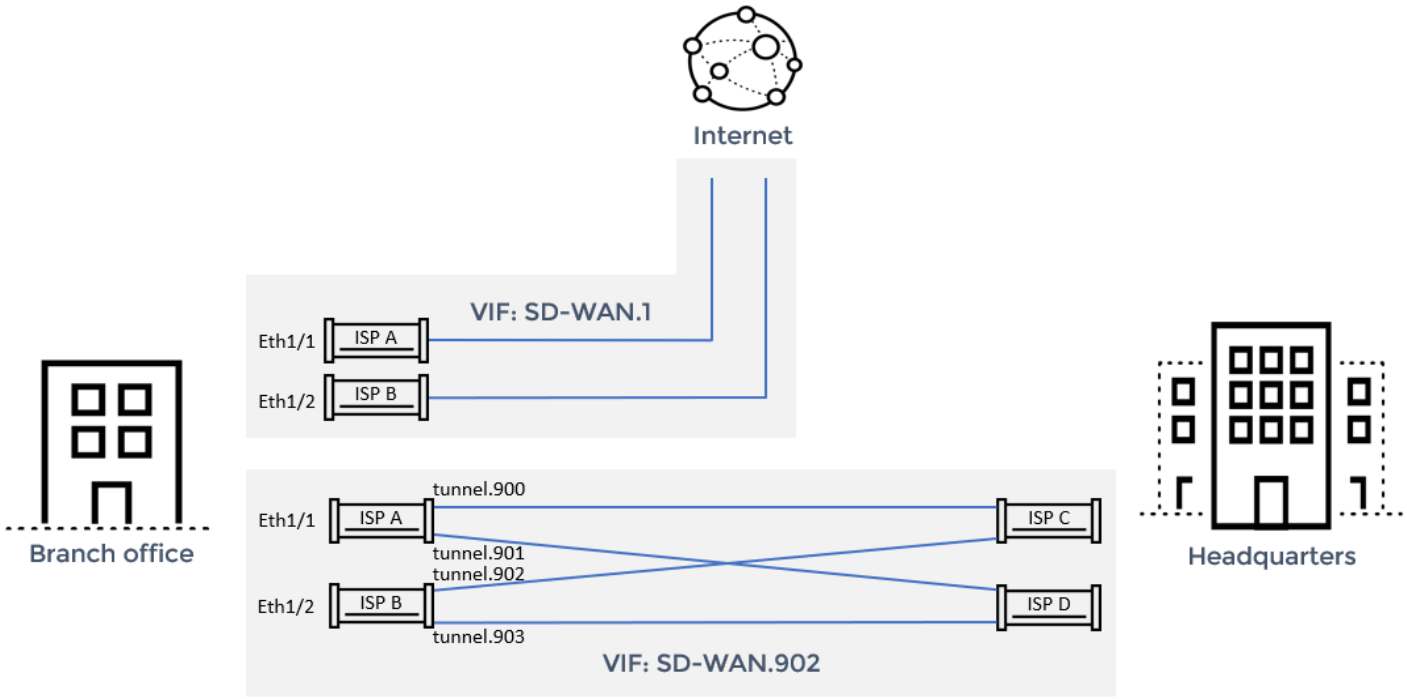
PAN-OS SD-WAN uses virtual interfaces (VIFs) to send and receive SD-WAN traffic. The VIF interfaces are automatically created by Panorama's SD-WAN plugin for each firewall based on the VPN cluster topology and its devices. In general, a dedicated VIF is created between the originating firewall and all of its peers - other hub or branch locations. A DIA VIF is automatically created and associated with the router's default route to handle non-local traffic to the internet. VIF members can be DIA interfaces and VPN tunnels.

As shown in the illustration below, the DIA VIF is created to support internet access and each ISP circuit that is defined as an SD-WAN link will automatically be a member. Similarly, the VPN tunnel VIF is created to support connectivity between the branch and the DC hub located at HQ. The SD-WAN interfaces are automatically added as VIF members and full mesh "MxN" VPN topology is automatically created between the two locations.



The number of tunnels created in the full mesh VPN topology will depend on the number of SD-WAN interfaces (ISP circuits) configured on the two firewalls. For example, a branch office has two ISP circuits and the DC hub it is connecting to also has two ISP circuits. In a full mesh M x N topology between two sites, a total of four tunnels are created and added to the tunnel VIF.

WARNING: The term "full mesh" between two sites is not the same as a "full mesh branch-to-branch" SD-WAN topology where every branch location is directly connected to every other branch in the VPN cluster.



Panorama's SD-WAN plugin automatically creates the necessary VIFs on the branch office firewall for DIA and VPN tunnel connections.

1. SD-WAN.1 (DIA VIF)
2. SD-WAN.902 (tunnel VIF to HQ)

In the VPN tunnel VIF, four tunnels are added to create the M x N full mesh between the branch and HQ firewall and are labeled tunnel.900 - tunnel.903 on the branch firewall. The auto provisioned VIFs and tunnels can be reviewed in the branch firewall's network interface SD-WAN tab and IPSec Tunnels respectively.

Path Quality and Traffic Selection Components

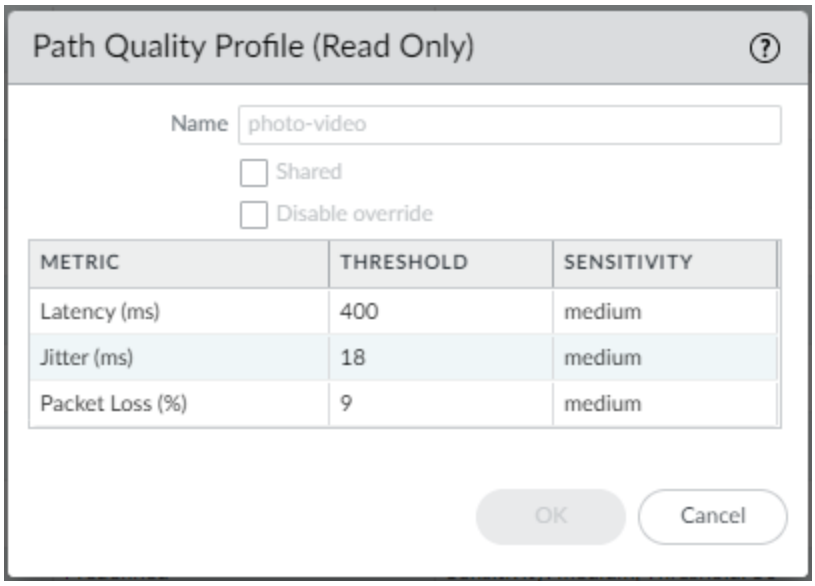
SD-WAN application failover is controlled by multiple path monitoring components as well as path selection criteria, and these technologies work together to provide one or more paths for application traffic to egress the firewall.

Path Quality Profile

The Path quality profile (PQP) is a key component of the PAN-OS SD-WAN feature and it is used in an SD-WAN policy to instruct the firewall to look for a better path when its thresholds for latency, jitter, or packet loss have been exceeded. To ensure that each of your business-critical and latency-sensitive applications have the best user experience, create and apply a unique PQP profile for each application or group of applications that have similar behavior.

The firewall is constantly measuring the health of its SD-WAN links, VPN tunnels to the hub or another branch, and to SaaS

locations on the internet and compares these values to the thresholds defined in the PQP. If all health measurements are under the PQP thresholds, the path is deemed "qualified". If any of the health measurements exceed one or more PQP thresholds, it is deemed "disqualified". PQP uses three metrics for its thresholds and each metric is measured independently.



Latency - The round-trip delay time for a data packet sent across an SD-WAN tunnel (measured in ms). By default, the SD-WAN healthcheck calculates latency every 200ms and uses a sliding window mechanism which includes the last three measurements in its threshold comparison. Taking the average of the sliding window's three measurements prevents "false-positive" path failovers when there is a very short spike in one of the health measurements.

The following example describes the sliding window latency concept and uses a PQP latency threshold of 100 ms for illustration purposes.

Time Slot	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Latency Measured	60	130	75	78	110	115	98	72	82	76	133	80	72	68
Window Average	N/A	N/A	88.3	94.3	87.7	101.0	107.7	95.0	84.0	76.7	97.0	96.3	95.0	73.3

1) The firewall doesn't initiate a path failover in time slots 2 and 11 even though the latency measured was over the PQP's 100 ms latency threshold because the window average (last three measurements) was not over the 100 ms threshold.
2) In time slots 5 and 6, the threshold was exceeded and caused the window average to go above the 100 ms threshold in time slots 6 and 7 and this triggered the firewall to perform a path failover.

Jitter - The variance in the delay compared to the average delay for a data packet sent across a SD-WAN tunnel (measured in ms). By default, the SD-WAN healthcheck calculates jitter every 200ms and also uses a sliding window mechanism like the latency measurement for the last three measurements.

Packet Loss - The percentage of packets lost compared to total packets sent across a SD-WAN tunnel. By default,

the SD-WAN healthcheck calculates packet loss every 200ms and also uses a sliding window mechanism that takes the last 100 measurements into its calculations. Unlike latency and jitter measurements, packet loss takes more packets into consideration to obtain a "normalized" reading for the initial measurement. Using the default setting, the firewall will take 20 seconds to obtain the first packet loss reading, then a sliding window is applied to support the rapid detection and path failover.

The SD-WAN path-monitoring component can measure the health of the underlay network, overlay network, and application traffic for DIA applications. The firewall will also monitor the ISP's next hop gateway address to determine the operational status of the ISP circuit. To accurately measure health on each of the different network types, path monitoring can use either active or passive health check mechanisms. With active monitoring, the firewall sends ICMP probes to a specific destination to calculate the PQP metrics for a path.

For example, the SD-WAN VPN tunnels use active probes sent between the firewall's IPSec interfaces to measure latency, jitter, and packet loss. Direct Internet Access (DIA) paths to specific SaaS applications can also use an active probe to a specific public IP address to measure path health. But DIA healthcheck can also take passive measurements by using the conversation between the client and the server to determine the path health.

Health measurements are used in conjunction with the PQP to provide the following:

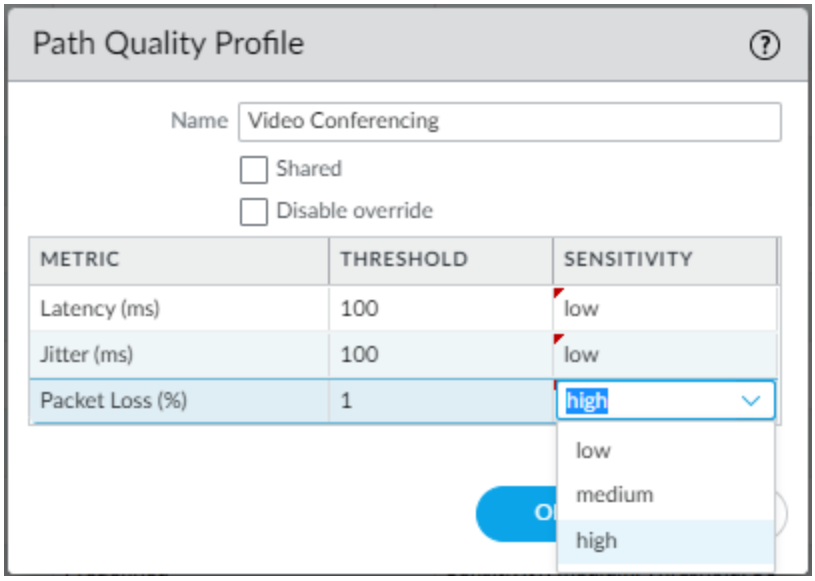
1. Probes provide the PQP with real time health measurements and allow the PQP to determine if the path is qualified or disqualified to carry the application. The threshold set for each metric is the maximum upper limit for that metric. If the health measurements are all below the PQP's metric thresholds, the path is qualified and operating within acceptable limits. If the health measurements exceed any of the PQP's metric thresholds, the path is degraded and is not qualified and this triggers the firewall to look for a better path.

   TIP: Health probes initiated by the firewall use the egress interface's IP address as the source address. The firewall automatically permits its health probes without the need to configure security policies to permit the probe traffic.

2. The PQP profile can also determine which path to use when more than one path is qualified. When this occurs, the PQP's sensitivity setting can be used to further tune how the best path is selected and there are three sensitivity settings to rank the prioritization of a PQP metric - low, medium, or high.

   If you configure a metric to have a higher sensitivity than the other metrics, the path selection algorithm considers that metric first. If all metrics have the same value, then the path selection algorithm first considers packet loss, followed by latency, and then jitter. In the example below, the video conferencing application is more sensitive to packet loss, so a "high" sensitivity setting is applied to that metric. When multiple paths qualify in the egress VIF, the packet loss for each path is examined and the path with the least loss will be selected, even if that path has higher latency than the other paths.



3. PAN-OS SD-WAN uses a ”one-side” probe technology that originates from the branch firewall and the probe can accurately measure branch-to-hub and hub-to-branch health. By eliminating the need to originate health checks from the hub to other locations, the number of probes used between the SD-WAN sites is greatly reduced.

Default vs Custom Path Quality Profile Settings

As a best practice, leveraging PAN-OS's predefined path quality profiles can help administrators get started quickly without having to figure out the correct latency, jitter, and packet loss thresholds for each application they need to maximize usability for.

Palo Alto Networks created the SD-WAN predefined path quality profiles by testing multiple applications from each application category in a controlled environment. The latency, jitter, and packet loss conditions were adjusted independently to see where the application's user experience dropped below acceptable levels and the results were compiled with other applications from the same category to determine an acceptable starting point.

If needed, the predefined category thresholds can be tuned afterwards when more application testing has been performed in the environment. Path failover can be accelerated or postponed to achieve the desired user experience. Simply make a copy of the predefined profile and adjust the latency, jitter, and packet loss values as needed. In general, raising the values delays SD-WAN path failover and lowering them accelerates path failover.

If you don't know which category an application falls under, use either the Palo Alto Networks Applipedia service or the PAN-OS Objects > Applications to identify the application and its category, and then use the corresponding predefined path quality profile in the SD-WAN policy.

TIP: If needed, start off with higher PQP values and test the application's tolerance to the thresholds. If these values are set too low, the application may flip back and forth between multiple paths too often. Also remember that many modern cloud applications are tolerant to latency and jitter and can recover lost packets without greatly impacting the user's experience. Only use aggressive thresholds for applications that really need them.

A lab environment can be used to experiment with latency, jitter, and packet loss percentage to identify when an application's behavior becomes unacceptable. There are open source tools that can be used to introduce latency, jitter, and packet loss into the internet connection, and these can be valuable for testing application tolerances. For example, linux's tc command can be used as shown in this example:

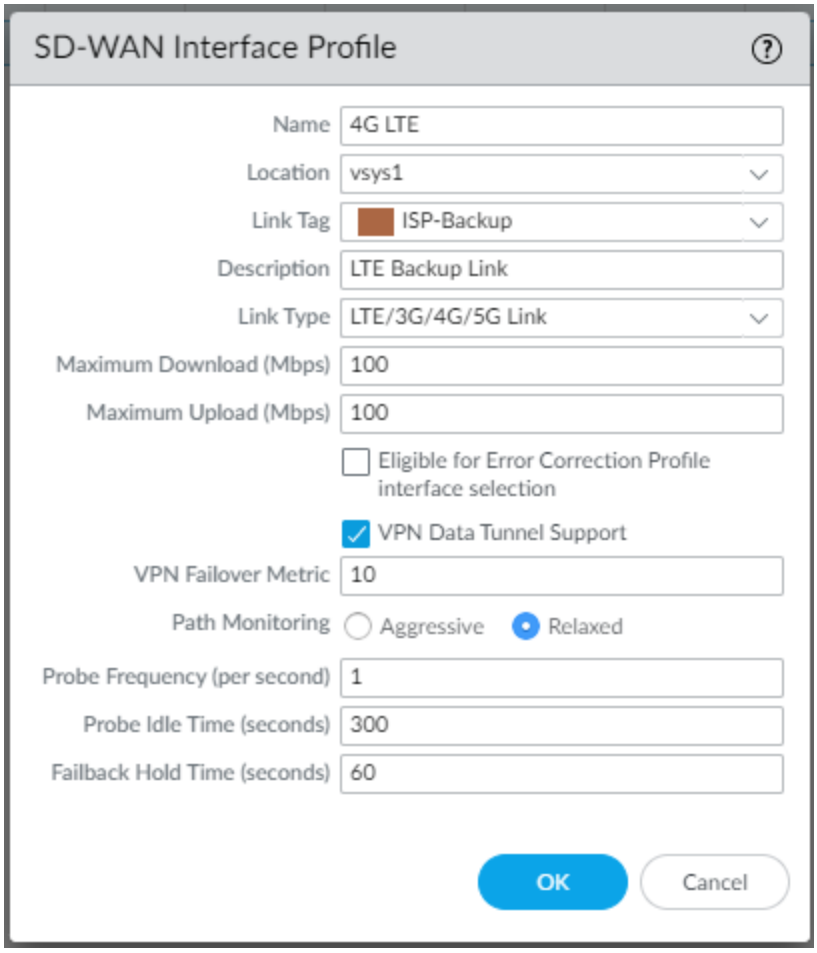
```
jitter-test.sh
#!/bin/bash
# set delay 100ms delay and 10ms jitter
sudo tc qdisc add dev eth0 root netem delay 100ms 10ms
sleep 90
sudo tc qdisc del dev eth0 root netem

#set 40% packet losudo ss
sudo tc qdisc add dev eth0 root netem loss 40%
sleep 90
sudo tc qdisc del dev eth0 root netem
```

Default vs Custom Health Check Settings

PAN-OS SD-WAN's health check default values are designed to provide rapid detection of path degradation and provide sub-second failover to a better path. The frequent health checks use more bandwidth and are a trade off for faster detection. These settings are good if you have high speed internet circuits, but may not be suitable for low bandwidth or expensive link types - such as satellite, remote DSL, and LTE connections.

PAN-OS SD-WAN allows you to change the default path monitoring behavior between an "aggressive" or "relaxed" mode and each mode has configurable options to control probe frequency.



Aggressive Mode - Sends a health probe at every probe frequency interval. The health measurement is performed consistently over time with no idle time permitted. Aggressive mode is used for connections that have ample bandwidth, constant traffic, where cost is not a factor, and when fast identification of link degradation is needed.

Relaxed Mode - Sends a series of health probes for seven seconds at the top of the probe frequency interval and then stops for the probe idle time. The probe frequency controls how many probes are sent per second. Relaxed mode is generally used for expensive links and/or backup links that don't normally have consistent traffic loads and immediate identification of link degradation is not required.

With either mode, the following parameters can be tuned to achieve the right balance of path degradation detection, bandwidth use, and failover speed.

Probe Frequency - The number of probes to send per second. The default is five and this provides a sub-second detection and failover frequency that is ideal for real time applications. Decrease this value to extend the detection and failover duration and to lower bandwidth consumption. The lowest value is one second and the firewall's sliding window measurement will take the last three probes into account to determine the metric's health value.

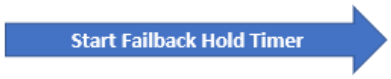
Probe Idle Time - The amount of time the firewall waits in Relaxed Mode before sending out the next round of

probes. A series of seven probes are sent at the beginning of the cycle and then the idle time is respected and no probes are sent. For LTE backup links that can be expensive to operate, lower the probe frequency to extend the probing duration and add a long probe idle time. For example, the probe frequency can be set to 1 second and the idle time can be set to 300 seconds to delay the next set of probes.

Failback Hold Time - The amount of time the firewall waits to ensure the path is qualified (healthy) before making it a preferred link for new traffic. The default is 120 seconds and this is the longest duration. In busy networks, you may want to have recovered links reinstated faster to help share the load. Have a look at the link's historical trends to see if it has high failure rates or not. If the link has a low failure rate, lowering this value should be fine.

Using the latency sliding window example, when the window average exceeds the PQP's threshold the firewall performs a path failover action and the failback hold timer is started for the impacted path. In order for the firewall to reinstate the impacted path, the path cannot encounter another sliding window measurement that exceeds the PQP's threshold within the timer duration.

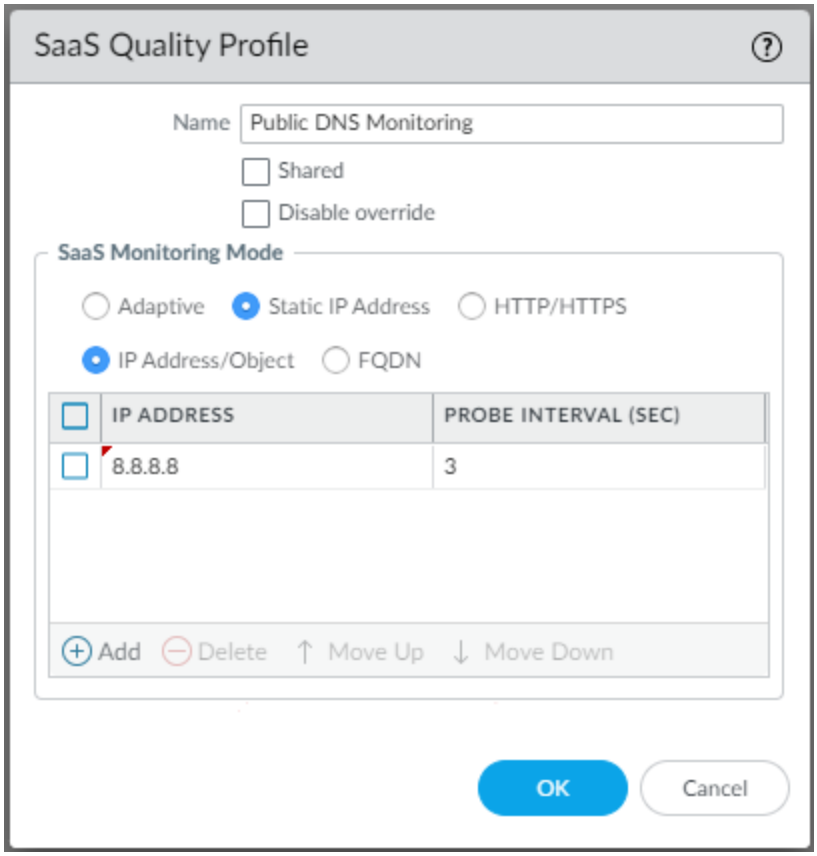
Time Slot	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Latency Measured	60	130	75	78	110	115	98	72	82	76	133	80	72	68
Window Average	N/A	N/A	88.3	94.3	87.7	101.0	107.7	95.0	84.0	76.7	97.0	96.3	95.0	73.3



TIP: When changing the probe frequency defaults, thought as to how packet loss is measured should be taken into account. With the aggressive mode's default of five probes per second, packet loss can be measured faster as more probes are sent. If the frequency is changed to a longer period (less probes), the PQP's packet loss threshold should be adjusted to match the ratio of the probe frequency. For example, if an application starts to fail at 5% packet loss with the default five per second (200ms) probe frequency, changing the frequency to once per second may need the PQP's packet loss percentage to be reduced to 1%. The packet loss measurement uses the health probes to detect lost packet loss.

SaaS Quality Profile

The SaaS Quality Profile (SQP) is used to measure the path health of Direct Internet Access (DIA) applications and can use either an active or passive approach. The active method is a one-sided health measurement that sends a probe from the firewall and it measures bi-directional health to a static IP address or a monitored URL. The passive method doesn't send any probes. SaaS applications will not have a destination side PAN-OS firewall frontending the traffic, so the SD-WAN VPN tunnel health probes are not used with these monitoring types.



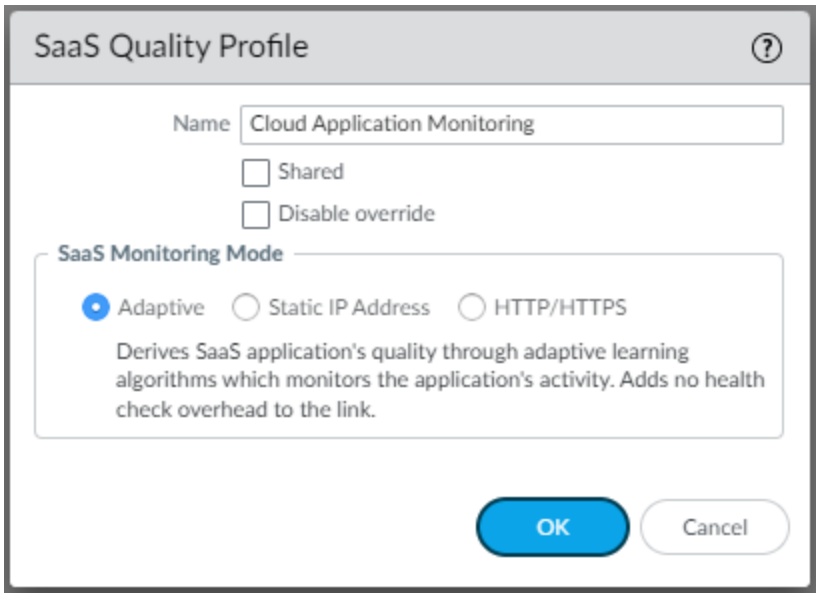
For the active monitoring mode, the system supports up to a maximum of four monitored IPv4 addresses and they can be an IP address or an FQDN. The probe interval can be between 1 and 1000 seconds and a waterfall approach is used to monitor the path if multiple IP addresses are configured. The firewall will use the first IP address on the list to monitor the path and if that IP is unreachable, the second IP on the list is used, and so on. When all IP addresses are non-responsive, the path is deemed "down". As long as the monitored IP address is reachable, it will be used by the SQP and the firewall uses the egress interface's IP address as the source address.

For the HTTP/HTTPS monitoring mode, a URL is specified for the firewall to monitor and the first IP address that is resolved by DNS is used as the monitored IP address. As DNS resolutions can return a different IP address with each request, multiple IP addresses can be used to monitor the URL and the closest IP address to the application may not be the one being monitored. When HTTP/HTTPS is enabled, the firewall initiates the necessary handshakes to establish a connection with the destination host and this can use more resources than the ICMP ping used in the active IP address/FQDN monitoring.

TIP: If an HTTPS URL is used for monitoring, be aware that an SSL handshake is initiated by the firewall. If the probe interval is set too low, it can create a lot of SSL handshake traffic to the destination IP address and this can also use more firewall CPU resources. A best practice is to use a longer probe interval with both HTTP and HTTPS monitoring methods.

The advantage of using active probes is higher accuracy as the firewall is proactively sending health probes and not relying on traffic generated by the two endpoints. The downside is additional traffic being generated by the firewall which can result in more bandwidth being used. Another issue can be the destination that is being monitored. It may not like the frequent probing and a security policy can be created to block the probes or rate limit them. The active method is best when you need accurate and fast health statistics to detect path degradation and the bandwidth used by the probes is not an issue.

In addition, the SQP can also measure path health using a passive approach called "Adaptive" monitoring. Adaptive monitoring doesn't send active probes and it uses the TCP traffic sent and received between the source device and the application server to determine the path's health. The firewall can measure both client to server and server to client activity to determine latency, jitter, and packet loss in both directions.



Adaptive monitoring is best when rapid detection of path degradation is not the top priority, but reducing the bandwidth on the path is. In general, the more traffic there is between the devices the higher the accuracy is for the health measurements to the monitored application. Typically, there will be a delay in the path measurement's accuracy at the beginning of the session as the firewall starts analyzing the new TCP flow. Once enough traffic is analyzed, the latency, jitter, and packet loss is accurately reflected. Another situation that can affect the adaptive monitoring's accuracy is a pause or lag between transmissions. During these periods of inactivity, the health measurements may delay failover of the initial sessions.

As SaaS path monitoring is designed to monitor the path to DIA applications, the SaaS quality profile applies monitoring on the SD-WAN interface that is servicing the default route. For both active and passive monitoring, the SQP's measurements are used in the SD-WAN policy to determine when a better path should be selected for the matched application. After a High Availability (HA) failover, the new active firewall will begin the SaaS path monitoring from a new starting point. No monitoring state is synchronized between the HA firewalls.

TIP: All applications in a single policy are treated as one application group and an application can be included in multiple SQPs. Care must be taken to ensure that the application is not placed into too many SQPs as this can cause unexpected path failover results. In addition, SQP monitored applications are not offloaded.

Error Correction Profile

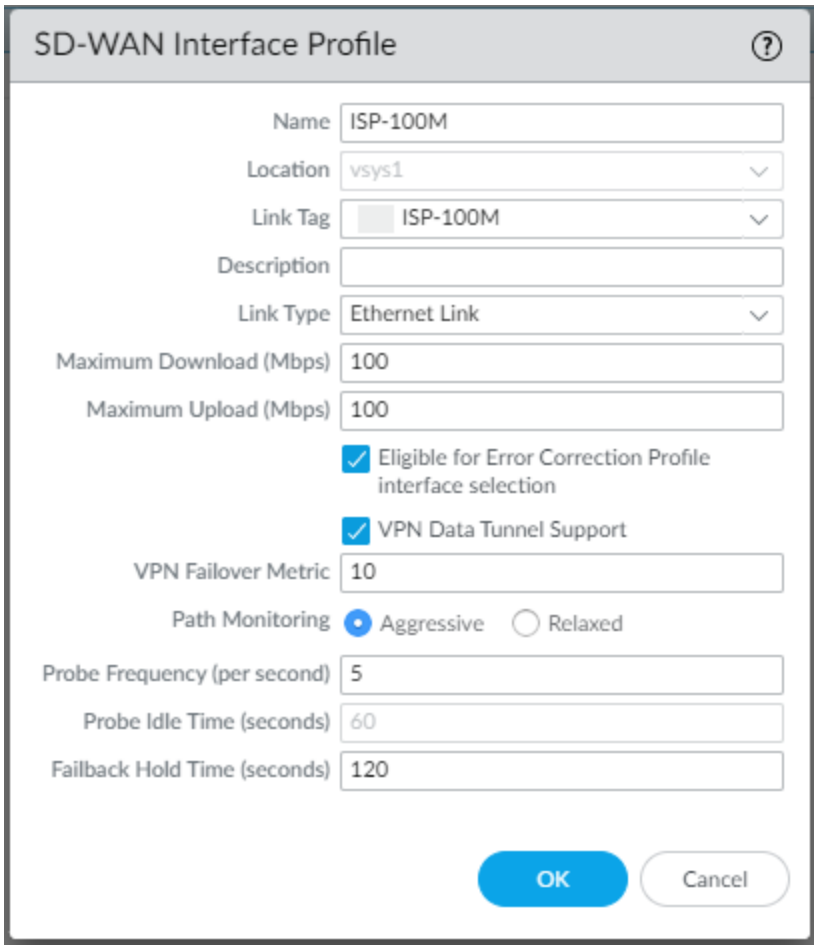
PAN-OS SD-WAN has two different mechanisms for recovering lost packets - Forward Error Correction (FEC) and Packet Duplication (PD). Although these two recovery mechanisms are not directly tied to path health checks and path selection, they can indirectly influence path selection by introducing additional traffic on SD-WAN paths. For example, if the FEC or PD traffic is added to an already congested link, it can cause the path quality profile thresholds to be exceeded and in turn, trigger a path selection event.

FEC uses parity packets to allow the receiving firewall to reconstruct lost packets and you can configure the ratio of parity packets to data packets to either increase or decrease the chances of recovery. The higher the ratio, the more bandwidth is consumed as more parity packets are generated and sent to the receiving firewall, but a better chance of recovery will also be achieved.

With PD, a separate path is selected to duplicate the entire traffic flow so two copies of the application traffic are sent to the receiving firewall. The receiver will take the first "healthy packet" and discard the other. Although this packet recovery mechanism provides the best protection against packet loss, it is also the most expensive in terms of bandwidth use.

TIP: Turning off packet duplication on high cost LTE links or low bandwidth links will make the network more efficient and lower operating costs. Error correction is a resource-intensive feature and not all applications will benefit from FEC or PD. Before enabling on a production network, try to test the application's response to these error correction methods beforehand. In some cases, it is more beneficial to use QoS overlays instead of enabling FEC or PD.

Any interface can be included or excluded from being used as an alternative path for PD and you can configure this with the "Eligible for Error Correction Profile interface selection" checkbox in the SD-WAN interface profile.

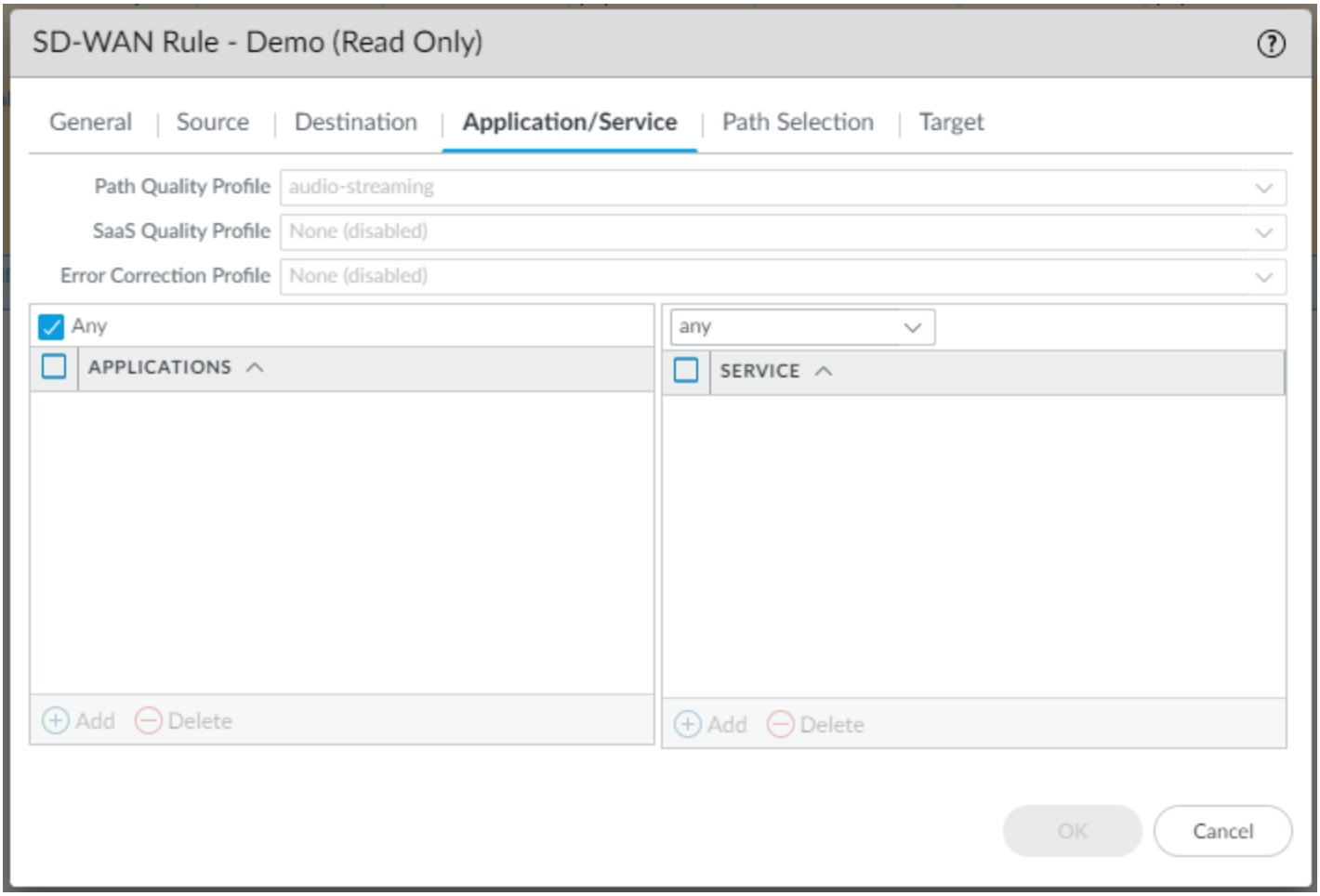


FEC and PD require compatible PAN-OS SD-WAN firewalls at both ends of the connection and are only supported on SD-WAN VPN tunnels. The firewall adds recovery information to the packets before sending them over the VPN tunnel to the receiving firewall, which uses this information to determine which packets were lost and to perform repairs. The receiving firewall must remove the recovery information from the header before forwarding the packet to the destination host.

For more information on how forward error correction and packet duplication features are configured, please refer to the SD-WAN administrator guide.

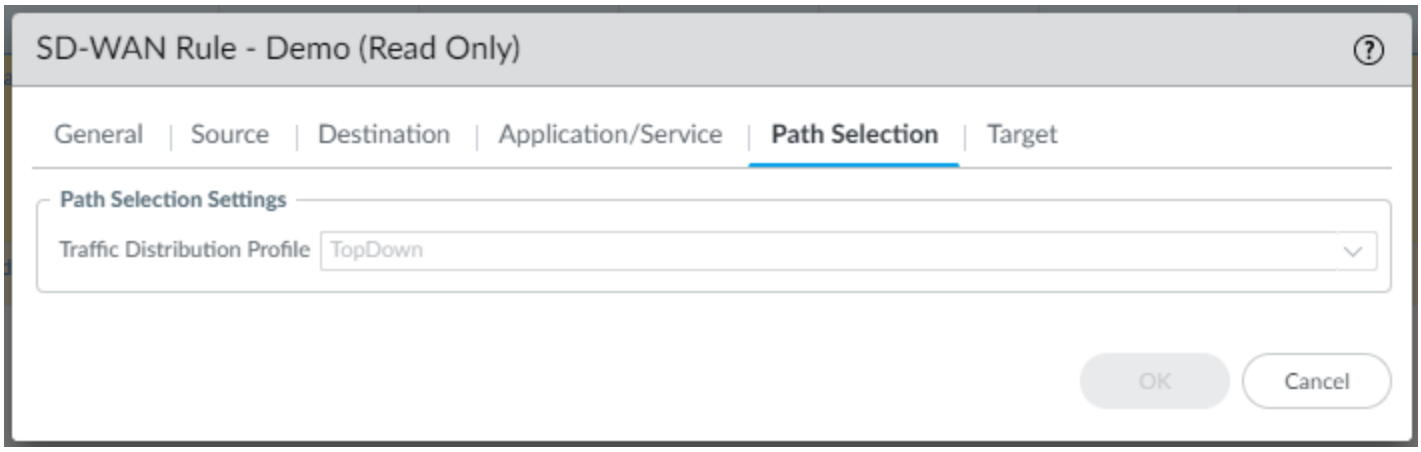
Traffic Distribution Profile

SD-WAN profiles operate in conjunction with the Traffic Distribution Profile to select the best egress path. The profiles are configured in the SD-WAN policy to provide context matching and different profiles can be used with different application types to achieve the highest user experience.



In addition to matching on one or more applications, the SD-WAN policy can also match on source/destination zone and IP address, a specific user or user group to provide granular SD-WAN policy control. Once a match is made, the path quality, SaaS quality, and error correction profiles assigned to the SD-WAN policy are applied to the traffic and a final path selection decision is made after comparing the VIF's qualified link health measurements.

If the SD-WAN logic determines that the existing path is not sufficient to meet the PQP thresholds, the assigned traffic distribution profile is used to determine the next "best path".

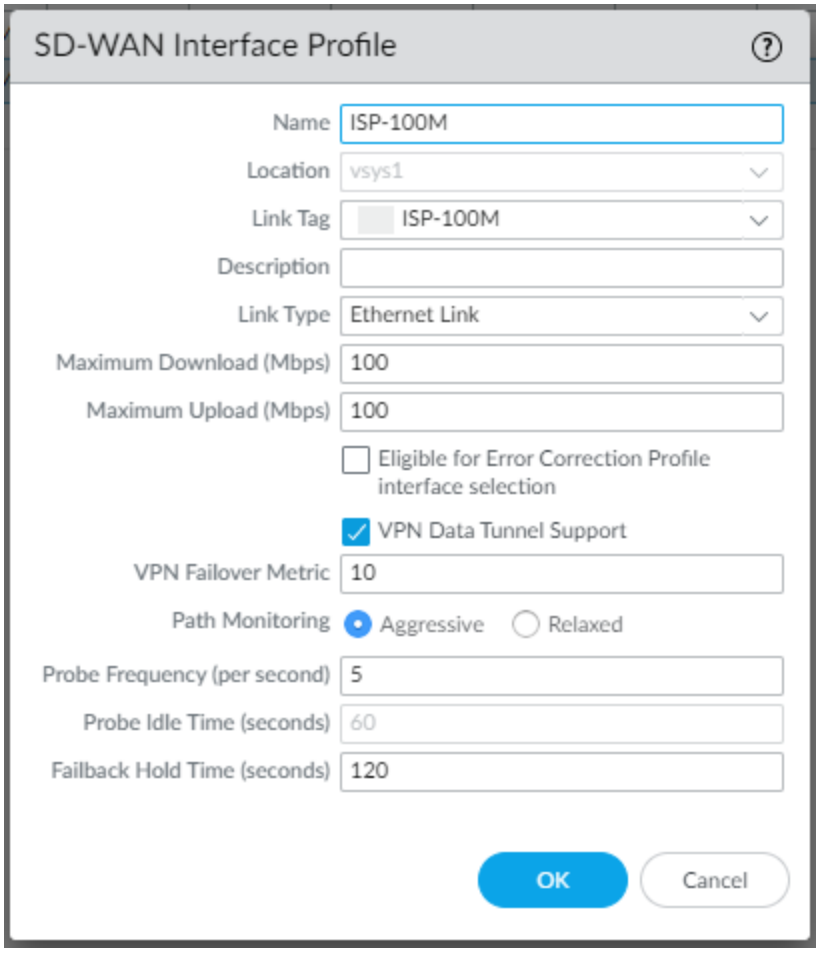


The traffic distribution profile provides three traffic distribution methods to control how the best path is selected. Link Tags defined in the SD-WAN interface profile bind the physical SD-WAN interfaces to the traffic distribution profile. An interface can have only one link tag assigned, but the link tag can be assigned to multiple SD-WAN interfaces. Link tags are created in PAN-OS's Object/Tags function and you can give the tag a descriptive name and color code it for easier identification.

When the same link tag is assigned to more than one ISP circuit, the firewall creates a Link Bundle and aggregates the bandwidth from each ISP circuit into a larger virtual interface. This allows multiple low bandwidth circuits to be combined to create a "fat pipe" that has more bandwidth for your applications and the firewall will automatically distribute sessions between the link bundle members using round robin order. For more information on link bundles, please see the "Other Path Selection Considerations".

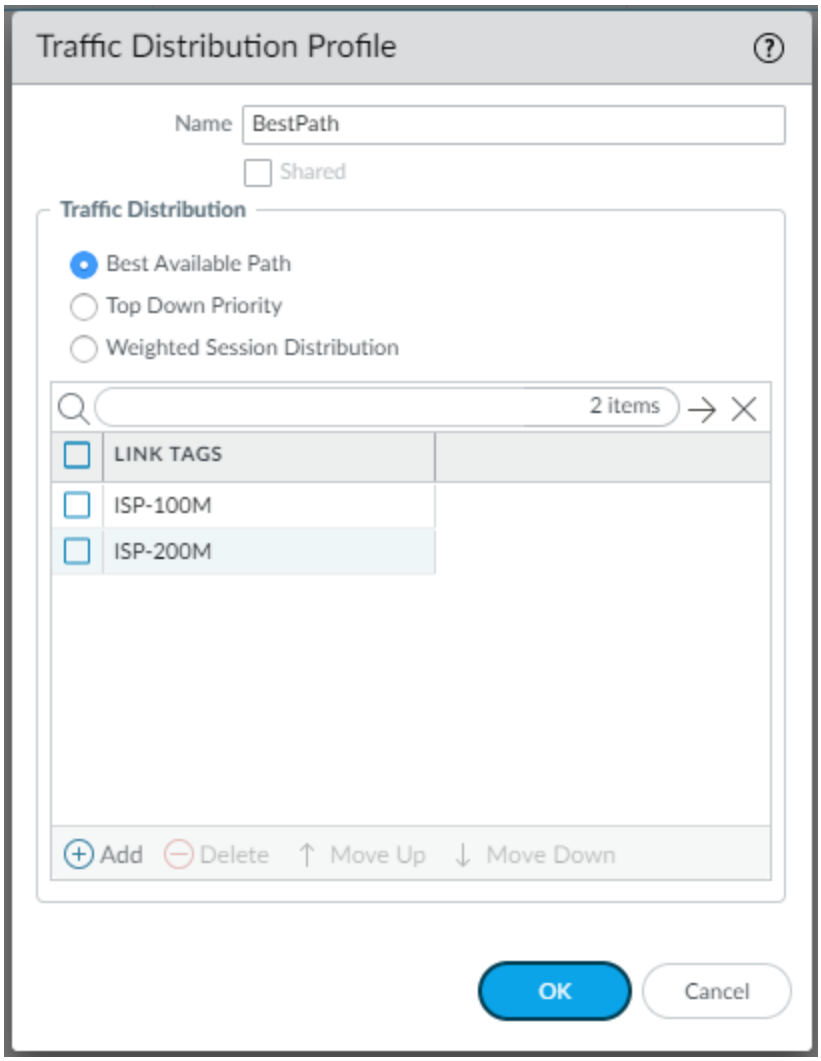
TIP: Best practice is to combine ISP circuits that have similar characteristics into a link bundle. For example, combining multiple broadband services with similar speeds to increase bandwidth and support round robin session loading will improve capacity for an egress interface. Link types that differ significantly should not be combined as it can cause excessive application failover. For example, combining a satellite link together with a high speed fiber circuit in a link bundle would not be a good idea as the low bandwidth, high latency nature of the satellite link doesn't support round robin session distribution well and this would create unnecessary path selection activity and overhead on the firewall.

In the example below, a link tag "ISP-100M" is used in the SD-WAN Interface Profile which is applied to the ISP circuit providing 100 Mbps capacity.



Three Traffic Distribution Methods

To provide flexibility in how a new path is selected, the traffic distribution profile supports the following three distribution methods.



Best Available Path - This method will select the best path from all link tags defined in the traffic distribution profile and the order of the link tags are not taken into consideration. This is the default path selection method. Path selection uses the PQP exclusively to choose the best path from all link tag interfaces by comparing the health measurements of each path to the PQP latency, jitter, and packet loss thresholds. Any path that exceeds the thresholds is excluded from the selection process and is disqualified.

If multiple paths qualify, the path selection algorithm uses the sensitivity value assigned to the metrics to prioritize which metric is used first in the selection process. If the sensitivity values are identically set, the firewall will use the healthiest path with the least packet loss, latency, and then jitter.

This distribution method also prevents excessive session "flip flopping" between paths and can help in situations where ISP circuits become sporadic. The firewall will attempt to anchor new sessions to the path that has the best quality, and once anchored, the session remains on this path. When congestion arises on the link, the firewall will distribute new sessions to a

healthier path in an attempt to offload the congested link and allow the link to recover and in turn, improve the application performance.

However, if the link further degrades to a "brown out" condition (exhibiting poor performance for more than one minute and moving new sessions to another path did not help), existing sessions on the impacted link are gracefully moved to better performing paths. The firewall will not move every session on the impacted link immediately as that may cause an oversubscription on the healthier links and cause additional session redistribution activity that may further degrade user experience. The graceful migration period can last up to one minute and up to 1000 sessions can be migrated to offload the impacted link. In a complete outage, all sessions will be migrated off the impacted link.

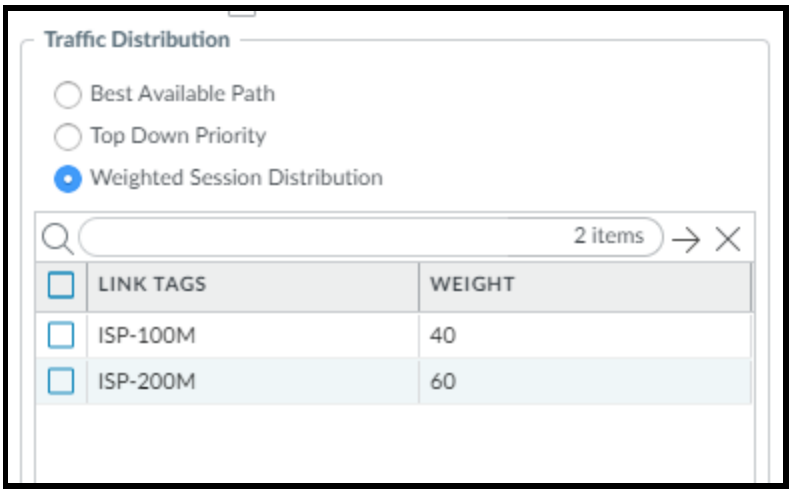
This distribution method is best suited for deployments where link cost is not a factor and you want the application to always egress out of the best path. For example, a branch is located in a high tech metropolitan area where cheap fiber and broadband services are readily available and cost is not a factor. The network team's goal is to have the best application experience possible at all times for the company's mission critical applications.

Top-down Priority - This method will select the best path using the tag order defined in the traffic distribution profile, and the link tag at the top of the list has the highest priority and is used first. The PQP thresholds are compared against the health measurements of each path in the tag and all qualifying paths are used for the path selection. As a path becomes disqualified, the path selection algorithm will look to the remaining paths within the link tag and choose the next path to send new sessions to. If no paths in the link tag are qualified, the next tag on the list is used. If there is a qualified path in the second link tag, it is selected.

If multiple paths qualify, the path selection algorithm uses the sensitivity value assigned to the metrics to prioritize which metric is used first in the selection process. If the sensitivity values are identical, the firewall will use the healthiest path with the least packet loss, then latency, and then jitter. This top down waterfall selection method will continue until the last link tag is used. If all paths are disqualified by not meeting the PQP thresholds, the path selection will choose the least impacted path - changing to the Best Available Path default selection method.

This distribution method is best suited for deployments when there are significant cost differences between paths or when there is a low priority link that typically has high failure rates. Link tags associated with the high cost/low reliability circuits are placed at the bottom of the link tag list to lower the selection priority and cost effective/reliable circuits are placed at the top of the link tag list for highest priority. For example, a branch is located in a rural area where there are limited ISP choices and it can only obtain DSL circuits from a single ISP provider. The decision is made to implement two DSL circuits in a link bundle to act as the primary path and deploy a more expensive 5G LTE circuit as a backup link. The 5G LTE circuit is only used if the DSL links are disqualified.

Weighted Session Distribution - This method distributes sessions across each of the link tags based on the percent weight assigned. The total percentage allocated across all link tags must add up to 100%. Weighted session distribution path selection is static and provides a "manual" way to distribute traffic to the link tags and there is no link failover if the current path encounters high latency, jitter, or packet loss. The firewall uses the new session's first packet to determine the distribution path and all subsequent packets in the session are forwarded to the same path. When a link tag reaches its percentage limit, new sessions are distributed to the other tags that are under their percentage limit.



This distribution method is best suited for deployments where there may be many sessions belonging to a non-critical application that can use a significant portion of the link's capacity. For example, an application transfers hundreds of files between locations and uses a new session for each file transferred. The files being transferred are not mission critical and delays are acceptable as they're transferred during off peak business hours. Furthermore, the application uses TCP and other file transfer repair mechanisms to resend packets if they are lost. A traffic distribution profile is created to spread 40% of these sessions onto the first link tag (with 100 Mbps) and the remaining 60% of the sessions are directed onto the larger ISP path (with 200 Mbps). This allows the administrator to predictably control and load the ISP circuits.

TIP: When you are going through the SD-WAN planning stage and identifying the different ISP circuit types that each branch will support, think about the order the applications should take between the different circuits. Traffic and application planning should also be identifying the most critical business applications and what their sensitivity is to bad quality links. Don't try and identify every single application that's on your network. Start with the critical business applications that must be protected to ensure a high-quality user experience first. This will simplify your SD-WAN and QoS policy design and reduce the implementation time.

Path Failover and Path Recovery

For the best available path and top down priority methods, if there are no qualified paths in the entire list, the firewall will select the least impacted path to send the traffic. If the path conditions change and the best available path is disqualified, the path selection algorithm moves new sessions to the next qualified path. When path conditions improve on the original path and it becomes qualified again, the firewall will use the Fallback Hold Time (defined on the interface profile) to determine when to forward traffic onto the path again.

The link must remain healthy and qualified for the duration of the hold time in order for it to be reinstated. If it becomes disqualified during the hold time, the timer is restarted. This helps to prevent excessive path failover for links that have a frequent "flip flop" behavior and helps to stabilize the network.

The following chart from the SD-WAN administrator guide summarizes how each traffic distribution method handles an application's sessions when a path is degraded and when it recovers.

PATH CONDITION	TOP-DOWN PRIORITY	BEST AVAILABLE PATH	WEIGHTED SESSION DISTRIBUTION
Session on existing path failed a path health threshold (brownout)	Affected session fails over to better path (if available)	Affected session fails over to better path (if available)	Affected sessions don't fail over
Top-Down or Best Available Path recovered: existing path is still qualified (good)	Affected session fails back to previous path	Affected session stays on existing path, doesn't fail back	Affected sessions don't fail over
Top-Down or Best Available Path recovered: existing path fails a health check	All sessions fail back to previous path	Selective sessions fail back to previous path until affected existing path recovers	Affected sessions don't fail over
Existing path is down (blackout)	All sessions fail over to next path on list	All sessions fail over to next best path	All sessions fail over to other tags based on weight settings
Brownout with no qualified (better) path	Take best available path	Take best available path	Take best available path

In general, the firewall offloads congested links by first moving new sessions to a better path. If the situation worsens and the link experiences further brownout conditions, existing sessions will be switched to a better path and there are two conditions to be aware of based on the VIF type:

Tunnel VIF - The egress interface is an SD-WAN tunnel VIF that goes to another SD-WAN firewall within the organization. There are no NAT policies applied between the source and destination. In this network topology, PAN-OS SD-WAN will move existing impacted sessions from the disqualified path to a better performing path. In a complete link outage, the firewall will attempt to move all sessions to a better performing path.

DIA VIF - The egress interface is an SD-WAN DIA VIF that goes directly to the internet. There is a NAT policy in place to translate internal IP addresses to a public routable address. In this network topology, PAN-OS SD-WAN will keep the existing impacted session on the disqualified path and move new sessions to a better performing path. This is done to preserve TCP applications that may fail when moved to an alternate path that has a different NAT policy

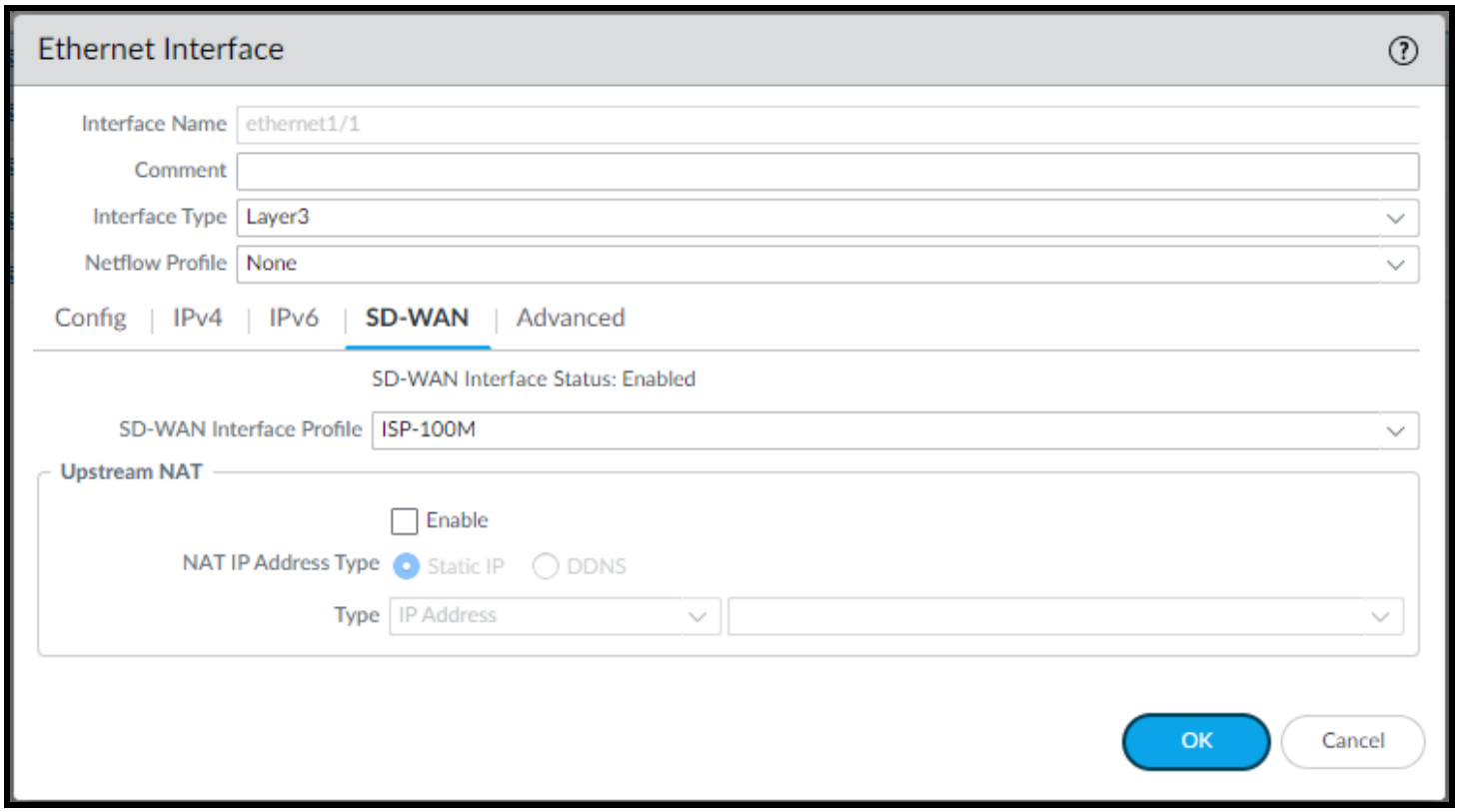
and translates the source IP to a different address mid-session. In most cases, moving new sessions to a different path will lower congestion and allow the impacted path to improve. In a complete link outage, the firewall will attempt to move all sessions to a better performing path.

Other Path Selection Considerations

In addition to the path measurement and traffic distribution profiles discussed in the previous sections, there are several other SD-WAN controls that can influence how the firewall distributes traffic and if the traffic is sent through a VPN tunnel or in clear text.

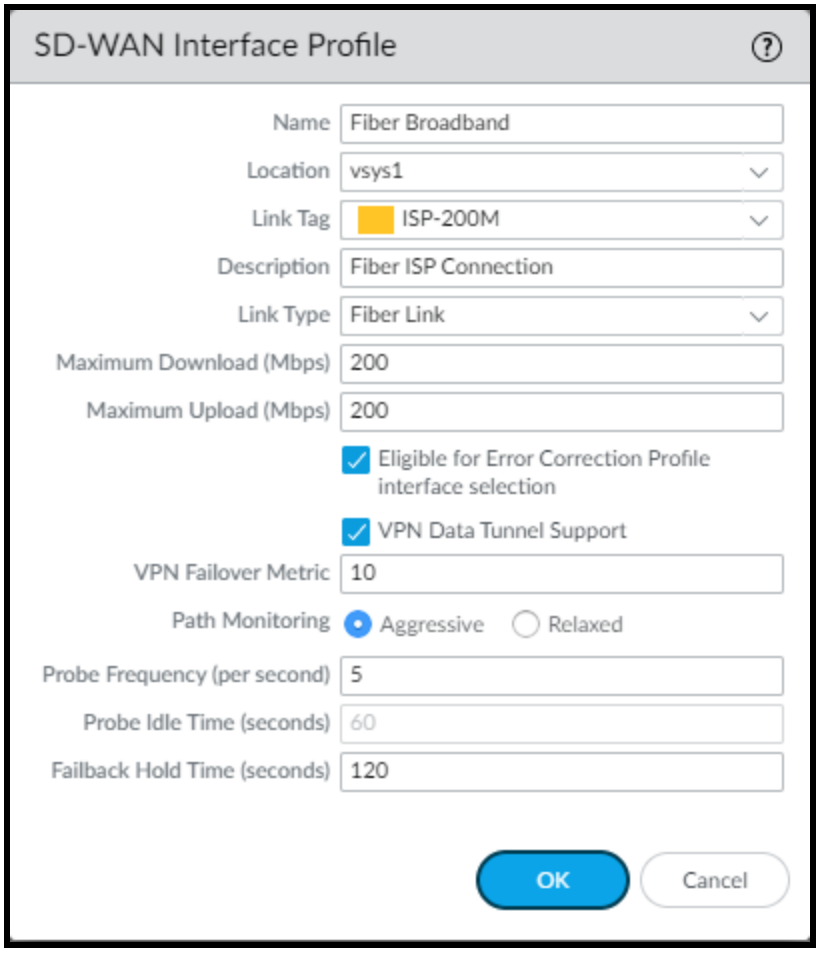
SD-WAN Interface Profile

The SD-WAN Interface Profile is used to define the link characteristics of the ISP circuit and it is assigned to a physical or virtual interface in the Ethernet Interface SD-WAN settings. Each SD-WAN interface can only have one interface profile assigned to it.



The following concepts have already been covered in previous sections so they will not be covered here. Please reference the other sections to get more information on:

- Path Monitoring
- Probe Frequency
- Probe Idle Time
- Failback Hold Time



The key items in the SD-WAN Interface Profile that can influence traffic steering include the following:

Link Tag - An identifier that is used to associate an ISP circuit with a traffic distribution profile. Each SD-WAN interface can only be assigned one link tag, but link tags can have multiple SD-WAN interfaces. As previously mentioned, SD-WAN interfaces assigned with the same link tags are used to form link bundles.

The following chart from the SD-WAN administrator guide summarizes how session distribution is handled with link tags and their respective paths.

PATH CONDITION	TOP-DOWN PRIORITY	BEST AVAILABLE PATH	WEIGHTED SESSION DISTRIBUTION
Multiple links with the same SD-WAN Tag	Share session load equally among links within SD-WAN Tag	Share session load based on best path within SD-WAN Tag	Share session load based on % weight assigned to SD-WAN Tag
Multiple links with different SD-WAN Tags	Share session load based on list priority, load link(s) in first SD-WAN Tag first.	Share session load based on best path from all SD-WAN Tags	Share session load based on % weight assigned to SD-WAN Tags

Eligible for Error Correction Profile Interface Selection - This parameter controls how SD-WAN's Forward Error Correction and Packet Duplication allocate packets to perform packet loss recovery. If enabled, the interface will be used to support FEC's parity packets if there is an FEC profile attached to the SD-WAN policy for the matching application. Similarly for PD, if enabled the interface will be used to duplicate the original packet flow for the applications that match an SD-WAN policy containing a PD profile.

TIP: You will typically disable this feature for links that do not have enough bandwidth to support the packet recovery features or on links that will not perform packet recovery. For example, a forestry company's branch office is located in a rugged mountainous area using low bandwidth satellite and expensive LTE connections for connectivity. They disabled error correction capabilities on the satellite link due to its low capacity and enabled it on the faster high capacity 5G LTE connection to ensure application quality. Although it's expensive, the need for application quality overrides the additional bandwidth cost.

If FEC or PD is configured in an SD-WAN policy, you must have at least one SD-WAN interface enabled with this capability, or the firewall will not be able to perform error correction.

VPN Data Tunnel Support - This parameter allows you to control how the data is sent over the path - through the SD-WAN VPN tunnel or outside of it in clear text. The default is enabled. There are some link types that may need the traffic sent in clear text in order to apply additional security features, perform traffic acceleration, or provide some other function. If this is the case and the link is already secure without the traffic going over the SD-WAN VPN tunnel, you can disable the VPN Data Tunnel Support feature and have the firewall send the traffic in clear text.

For example, some satellites can perform TCP acceleration or apply additional filtering when the traffic is in clear text. If the firewall forced all traffic through its VPN tunnels, this additional functionality will not be possible. Another use case justifying the disablement of the VPN data tunnel support is when private WAN links between locations are considered very secure and the extra VPN encryption overhead is not desired.

The firewall will still require the SD-WAN VPN tunnel to be created between the two sites even if you choose to send the data over the clear text channel. The SD-WAN VPN tunnel is used by the firewall to conduct its health checks and determine the latency, jitter, and packet loss of the connection. This parameter must be enabled to support DIA AnyPath failover (see DIA AnyPath section for more information).

VPN Failover Metric - This parameter sets the failover priority of the VIF's VPN tunnels and is used by the DIA AnyPath feature. With DIA AnyPath, any VPN tunnel can be used for DIA failover, so it's important to prioritize the order in which the

VPN tunnels are used. For example, a branch office has three ISP circuits and two of them are low cost broadband services while the third is an expensive 5G LTE backup circuit. Panorama automatically creates the M x N full mesh DC VIF interface to connect the branch to the data center hub and assigns three VPN tunnels to the VIF, resulting in three possible paths with one for each ISP circuit. If the VPN Failover Metric was left at the default value of "10", DIA AnyPath will use all of the tunnels equally as an alternate path for DIA failover.

In order to control when the LTE circuit is used, the failover metric can be set to prioritize the selection. Like router metrics, the lower metric value means higher priority for path selection. In this use case, we can give the two broadband interfaces the default metric of 10 and assign the LTE interface a much higher metric of 100 to delay its selection.

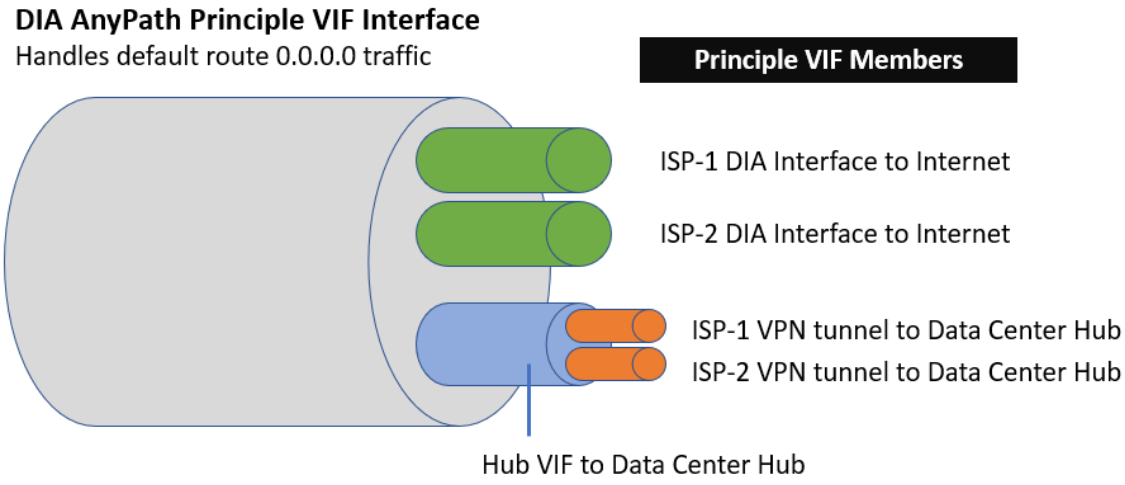
DIA AnyPath

The DIA AnyPath feature is designed to provide failover capabilities to DIA applications that normally go directly to the internet, by allowing them to be redirected through a hub location. DIA AnyPath is only applicable to internet bound applications and not to private applications that normally go through an SD-WAN VPN tunnel.

With DIA AnyPath, the default route is assigned to the DIA virtual interface (VIF). This powerful interface is known as the Principle VIF and it can contain other VIFs that go to hub or branch locations as members, and a maximum of nine interface and VIF members are permitted. VIF members can belong to different security zones to allow flexibility for failing over traffic that normally goes to an untrusted internet zone to another firewall in an internal trusted zone, such as a data center hub firewall.

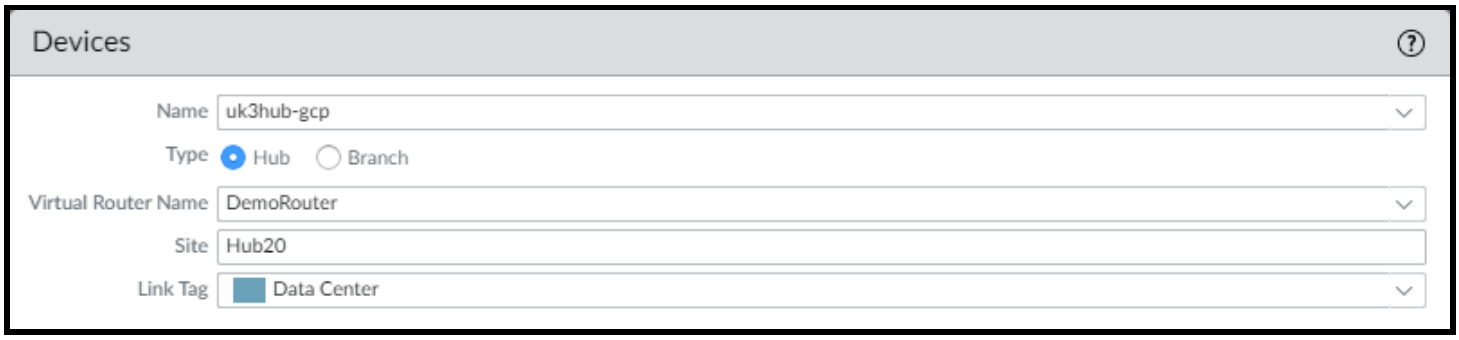
TIP: The DIA VIF is the same as the Principle VIF in the PAN-OS command line interface.

In the illustration below, the DIA AnyPath Principle VIF is handling all traffic that goes out of the firewall's default route. In addition to the two ISP circuits with DIA access, the data center hub VIF is also added as a member to the principle VIF and the firewall can use any member to send DIA traffic.

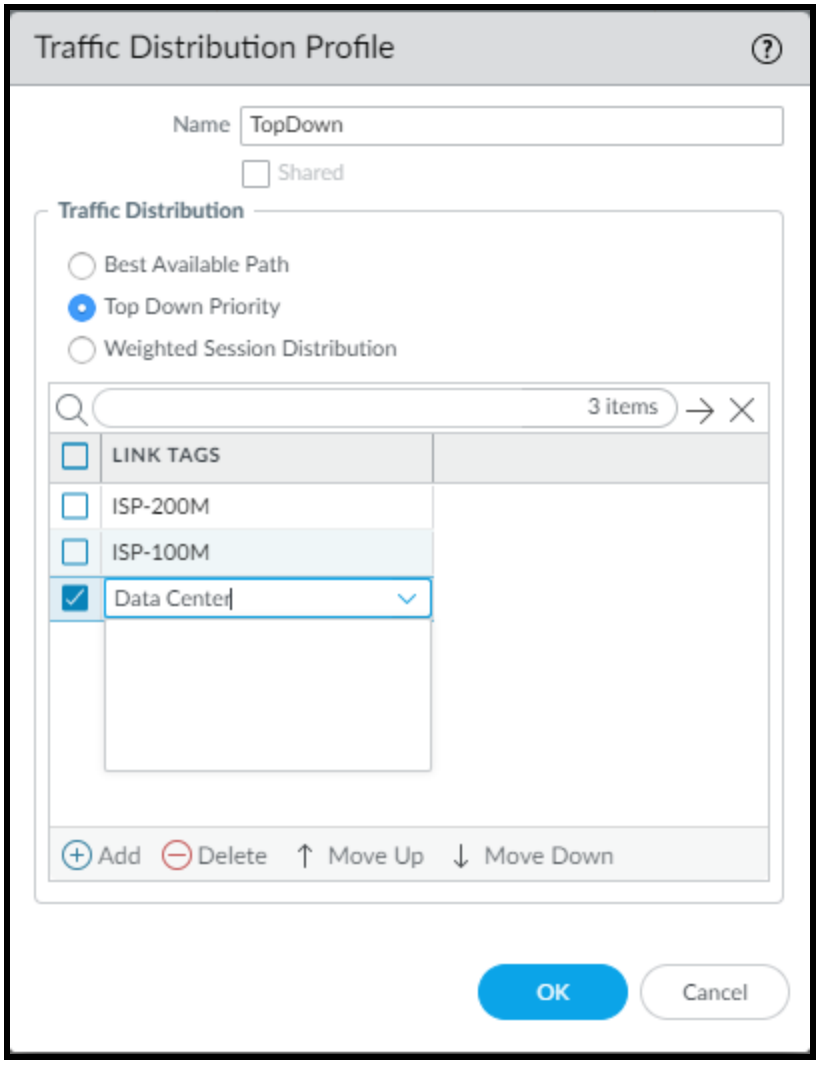


The order in which the principle VIF's members are selected is determined by the link tags in the traffic distribution profile. If the tunnel VIF going to a branch or hub contains multiple VPN tunnels, the selection order of the VPN tunnels can be controlled through the VPN Failover Metric.

When a DC hub device is added to SD-WAN, you can decide if DIA applications from a branch location can use this hub's internet connections for failover. If this is permitted, assign a link tag to the hub device as shown in the illustration below and Panorama will automatically configure and associate the link tag to the corresponding VIF interface. In this example, the link tag called "Data Center" is assigned to the "uk3hub-gcp" device.



During a DIA failover condition, SD-WAN will consult the traffic distribution profile's link tag order to select the path to fail the DIA application to. The hub's link tag is placed below the DIA link tags in the traffic distribution profile's top down distribution method as shown in the illustration below.



In this example, there are two DIA internet paths for the branch firewall to take, ISP-200M and ISP-100M. With a top down distribution method, the SD-WAN policy will forward matching application traffic through the interfaces associated with the ISP-200M link tag first. When there are no qualifying paths in the first link tag, the interfaces associated with the second link tag are considered next. When there are no good DIA paths from the first two link tags, the branch firewall will use the Data Center link tag that goes back to the DC hub as an alternate route to the internet.

The health checks performed through the VPN tunnels connecting the branch to the hub will determine which healthy tunnel to take. If there are expensive circuits that should not be used initially, the VPN Failover Metric setting in the SD-WAN interface profile can be used to control the DIA AnyPath tunnel selection order. Simply raise the metric higher to decrease the selection priority.

The data center hub firewall will receive the DIA failover traffic and perform a route table lookup. Based on the destination IP address, it determines that the traffic is headed for the public internet and its SD-WAN policy performs an application match and the traffic is forwarded based on the path quality and traffic distribution profile outcomes.

SD-WAN interfaces supporting DIA AnyPath failover must have the VPN Data Tunnel Support checkbox enabled on its interface profile. This allows untrusted internet traffic to be sent to the hub firewall as clear text over the hub VIF while the private internal traffic is sent through the hub VIF's VPN tunnels. The firewall separates internet traffic from private traffic to maximize security and resource utilization efficiency.

TIP: As the hub tags are not relevant to the hub itself, configure DIA AnyPath traffic distribution profiles under the device group (DG) in Panorama, and not as a "shared" object. Leverage the hub and branch SD-WAN zones to quickly set up security policies for DIA AnyPath failover traffic.

DIA AnyPath End-to-End Health Checks

As previously stated, the SaaS Quality Profile is used to measure the DIA path's health (latency, jitter, packet loss) from the firewall to an IP address or a URL on the internet. In the use case where the DIA traffic needs to failover to a hub location to get a better internet path, an end-to-end health measurement is needed. The branch does not initiate a separate probe through the hub to monitor the SaaS application's path health and this reduces the number of probes going to the hub firewall. To maximize bandwidth efficiency, two path measurements are added together to obtain the end-to-end measurement.

End-to-end health = branch-to-hub health + hub-to-application health

To instruct SD-WAN to combine the two separate health measurements, two identically named SaaS quality profiles are configured - one on the branch firewall and one on the hub firewall that receives the DIA failover traffic. The hub firewall sends its measurements back to the branch firewall to perform the end-to-end calculation.

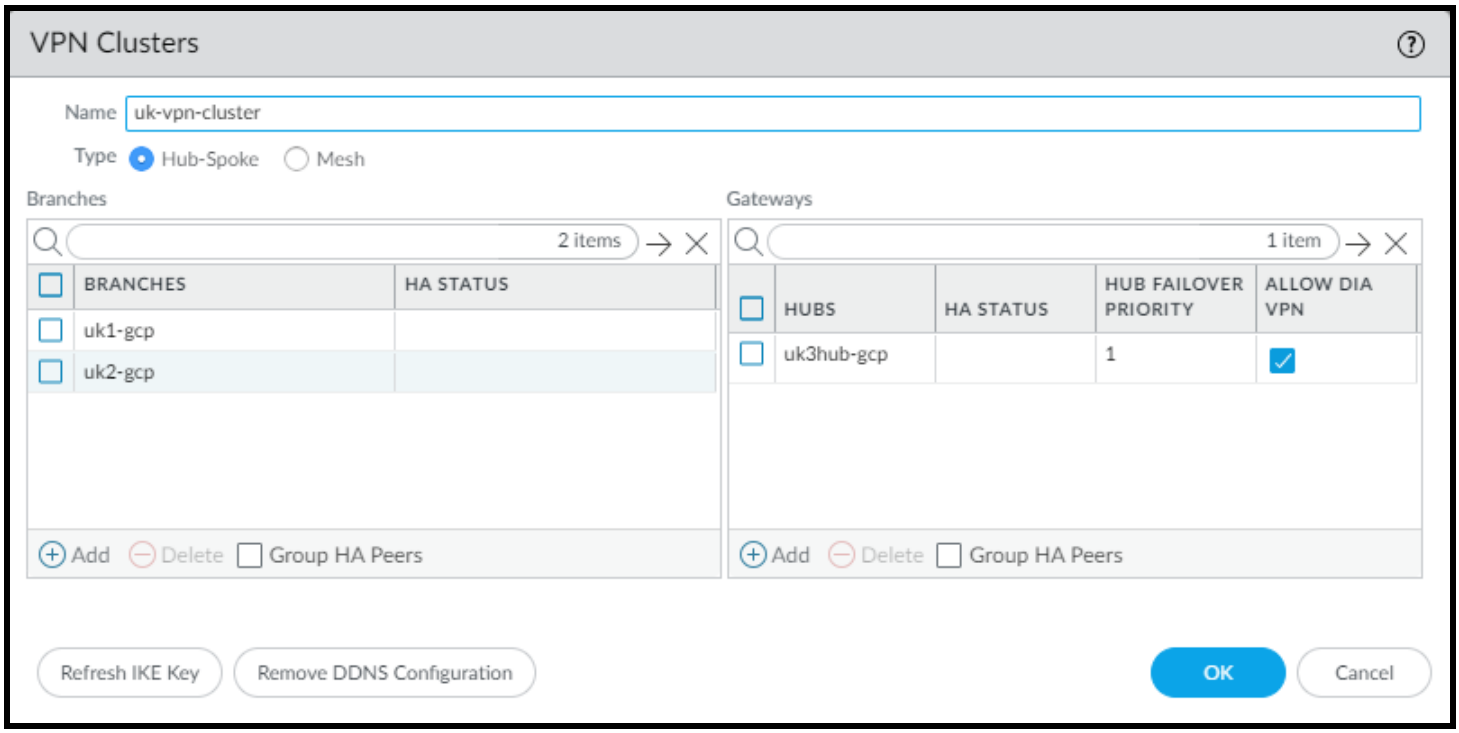
Only the names of the SaaS Quality Profile needs to be the same; the IP or URLs monitored can be different. An example of when you may want to use different IP addresses for each SaaS quality profile is when the SaaS application can be serviced from multiple cloud locations and the failed over sessions don't have to go back to the original data center to be serviced.

In this use case, you can put a URL or IP address in the monitoring profile that is closer to the hub's geographical location to get a more accurate end-to-end measurement. For more information on this use case, reference the SD-WAN administrator guide for Failover SaaS Monitoring.

Blocking DIA AnyPath Failover Traffic

There may be times when you need to stop the branch's DIA failover traffic from going to a hub. For example, in a hierarchical administration structure, the hub location's network administrator wants to prevent the branch offices from sending their internet traffic to the hub site and using the limited internet bandwidth they have. Or the hub's network administrator needs to temporarily suspend the DIA AnyPath failover traffic as one of their internet circuits has gone down and they need to preserve bandwidth for the hub's internet traffic.

To control a hub's ability to accept DIA failover traffic from another location, enable or disable the "Allow DIA VPN" option in the SD-WAN VPN Cluster configuration menu as shown below.



If the Allow DIA VPN checkbox is selected, the hub will allow DIA failover traffic from the branches that are connected to it. If unchecked, it will reject the DIA failover traffic. Any hub that is permitted to handle DIA failover traffic from the branch locations must have at least one path to the internet. Otherwise, the traffic will be blocked and dropped by the hub firewall.

Per-Application Split Tunneling

The SD-WAN policy allows a different traffic distribution profile to be assigned to any set of matched applications and with DIA AnyPath's principle VIF permitting up to nine VIF members, applications can have different default route paths. This is especially important if there are use cases requiring different default route handling for specific applications.

For example, a small number of internet applications need to be backhauled to the data center to have additional traffic inspection performed by a specialized security appliance located only in the data center. The majority of internet traffic will

go directly to the internet and be secured by the branch's SD-WAN firewall. For this use case, two traffic distribution profiles are created to control default route selection:

Hub Tag Priority - This traffic distribution profile uses a top down distribution method with the DC hub's link tag positioned at the top of the list. It is assigned to the SD-WAN policy that matches the applications that need to be redirected to the data center for the application's default route. A backup link tag can be placed under the DC hub link tag to provide failover if required.

Internet Priority - This traffic distribution profile uses a top down distribution method with the DIA VIF tag positioned at the top of the list. It is assigned to the SD-WAN policy that matches the remaining internet applications and will route the traffic directly to the internet. A backup link tag can be placed under the DIA VIF tag if internet failover is required.

Summary

This paper demonstrates the concepts used in PAN-OS SD-WAN's health monitoring and path selection components and shows how each component is used in the SD-WAN policy. Security concepts were outside of this paper and were not discussed, but security policies must be part of the overall SD-WAN design and implementation to control and inspect traffic between each location and the internet.

For more information on SD-WAN concepts, please visit the following links:

- PAN-OS Administration Guide
- PAN-OS SD-WAN
- Panorama