WHITE PAPER

Gaining Visibility to Distinguish Server vs. Network Issues

A Guide for the Server and Network Team

Introduction

In data center infrastructure, the lack of full network visibility down to the server itself creates serious limitations for IT management when troubleshooting problems. Is the problem in the server, network interface card (NIC), top-of-rack (ToR) switch, or something deeper in the data center network?

This can result in mean-time-to-resolution (MTTR) challenges, and it can be intensified by new data center paradigms. For example, hyper-converged infrastructures promise to optimize application delivery, coupled with increased resource and operational efficiency, but reduce visibility for network teams as they work to resolve issues.

In general, enterprise data centers are adopting turn-key technologies to drive agility and cost efficiencies. The network layer, when contrasted with compute and virtualization infrastructure, is typically the least agile part of the data center to design, configure, and operate. When it comes to troubleshooting and driving operational efficiencies, network issues such as congestion and packet errors are heightened due to the complex interaction between the network and the teams managing multi-vendor hyper-converged solutions in both large-scale data centers and remote office IT environments.

This document outlines the challenges for two key data center groups—server administrators and network managers—and the paths each can take to quickly identify the root cause of I/O performance issues.

For a server administrator, the challenge may be to quickly identify that a packet-loss issue in the path of a particular flow is not caused by the server I/O components (e.g. driver or NIC hardware). As an example, consider a troubleshooting ticket that's about to be opened due to poor application performance, and the IT team needs to collect data:

- Identify where the workloads exist across within the data center.
    - Are the workloads virtual machines, bare metal, containers, or a mixture of all of them?
    - Which physical servers host the workloads?
- Which top-of-rack switches do the servers connect?
- Take a bi-directional full packet capture for all flows involved in the transaction.
- For all servers in the path of a business transaction, identify if there is packet loss happening:
    - at the PF or VF to/from PCIe bus
    - within the NIC (any internal components such as the ASIC between the PCI bus and physical network uplinks)
    - on the physical uplinks connected to the top-of-rack switches
- Present the data to the network team demonstrating that an I/O issue in the servers is not the cause; suggesting the network team researches any network performance issues that may be at fault.

From a network administrator's perspective, the challenge is to understand problems that may lie beyond the switch port; hyper-converged solutions add to the challenge of decreased visibility on the server side.

In either domain, the time and difficulty level of gathering meaningful data can affect MTTR if conflict arises after a ticket is opened, and hamper predictive maintenance. Connecting the dots at the edge for an administrator becomes fast and painless with the Pensando Distributed Services Platform, which can be gradually introduced into an enterprise without requiring a significant upfront investment financially or a steep learning curve to see the benefits.

The Pensando Distributed Services Platform

The Pensando platform enables various network, security, storage, and visibility services, implemented by a set of Distributed Services Cards (DSCs), which are centrally managed and monitored by the Policy and Services Manager (PSM) controller. DSCs are PCIe host adapters that are deployed in the data center servers. The PSM is a centralized management platform, leveraging an intent-based model that performs lifecycle management while delivering pervasive visibility, network, storage, and security policies to the DSCs for services implementation at the edge. The PSM provides both a secure API and a GUI management framework to enable maximum provisioning flexibility.

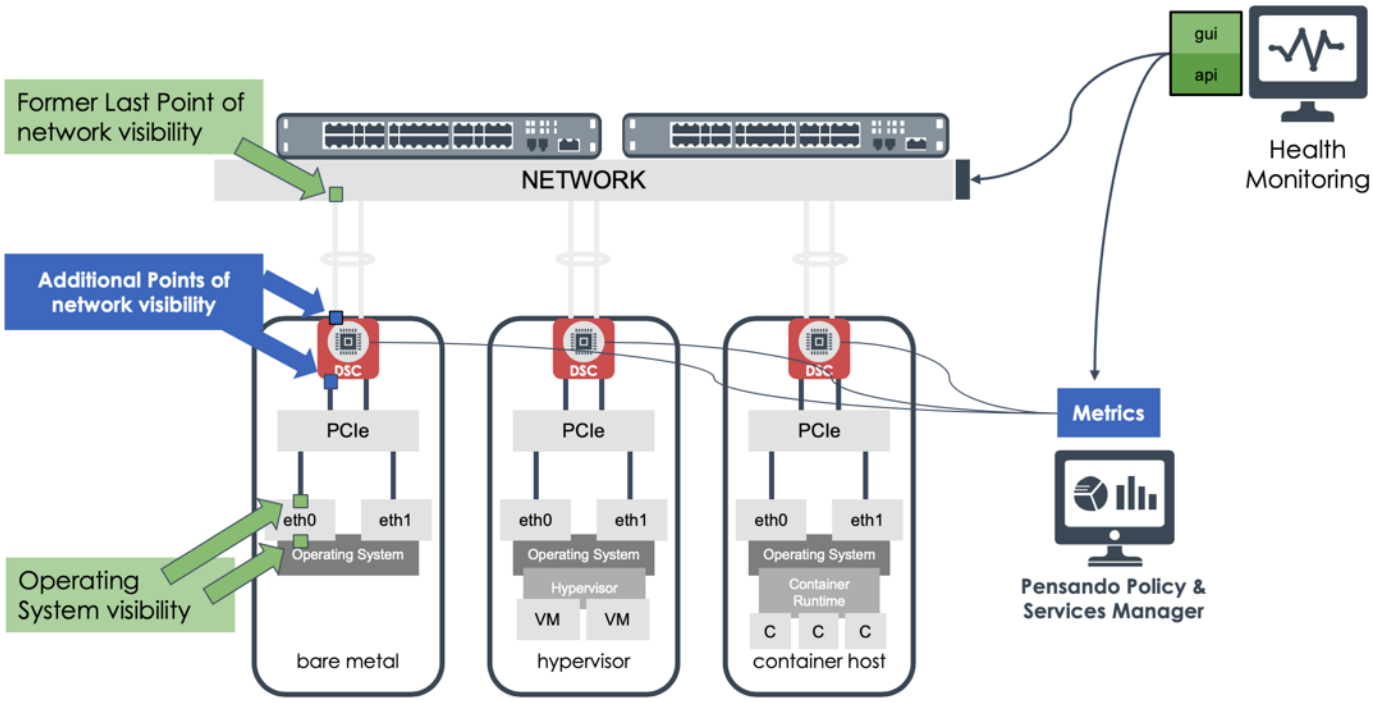


Figure 1. DSCs add additional points of visibility closer to workloads in each host, with consolidated observability coordinated via the PSM

The remainder of this document describes the journey in which the Distributed Services Platform can solve visibility and troubleshooting challenges in enterprise deployments.

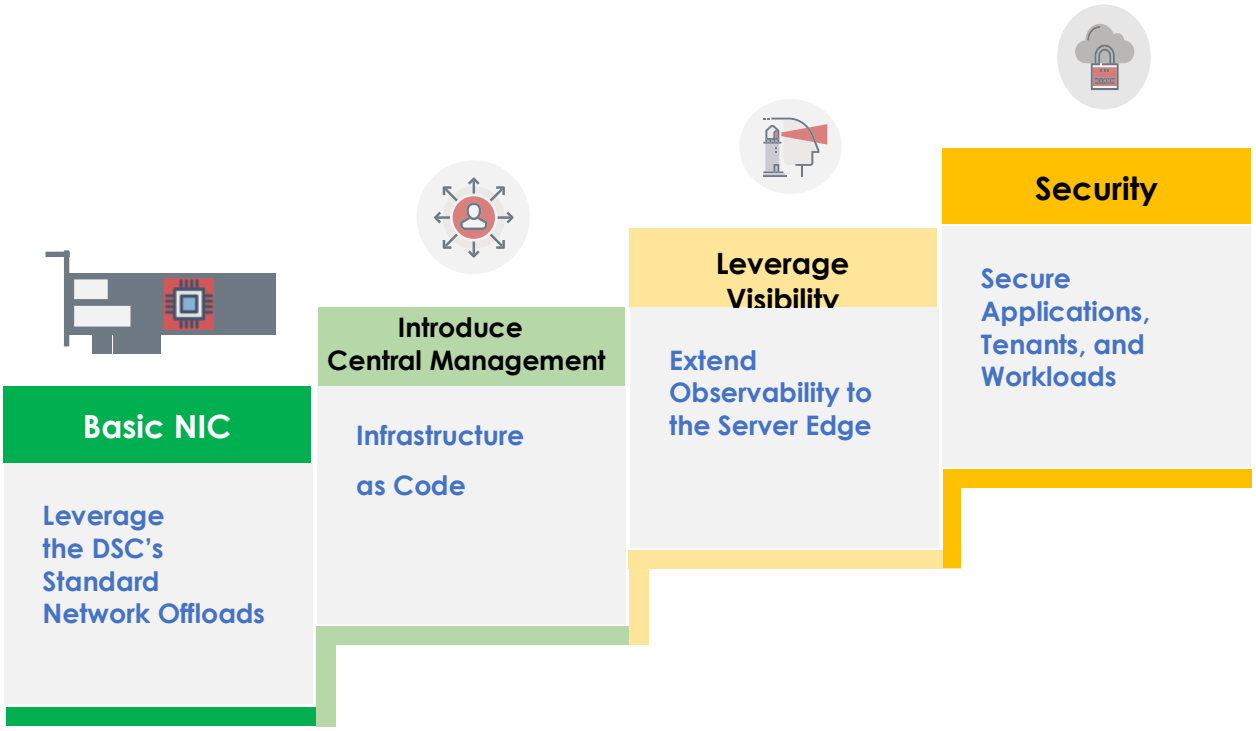


Figure 2. Adoption model, showing the non-disruptive introduction of key visibility, manageability, flexibility and security capabilities

From the Host, DSCs Look Like Just Another NIC

The first step to introduce the Pensando solution in an enterprise network consists of deploying DSCs in place of legacy NICs when ordering hyper-converged infrastructure solutions. Once the server is powered on, and the driver for the DSC is installed, it can be deployed and managed like any other server. Pensando offers drivers for all major contemporary operating systems.

Introducing the PSM and Gaining Visibility

The next step to gain more benefits is to use the PSM to centrally control the DSCs and unlock additional lifecycle management and observability functionality.

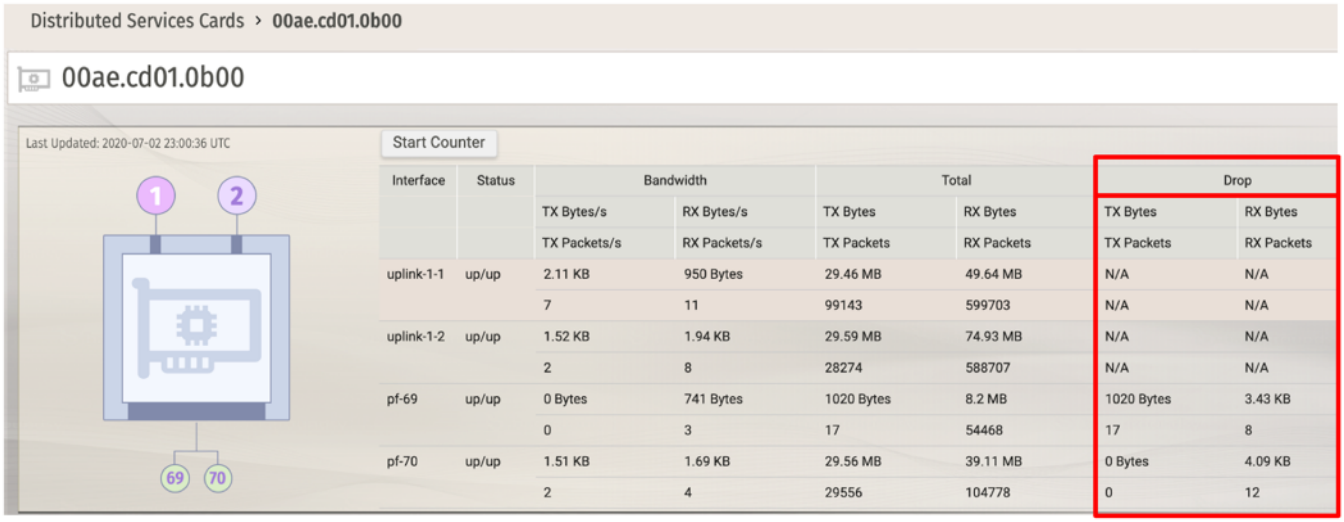
Once a PSM is deployed as a central manager, each DSC discovers the PSM to which it is assigned and is automatically assigned to one of the PSM's policy groups.

Among other capabilities, the PSM provides full life cycle management for all its DSCs, including firmware upgrades, health monitoring, centralized events, alarm reporting, and a robust set of metrics, which can be displayed to help with troubleshooting and provide pervasive visibility. The administrator can use the PSM to access telemetry data collected by the DSCs ("Fields" in the figures below), organized in various categories ("Measurement" in the figures below).

Server and network administrators can use this powerful distributed monitoring capability to take the pulse of the network, identify potential performance bottlenecks, and remediate developing issues before they become a problem.

Figure 3 displays the first step in getting details about the DSC and its connections to the network and the host OS. The initial output is very similar to what would be seen with a `show interface Ethernet 1/1` command on a typical top-of-rack switch, or `ethtool -S` on a host system. The passing of bi-directional traffic through the DSC and its uplink can be quickly identified, as well as determining if the interface is experiencing any drops. If the physical network is not experiencing any drops and the DSC doesn't report drops for the physical

network connections(s) or towards the PCIe bus, then the output below can help narrow down the possible root cause of packet drops to a kernel function within the workload's Operating System.



Interface	Status	Bandwidth		Total		Drop	
		TX Bytes/s	RX Bytes/s	TX Bytes	RX Bytes	TX Bytes	RX Bytes
		TX Packets/s	RX Packets/s	TX Packets	RX Packets	TX Packets	RX Packets
uplink-1-1	up/up	2.11 KB	950 Bytes	29.46 MB	49.64 MB	N/A	N/A
		7	11	99143	599703	N/A	N/A
uplink-1-2	up/up	1.52 KB	1.94 KB	29.59 MB	74.93 MB	N/A	N/A
		2	8	28274	588707	N/A	N/A
pf-69	up/up	0 Bytes	741 Bytes	1020 Bytes	8.2 MB	1020 Bytes	3.43 KB
		0	3	17	54468	17	8
pf-70	up/up	1.51 KB	1.69 KB	29.56 MB	39.11 MB	0 Bytes	4.09 KB
		2	4	29556	104778	0	12

Figure 3. PSM visual display of the interface counters of a DSC. Each uplink connecting to the top-of-rack switches can be easily identified, as well as the physical or virtual functions that connect to the PCIe bus. Each interface displays its status, bandwidth, total throughput and drops.

Each DSC collects a broad set of metrics organized in various categories. A subset of these metrics can be selected to create custom graphs on the PSM (such as the graphs shown in the figures below) showing the latest values of the metrics, offering a dashboard of the health and performance of the host's DSC. Detailed performance and error metrics are built into user-customized charts that aggregate the data over time for a comprehensive view for all data points collected.

Figure 4 illustrates the metrics collected for a given DSC for a node experiencing drops related to problems such as TCP RSTs, and TCP out-of-order windows.

- Session: (# CPS, # TCP/UDP/ICMP, TCP RSTs, Window Size Zero, etc.)
- Interface: (tx/rx unicast, multicast, broadcast, error stats, etc.)

Figure 4. DSC Interface Metrics

In a future software release, the PSM will allow administrators to configure policies that define a metrics-based alerting behavior for their DSCs, enabling sites to customize the thresholds for their specific use cases. As an example: upon exceeding a configurable metric error threshold, the PSM operator may request a full-packet capture for that DSC node, to archive to a remote system for auditing and further analysis. A PSM operator will be able to run the troubleshooting tool to focus closely on metrics between two workloads and related flows, as shown in the figure below.

Figure 5.

In Figure 6, the measurement "Uplink Interface Packet Statistics" provides rich details about the interface statistics. Custom metrics can be created to monitor specific details for all interfaces on the DSC connected to the host operating system as well as connected to the network being monitored or undergoing troubleshooting. The display allows charting and monitoring of these specific metrics for any specified interval: hour(s), day(s), week(s) or month(s).

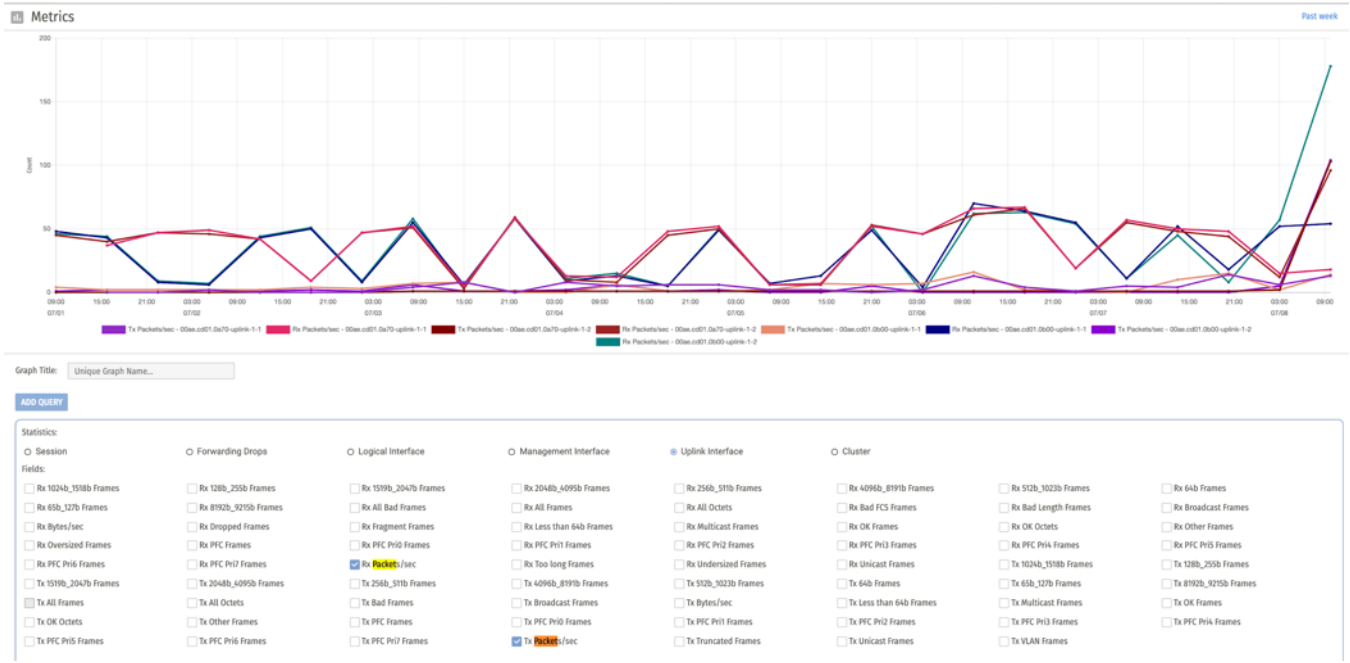


Figure 6.

Figure 7 represents metrics that summarize specific DSC sessions such as TCP RST sent, TCP sessions, half-open TCP sessions, and drops.

Figure 7.

Figure 8 presents valuable metrics relating to ASIC health on the DSC.

Figure 8.

Figure 9 presents the connections per second sustained by a DSC or group of DSCs.

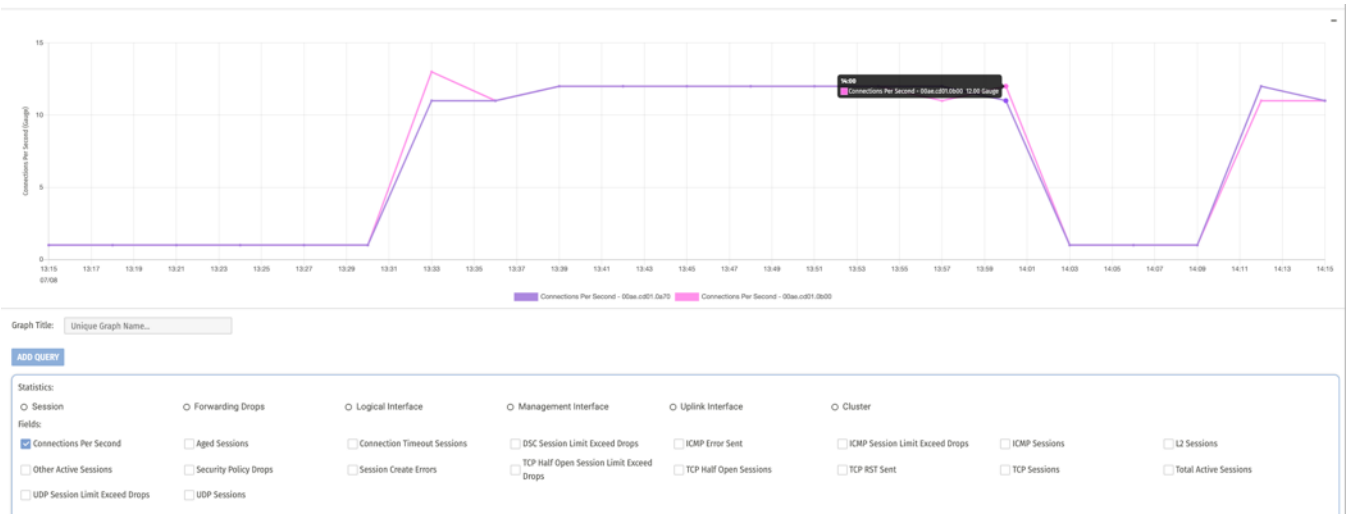


Figure 9.

Packet Mirroring

If the server network I/O looks clean from the DSC metrics reviewed, an administrator can use the DSC as a virtual TAP (Test Access Point) for full packet captures. Dynamic Flow Mirroring provides the ability to inspect packet content at line-rate, isolating and extracting application-specific traffic which can then be delivered to appropriate tools for further processing. Flow Mirroring uses traffic replication and spanning to a production network or packet broker network, which then delivers packets to an analytics engine that can perform content inspection and correlation for compliance.

If deeper and richer visibility is needed when troubleshooting, wire-rate bi-directional complete packet captures can be enabled on the fly. A DSC can make a copy of each packet matching a given mirroring policy and send it to a collector (e.g., Splunk, ELK, Wireshark) using ERSPAN (Encapsulated Remote Switched Port Analyzer) encapsulation.

The administrator can choose whether to send all of the traffic transiting one interface of a DSC (which is called *interface-based mirroring* or *bidirectional ERSPAN*), or only packets that match mirroring policies configured on

the PSM. As shown in Figure 10, mirroring policies can be defined as a flow identified by a 5-tuple: source and destination IP address, transport protocol, source, and destination transport port.

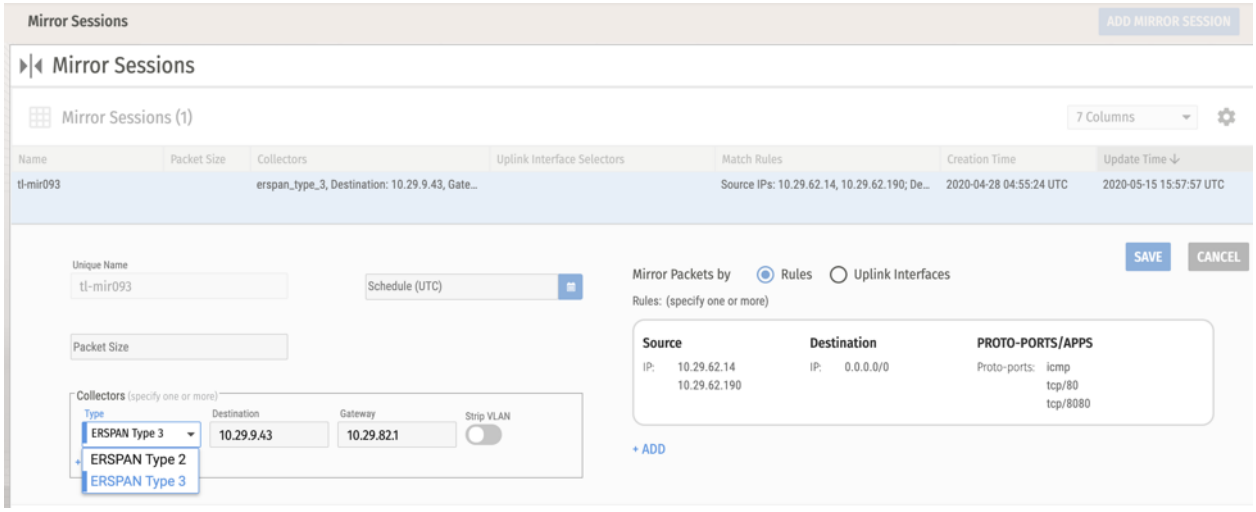


Figure 10. Mirroring Policies

Flow-Level Visibility

An administrator can also enable flow visibility to collect metadata associated with every flow that traverses a DSC. Flow exporters on the Pensando DSC can be enabled to export flow information and statistics to flow monitors on remote NetFlow collectors in IPFIX (NetFlow.v10) format. Centralized configuration allows tuning of export timeouts, export intervals, export formats, and specifying output to more than one collector. In addition to exposing the standard set of NetFlow fields, the DSC also reveals flow start and last seen time, maximum segment size, and state, offering a wealth of information for troubleshooting or accounting use cases. As shown in Figure 11, each flow is identified through the 5-tuple: source and destination IP address, transport protocol, source, and destination transport port. In a future software release, flows will be tagged by labels ingested from the VM or container orchestrator.

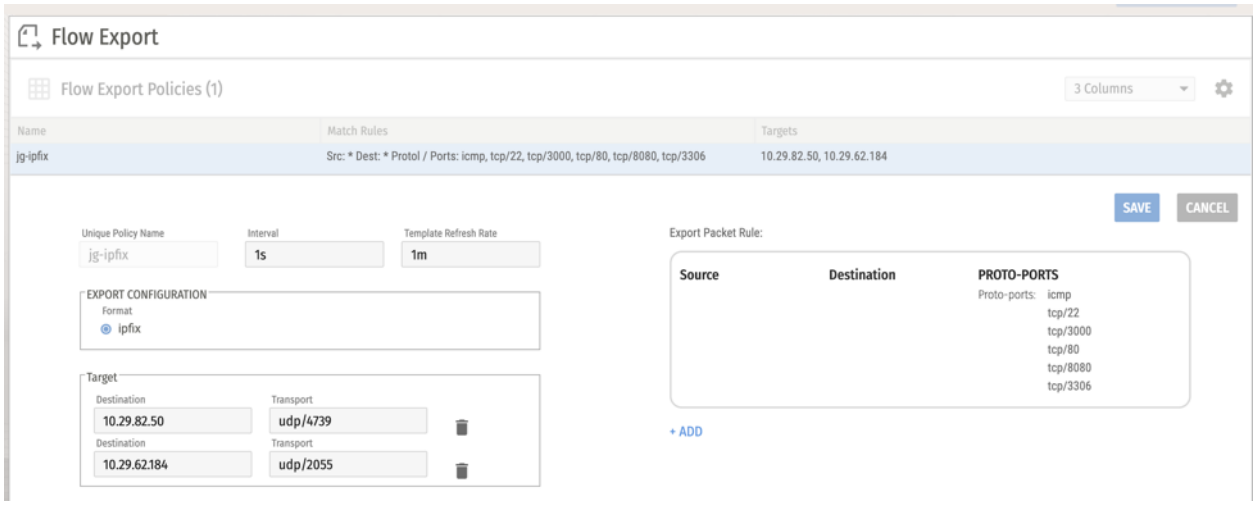


Figure 11. IPFIX / NetFlow configuration

The PSM administrator can specify the collector (or target) that should receive the flow information in IPFIX format, and the transport protocol.

The administrator does not have to worry about *which* DSC should collect the flow data. The PSM and DSCs work together to ensure that the collection and export policy is available to all DSCs, and those nodes involved in the flow will automatically collect and export the relevant information. If the workload generating a flow is relocated on the network (for example, because a VM moves through vMotion or a server is migrated to a different hardware host), the Pensando Platform will detect this and continue to seamlessly collect information without any intervention by the administrator.

Through these visibility benefits, administrators can gain insight into the traffic patterns in their enterprise data centers to troubleshoot performance issues or identify performance bottlenecks before they become an issue.

Flow Logging and Tracking

The Pensando DSC supports full TCP connection state tracking, providing application-centric deep information inspection and telemetry—a function typically offered only by high-end stateful firewalls.

Once connection tracking is enabled, the flow/session entries for the following protocols are validated in the pipeline:

- TCP state and connection tracking
    - Perform TCP SYN validation, evaluating security policy prior to pipeline flow programming.
    - Validate TCP sequence and ACK numbers are within the expected TCP window, for all packets.
    - Perform session closing state tracking, FIN/RST, and adjusting TCP state to closing.

- ICMP request and response tracking
    - Using ICMP ID and sequence numbers, invalid ICMP responses can be filtered, and requests and responses can be correlated.
    - ICMP sessions can be aggressively aged out instead of waiting for the inactivity period to expire.

The Policy and Services Manager collects flow logs that indicate source and destination IPs, ports, action, rule id, direction, making it easy for users to correlate and search.

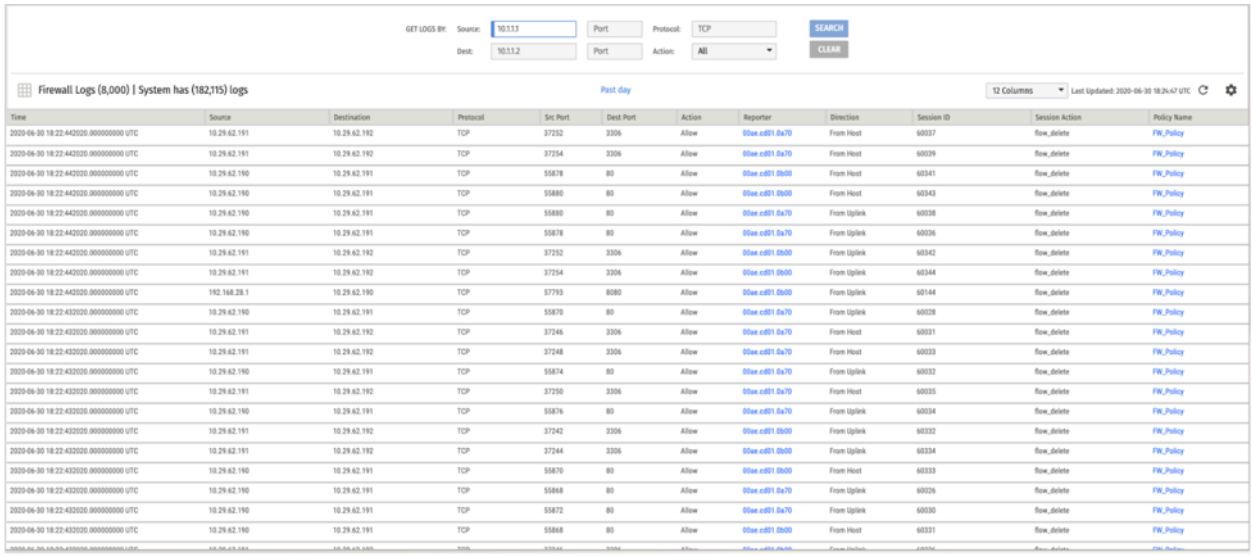


Figure 12. Firewall logs

Summary

The place to understand application behavior and identify sources of application outages is at the server edge: closer to the workloads, and directly in-line with network traffic.

Server administrators who typically capture "pcaps" directly on the host OS for troubleshooting can offload this function to the DSC as a consumable service available to both network and server teams. Server administrators can now make data-driven decisions at scale to quickly identify if server network I/O is problematic, or now have a wealth of data to prove otherwise.

Network administrator visibility no longer stops at the switchport connecting to the server—the demarcation between network and server moves all the way to the server's PCIe bus.

The Pensando Platform reduces MTTR and reduces the risk of negatively impacting business traffic in hyper-converged environments. The key benefits are:

- Zero performance hit on production traffic
- Application location awareness
- Operational simplicity
- Workload-based stats
- Complete visibility, capturing bi-directional (Tx/Rx) data

A federation of DSCs managed by the Pensando Policy and Services Manager (PSM) is designed to address barriers to visibility and other network administration challenges found in legacy monitoring platforms with hyper-converged solutions. Pensando's DSC-derived workload-based metrics will turn on the lights for server and network administrators, allowing them to see and understand network statistics at the server edge; Pensando's Dynamic Flow Mirroring solution will reduce the need for costly network TAP appliances in data center infrastructure, eliminating the need to reconfigure top-of-rack switches and apply traffic spanning sessions.

About Pensando

Founded in 2017, Pensando Systems is pioneering distributed computing designed for the New Edge, powering software-defined cloud, compute, networking, storage and security services to transform existing architectures into the secure, ultra-fast environments demanded by next generation applications. For more information, please visit www.pensando.io