White paper **AI Data Center Security**

Delivering a Secure, End-to-End AI Data Center Solution

See how Juniper Networks enables customers with the tools to build robust, performant, and highly secure AI data centers.

JUNIPER NETWORKS

www.juniper.net

 2000833-001-EN May 2025

Contents

01 Abstract

Networks for AI workloads

The largest AI data centers require extremely high performance, scale, and availability along with robust security controls to protect the infrastructure, GPU clusters, AI models, and tenants in multi-customer environments.

Juniper is the leader in networks for AI workloads with several top AI data center deployments around the world, as well as a published Juniper Validated Design (JVD) for Building an AI Data Center, which describes best practices for the backend training network architecture. This reference architecture extends our AI leadership to provide guidance on security for AI, where various security use cases are discussed, including:

- Anonymous inference at scale
- Private training frontend for multiple tenants
- Private inference with dedicated RAG per tenant
- Network management protection

01 Introduction

How AI data centers are different

As a leader in networking for AI and security for AI, Juniper has been deployed in the largest AI data centers around the world.

As a leader in networks for AI workloads and security for AI, Juniper has been deployed in the largest AI data centers around the world.

In this paper, we leverage our extensive AI data center experience to discuss the unique requirements and use cases for designing and securing scalable AI data centers.

AI data center deployments are comprised of several different types of networks, as seen in (**Figure 1**). The backend network that services the high-speed GPU clusters and dedicated storage requires a high-performance, scalable switching fabric to provide high speed, low latency, and low buffering/loss capabilities. This enables predictable and shorter AI model training timeframes.

Since the backend network is isolated from the frontend and the internet, and the focus is on speed, there is typically minimal security infrastructure deployed—other than management security. Our Building an AI Data Center solution brief describes the backend architecture in more detail.

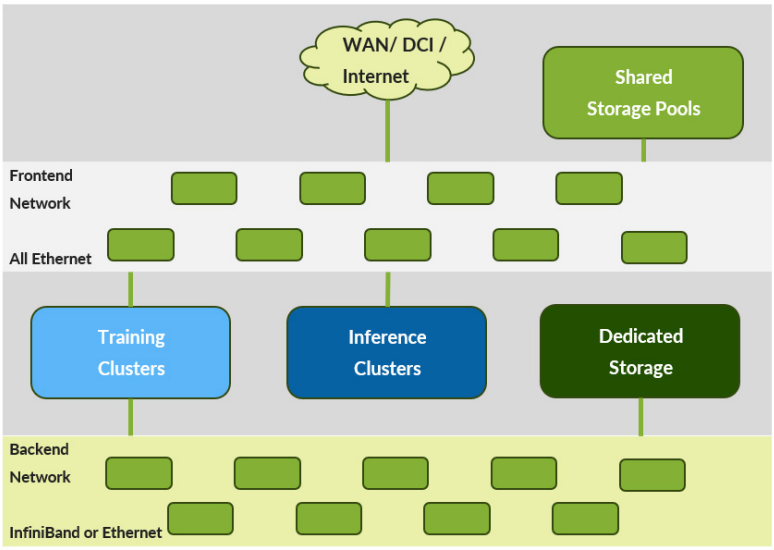


FIGURE 1

Typical AI cluster high level network layout

The frontend requirements for an AI data center are quite different. From a networking and security perspective, the frontend network is similar to most commonly deployed enterprise networks and has many of the same security requirements.

In the case of AI data centers, the frontend provides public and private services, such as frontend interfaces for customers to train their models, anonymous inference services, and even private inference services. The latter may require a customer's own Retrieval Augmented Generation (RAG) database and storage—either within the data center or securely accessible from the customer's location. Protection for remote management of the infrastructure is also required.

AI data center use cases require secure access, encryption, native fabric segmentation, and Layer 7 inspection. These frontend use cases and their particular security requirements are discussed in this paper.

01 Introduction

Threats in AI data centers

In terms of AI services, the potential associated security risks include:

- Sensitive information leaks through the exposure of training data
- Intellectual property theft in the event of a model leak
- The ability to tamper with models and bias the results
- Abuse of expensive computational resources for other tasks
- Compromised services targeting high-profile AI client assets (e.g., rogue RAG)
- Unauthorized access to low-level infrastructure, such as specialized hardware, firmware, etc.

Juniper provides an integrated and robust network and security solution to address these risks for large scale AI data center use cases.

01 Introduction

Juniper security for AI data centers

The first mission of securing AI services is to ensure that multitenant access for both humans and systems is safe and that the network prevents the elevation of admission rights as much as possible.

Although, we should clarify that effective tenant segregation depends on the particular application.

For example, if the offered service has the same login portal on the same internal host, the network doesn't necessarily provide AI services with tenant protection. In such cases, network security can still provide protections against exploitations of vulnerabilities and prevent reverse callbacks in the case of successful code execution. However, the network can be very effective in terms of separation and blast zone reduction if there are multiple separate AI service portal instances, allowing network security to steer and limit access of individual tenants.

Another consideration is that firewall segmentation of zones is a rudimentary measure that must be taken. Minimal segmentation would include the creation of zones for the management network, frontend network, associated middleware/backends, backup, and, potentially, AI/inference/storage resources. This is easily attainable on a high-performance, high-capacity device, such as Juniper's SRX Series Firewalls.

To minimize attack surfaces for both incoming and outgoing traffic—whether single or multitenant—extensive Layer 7 security features provided by SRX Series Firewalls can be utilized.

For **incoming traffic**, this involves limiting access to only what is necessary for services to operate and imposing Layer 7 (L7) security when appropriate to protect published services.

For **outgoing traffic**, the focus is primarily on ensuring that no internal systems are reaching out to hostile services, while allowing and inspecting only necessary services, such as software updates, and prohibiting regular internet browsing. And it provides protections in order to mitigate client-side application attacks if communication is eventually permitted.

These features are part of the Juniper AI-Native Security solution that's discussed later in this document, as well as in detail throughout Chapter 3 of this paper.

Transport security is the next focus, as encryption in the form of various VPNs or MACSec can be very effective in reducing attack surfaces while providing integrity and confidentiality for transferred data. It is expected that the inner control plane data will already be wrapped in TLS. VPN technologies can be particularly interesting for providing end-to-end multitenancy and segregation in AI data centers, especially in conjunction with technologies like EVPN/VXLAN. Remote access VPN also provides an authentication and authorization security layer. And finally, modern VPNs with extensions for post-quantum security can protect against the potential loss of confidentiality that will arise with the advent of quantum computing.

01 Introduction

HA and performance in AI data centers

In addition to security, AI data centers have a requirement for high availability (HA) and integration into today's typically L3-based topologies.

This makes the SRX Series Firewalls' Multi-Node High Availability (MN-HA) resiliency formation technology, with its sub-second failover capabilities, highly applicable. Initially, MN-HA was designed for service provider-grade IPSEC VPN deployments, which also fits perfectly into the previously discussed encryption-related topic. Accompanied by Juniper MX, PTX, and QFX hardware series, Juniper security solution for AI data centers becomes very robust.

Performance and scale depend on the use case. When managing high-profile AI assets, Juniper typically positions hardware acceleration-capable platforms, such as the SRX4600, the SRX5000 chassis line, and the new SRX4700 high-scale, fixed-form-factor firewalls. This is alongside Juniper MX, PTX, and QFX routing and switching hardware, as well as products for management and visibility. In high-end scenarios where a pair of elements may be insufficient, deploying a scale-out solution is also an option. Regarding VPN performance, the latest advancements include multi-threaded VPN processing, and in-line, hardware-accelerated IPSEC on the MX304 routers and the new SRX4700 firewalls. Accelerated IPSEC is particularly suitable for encrypting Remote Direct Memory Access (RDMA) traffic between sites.

02 AI data center use cases

Overview

This chapter provides an overview of potential AI data center use cases where Juniper security applies.

Currently, the use cases covered relate to the security of anonymous inference, private training frontend security (primarily focusing on VPN styles and tenant separation), security for Retrieval Augmented Generation (RAG), and network management security. There may be additional use cases outside the scope of this document, such as AI data center solutions "in a box," which is similar to what is being observed in the semiconductor manufacturing supply chain.

The generic topology of the perimeter is shown in (**Figure 2**), where "AI Workload" represents all components of AI clusters. "third-party security" might include Web Application Firewall (WAF), prompt-injection protection, model leakage protection, etc.

Ideally, Juniper JSA SIEM/SOAR/XDR and Apstra Flow Data would also be positioned to collect data from networking equipment and host operating systems for all the use cases, optionally facilitating automatic enforcement loops.

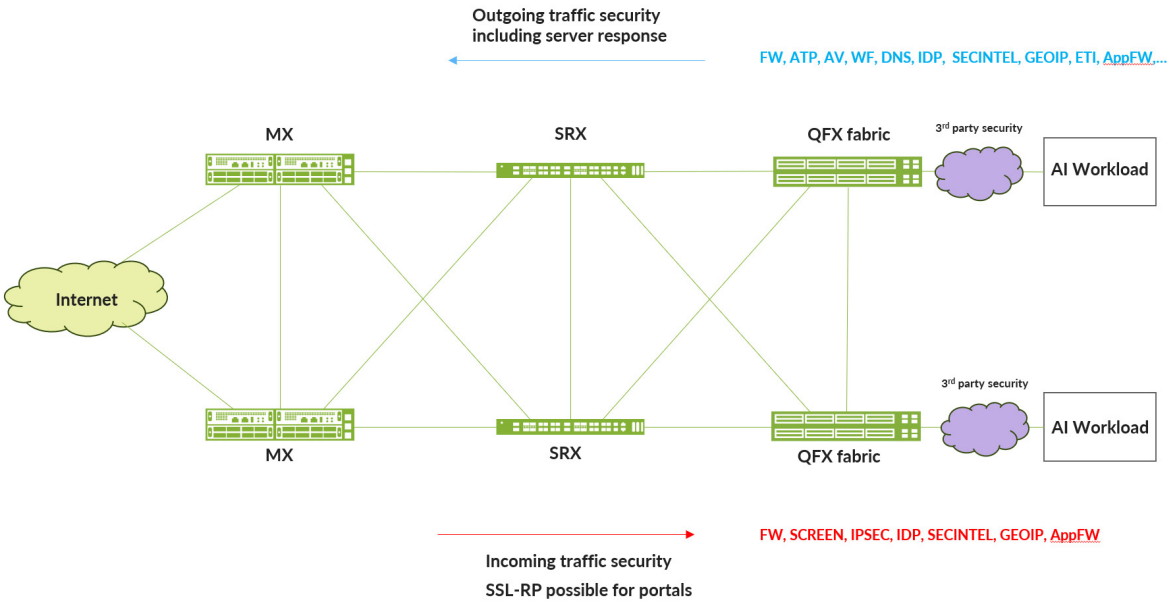


FIGURE 2

High-end perimeter L7 security options

For detailed information about the security tools mentioned in this use case, please refer to Chapter 3.

02 AI data center use cases

Anonymous inference at scale

***Mission*: Protect public access to the AI frontends—typically public inference clusters.**

The Juniper SRX Series Firewall(s) attached to the frontend fabric (**Figure 3**) safeguards front ends that interact with users and systems by primarily implementing outside-to-inside security measures. This includes limiting L3/L4 access and potentially employing L7 application filtering, intrusion prevention systems (IPS), DoS/DDoS prevention, and network traffic logging.

Third-party security solutions would typically implement application-level security, such as WAF, setting boundaries for data structures to protect frontends and provide reporting. Then, a specialized AI solution could establish guardrails around the models in terms of data leakage, jailbreaking the AI models, etc.

To a certain extent, inside-to-outside security can also be implemented, typically to prevent and alert on suspicious activities such as post-exploitation callbacks (e.g., SecIntel, reverse shell detection, DNS security kits). It is generally not possible to separate tenants by network, as the resources are shared.

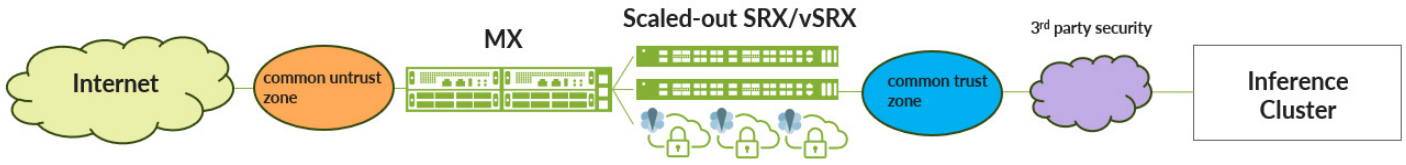


FIGURE 3

Public access to AI data center frontends

Our JVD for a scale-out firewall design enables a scalable pay-as-you-grow model that can scale to the highest performance and availability requirements. The design also provides extreme flexibility by allowing both virtual and physical SRX Series Firewalls to participate in the scale-out architecture.

02 AI data center use cases

Private training frontend

***Mission*: Segment multiple customers into virtual networks to enable secure and private model training frontends. These frontends typically provide shell access for customers to manage their training cluster via scripts.**

In this use case, the focus is on providing an additional authentication and authorization layer by using VPN technology, wrapping communication of users and systems into tunnels, and virtual LANs separating private AI workloads from other tenants and those publicly accessible.

The following breakdown of options for private frontends will cover multitenant VPN clients connecting to AI data center resources, modifications involving site-to-site VPNs (including third-party options), and the potential for end-to-end EVPN/VXLAN stretching using Juniper SRX VPN spokes and container SRX.

02 AI data center use cases

Private training frontend: Multitenant remote VPN access

***Mission*: Protect access to frontend portals/APIs for dial-up VPN clients.**

This scenario includes service access to the control planes and infrastructure of AI clusters. Additional technologies in use include Firewall/AppFW, IPS, application identification, SecIntel, remote access DoS/DDoS protections, and, potentially, SSL proxy.

The first option for multitenant access to AI data center resources (**Figure 4**) is the use of SRX Series Firewalls and Juniper Secure Client (JSC) remote access VPN, where tenants A and B have different portals for entering their credentials. In this more advanced scenario with EVPN/VXLAN Type 5, each tenant can unwrap their traffic in separate VRFs with distinct VXLAN VNIs due to different VPN profiles.

This setup permits scalable traffic steering throughout the entire infrastructure to tenant-specific resources. In this case, common resource access traffic could merge in a common VRF with policies prohibiting cross-VRF talk, except to protected resources. Authentication and authorization would utilize classical RADIUS, LDAP, PKI, or SAML.

At the leaf equipment connecting tenant-specific resources, the traffic would break out into VLANs, allowing third-party security solutions outside of the Juniper security framework described above to engage.

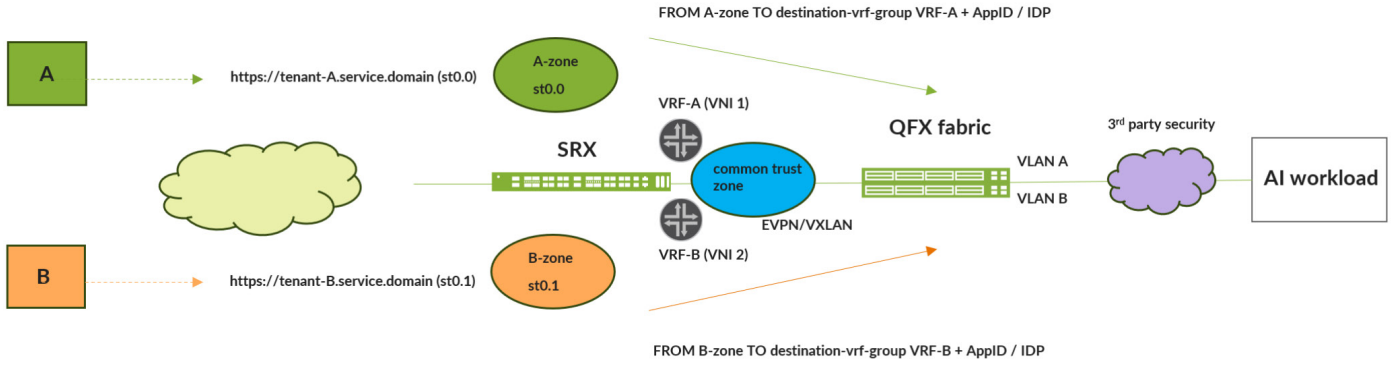


FIGURE 4

Multitenant remote access using JSC and EVPN/VXLAN Type 5

02 AI data center use cases

Private training frontend: Juniper end-to-end multitenant site-to-site IPSEC

***Mission*: Similar to the use cases above, here we protect access to AI frontends and infrastructure for users with broader applicability for AI resources, including API clients from remote locations.**

This use case functions without VPN level user authentication or authorization inherently coming with remote access VPNs. The technologies remain the same except for the removal of remote access and the addition of DoS/DDoS protections.

Instead of using the remote access client (JSC) on the customer's endpoints, a site-to-site IPSEC VPN is implemented (**Figure 5**). If the customer has their own remote access solution, remote traffic could effectively U-turn into the site-to-site tunnel. The advantage of this solution is that no additional client software needs to be installed on the endpoints, which is beneficial in cases where there are corporate policies, both formal and technical.

As this is an end-to-end solution with either physical or virtual Juniper SRX Series Firewalls located at customer premises, sub-multitenancy could be achieved by using EVPN/VXLAN Type 5 within the tunnels. With this approach using VLAN tagging (VLAN A1, VLAN A2), the separation of different admission levels could be decided by the customer placing traffic into corresponding VRFs. This effectively and transparently routes user and machine data to the resources within the AI data center.

The benefit of using EVPN/VXLAN with an IPSEC underlay is to avoid complexity with IPSEC when separation is needed into VRFs. Similar to the previous option, break out at the leaf level to VLAN would permit engagement of third-party security.

Regarding the high availability of such a solution, by using Juniper equipment end-to-end the SRX Series Firewall multi-node HA (MN-HA) capability of terminating two distinct tunnels with routing protocol within can be used for fast routing-protocol-driven failovers.

An alternative to this scenario (**Figure 6**) is the placement of private AI frontends into public cloud services for purposes of scaling, distribution, and agile deployment/development options. Virtual instances of Juniper SRX Series Firewalls can fulfill the role of the on-premises SRX.

In terms of frontend protections, the transport toward the on-premises location would be IPSEC and potentially also folding in EVPN/VXLAN for purposes of spawning multiple remote zones when needed. The on-premises SRX can serve as an additional security layer between them (for example, RAG security, management protection, etc.).

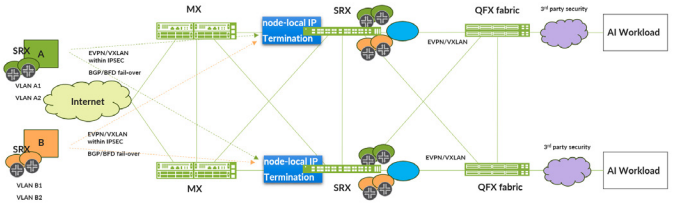


FIGURE 5

Multitenant access using IPSEC site-to-site VPN and EVPN/VXLAN towards fabric

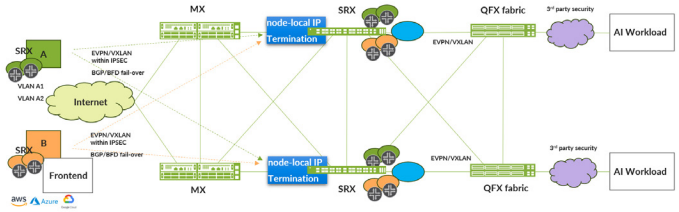


FIGURE 6

Alternative of the scenario where the frontend is placed in public cloud

02 AI data center use cases

Private training frontend: Juniper end-to-end multitenant site-to-site IPSEC with end-to-end VXLAN

***Mission:* Similar to the use cases above, here we protect access to AI frontends and infrastructure while preserving the EVPN/VXLAN fabric. This ensures the tunnel remains intact, increasing performance while preserving robust security.**

This third private training frontend option (**Figure 7**) involves an alteration of the previous methods with the same mission. The distinction being that the EVPN/VXLAN is brought directly to workloads terminating on a SRX Virtual Firewall (vSRX).

This is a very flexible design in terms of adding new tenant configurations independent of underlay fabric configuration where SRX security can still engage. This implies that potential third-party security solutions would need to either integrate with the EVPN/VXLAN fabric, act transparently, or can function as VNFs within the workload frontends.

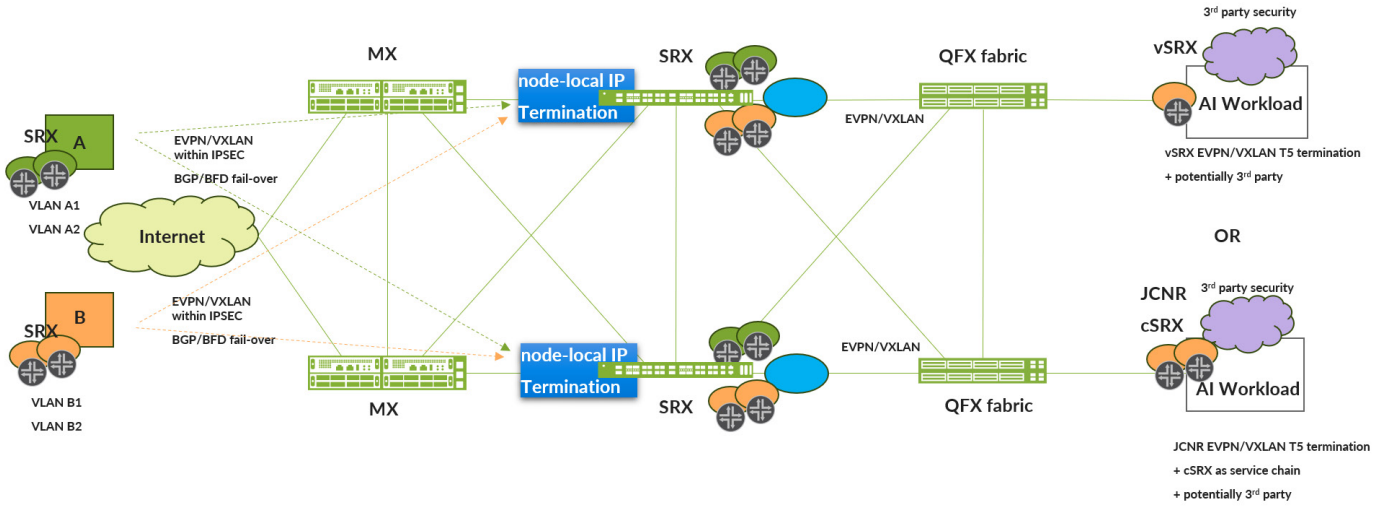
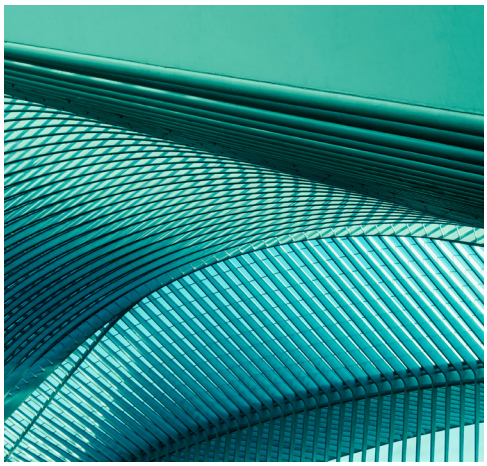


FIGURE 7

Multitenant access using IPSEC site-to-site VPN and end-to-end EVPN/VXLAN Type 5 transport

02 AI data center use cases

Private training frontend: Third-party IPSEC, multi-site



***Mission*: Secure off-site AI data center resources using high performance in-line VPN and firewall services.**

This option (**Figure 8**) involves using high-speed IPSEC tunnels to customers, potentially leveraging their own remote access infrastructure to redirect client traffic into a VPN. The tunnels are terminated on the Juniper SRX Series Firewall MN-HA loopback IP address, which represents the most interoperable approach to HA with network-driven failover. Customers are then assigned to their own VRFs with unique VXLAN VNIs and routed toward the fabric.

This scenario can also be expanded to include high-capacity site-to-site tunnels to remote outposts, utilizing ultra-high-speed tunneling technologies such as Juniper SRX fat-core or MX inline IPSEC. This use case could be the expansion of data centers and cost considerations (e.g., energy prices).

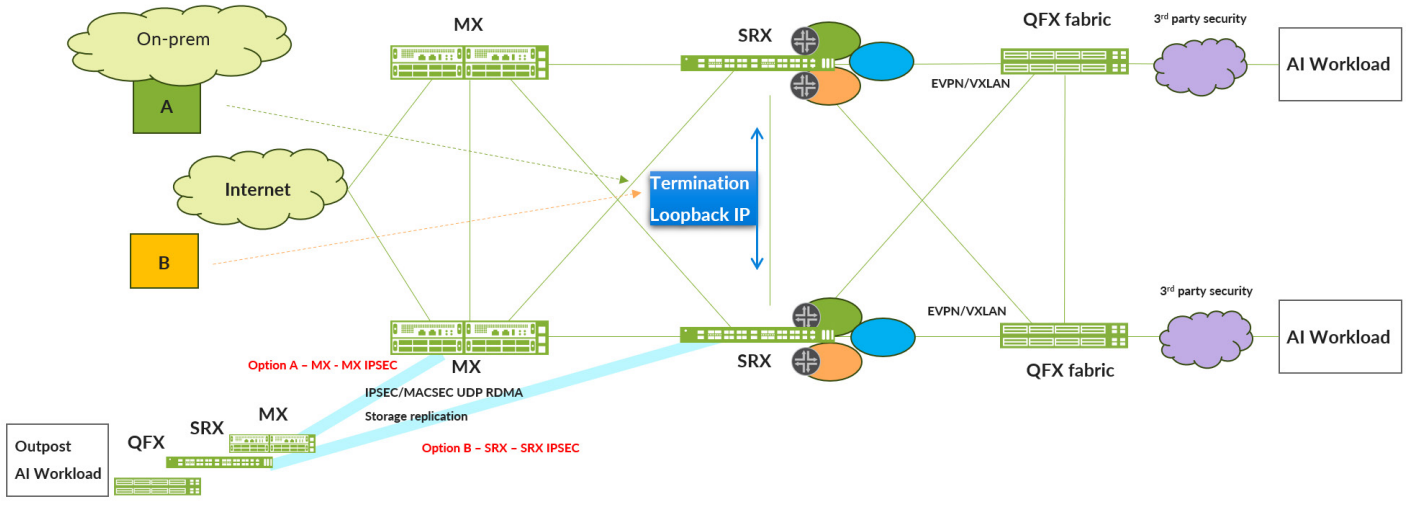


FIGURE 8

IPSEC with third-party and outposts

02 AI data center use cases

Private inference with RAG

***Mission*: Secure access to customers' Retrieval Augmented Generation (RAG) vector databases, whether local or remote. Additionally, consider attacks originating from hostile database services.**

Note: Any of the private training frontend designs mentioned previously can be used as a base and extended for this use case.

Juniper's SRX Series Firewalls can function as high-speed multitenant segmentation firewalls, isolating vector databases from other customers and the rest of the infrastructure. For instance, in the event of a single element compromise, it can create separate zones for each tenant to mitigate the risk of attack escalation within the AI data center's core infrastructure.

This can be achieved by permitting only service sockets and restricting access scope while also utilizing JSA to alert on any deviations from standard communication patterns. Indicators of a compromised database stack might include attempts to initiate callbacks outside of normal communication channels. Employing SecIntel, reverse shell detections, and a DNS security kit can help prevent these incidents.

In the case of known vulnerabilities, Intrusion Prevention System (IPS) functionality can protect the services until patched software is delivered. Additionally, vulnerable database client software can be safeguarded from rogue databases through the use of IPS and anomaly detection. For connectivity to off-site databases (such as those in the public cloud), IPSEC technology would be implemented and firewall rules—including advanced L7 security—could be enforced within the IPSEC tunnels.

With the insertion of the SRX Series Firewalls, implementing these features into the switch fabric is simplified with native EVPN VXLAN Type 5 support. This allows the segmentation and security functions to be easily separated between perimeter and internal firewalls, as shown in (**Figure 9**), or integrated into the perimeter SRX firewalls, if desired. Native fabric support in the SRX firewalls also provides reduced operational burden and enhanced automation capabilities to the network operations team.

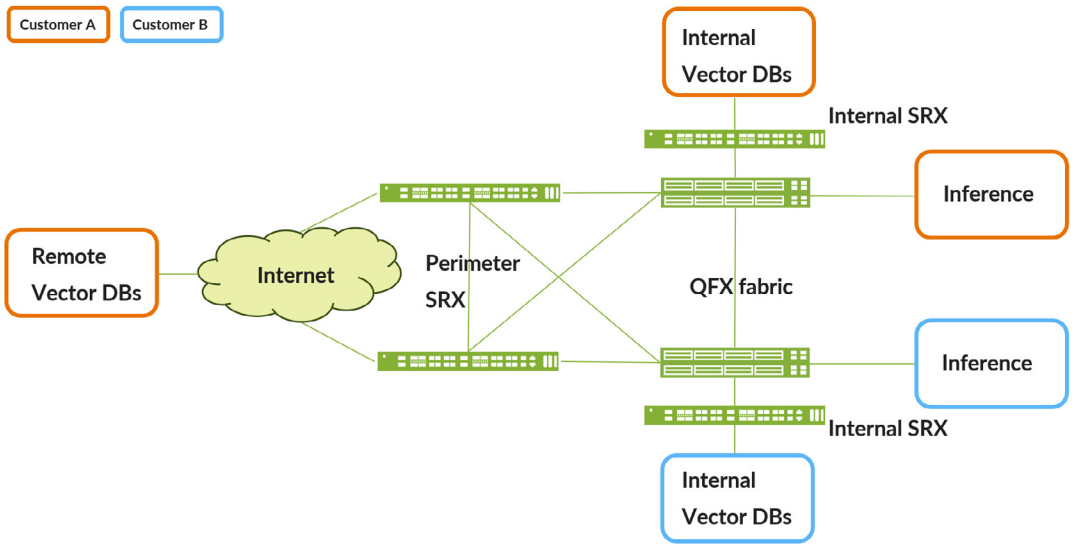


FIGURE 9

Segmentation/ protection of internal RAG services, secure connectivity to remote

02 AI data center use cases

Management protection

***Mission:* In the case of single element compromise, protecting management segments in AI data centers is a fundamental measure against attackers gaining access to all administrative interfaces of the infrastructure.**

For example (**Figure 10**), critical components include compute and networking equipment out of band (OOB) interfaces and data center infrastructure, like power control, cooling, and all central management systems. The ability to access those by attackers could eventually serve as a vector for taking control and installing ransomware and/or exfiltrating data, including loss of intellectual property in the form of private training data, models, and specialized hardware designs.

Best practice is to limit access to these interfaces by enforcing either local or remote authentication and authorization sources—even before accessing the administrative networks. In the case of multitenancy, Juniper SRX Series Firewalls can be used to separate tenant management networks and potentially even separate types of management resources for specific tenants. This would be achieved by configuring dedicated security zones on the SRX firewalls to protect the management interfaces while enabling the extensive L7 security functionality available on the platform.

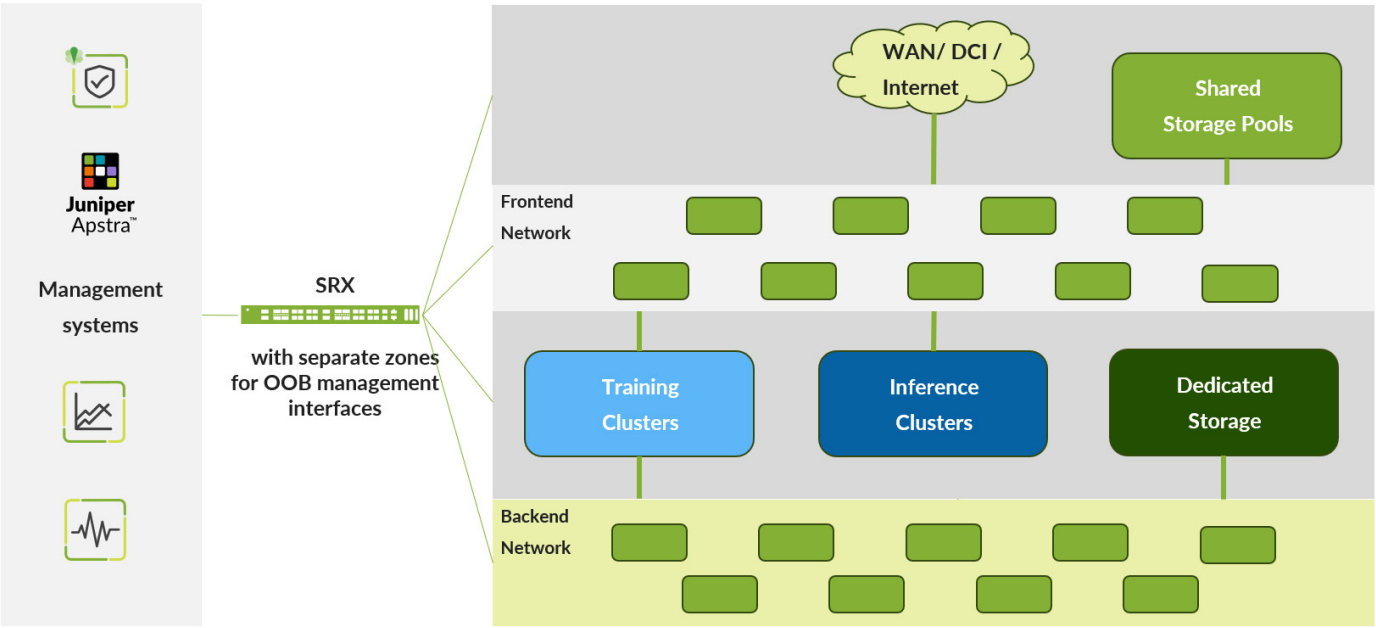


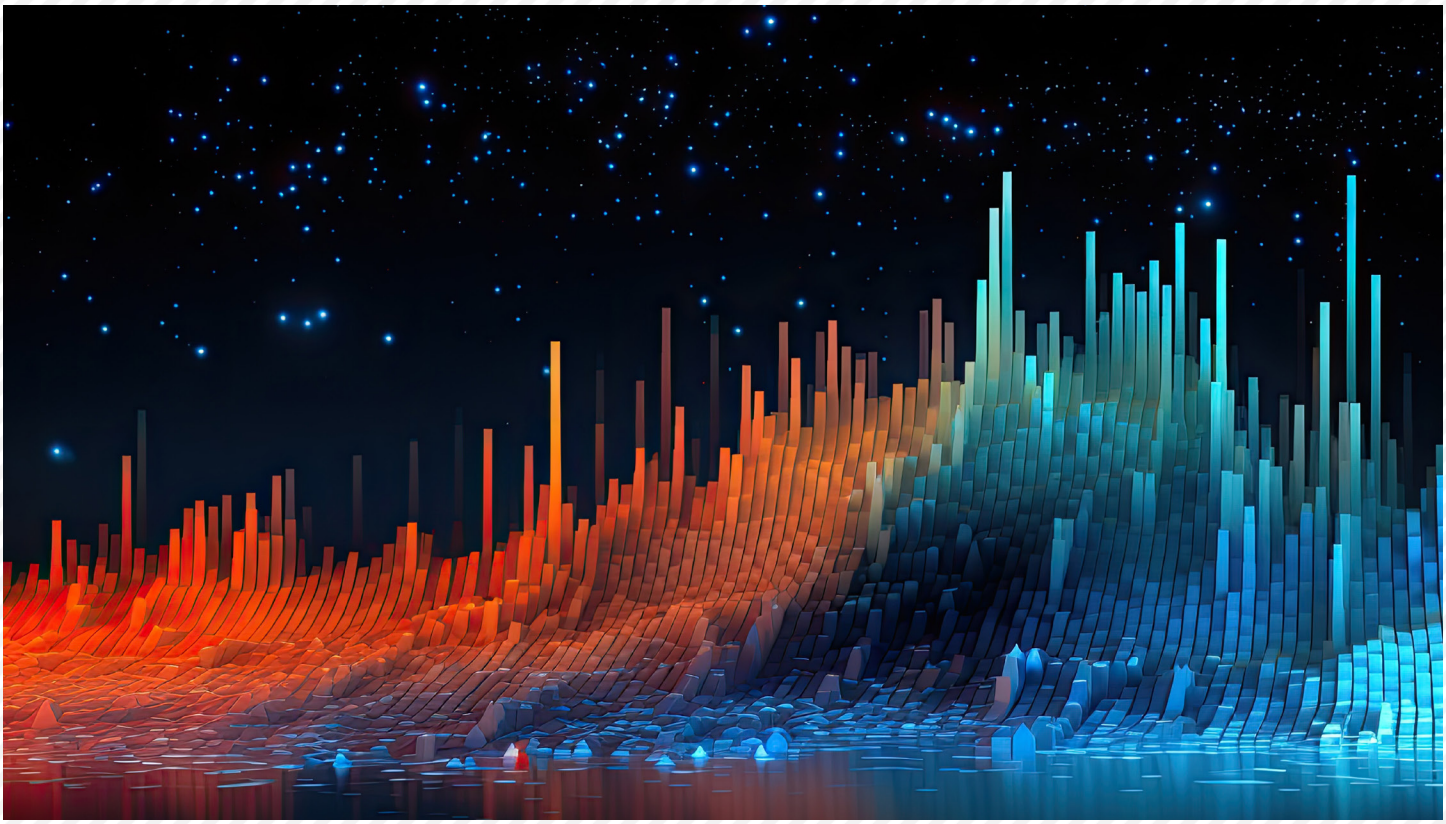
FIGURE 10

High-level OOB management networks zoning

03 Juniper security tools for AI data centers

Overview

This chapter discusses details of the SRX networking security feature kit from the perspective of traffic direction—incoming and outgoing—where some features naturally overlap and mitigate different subsets of attacks.



03 Juniper security tools for AI data centers

SRX security for outgoing traffic

For outgoing security (traffic from protected AI assets like frontend portals and inference clusters), the following SRX Series Firewall features can be used.

L4 / NAT firewall

A classical L4 / NAT firewall serves as a foundation for strict traffic flow control whenever possible. This can include GeoIP-based filtering to narrow access to at least the bare minimum by countries.

AppFW

As an extension to the L4 firewall, AppFW examines payloads and mitigates classical evasion techniques, such as running services on ports where they are not supposed to be, like various VPNs on ports typically associated with other services. AppFW can also simplify firewall deployment by avoiding strict destination control by IP prefixes. Instead, it can allow specific applications, such as OS updates, to access any CDN.

Juniper Advanced Threat Protection (ATP) Cloud and Anti-Malware

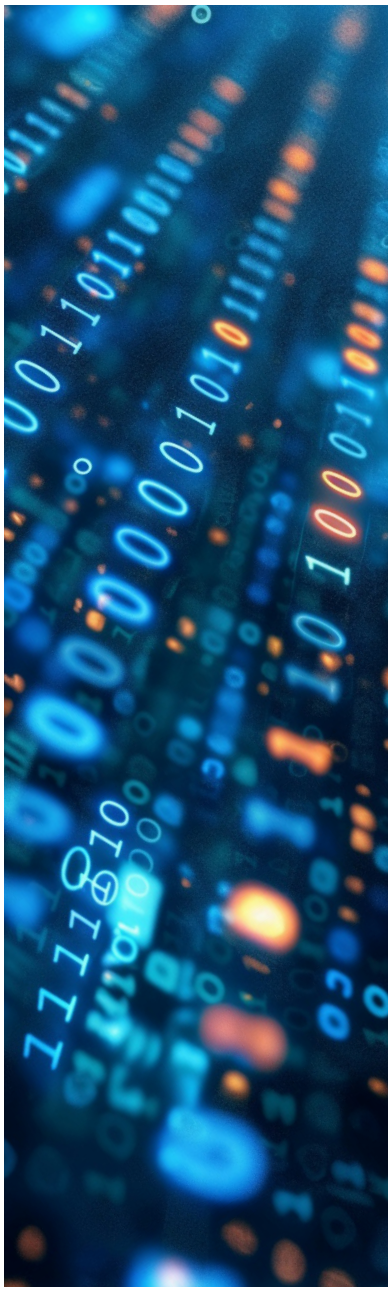
These features can analyze downloaded executables, libraries, and documents, including seemingly legitimate traffic such as system software updates that could be coming from hostile compromised sites. Multiple techniques, such as static analysis, multiple anti-malware scans, detonation in a sandbox, and the latest method of scanning executables via an on-box AI model, can be employed.

Web Filtering

Web Filtering is typically one of the first lines of defense, preventing systems and users from accessing anything outside of specific categories while also considering the reputation of the sites. This is done in conjunction with a cloud reputation service.

SecIntel

SecIntel is a reputation database of remote IP addresses and URLs, unlike web filtering, which also considers traffic beyond HTTP and HTTPS. Part of the SecIntel toolkit includes the infected feed, where suspicious hosts are listed, potentially preventing connections to external sites in an expedited manner.



DNS security

DNS security is the next important building block in network security, particularly against beaconing and command-and-control vectors. It analyzes requests to SecIntel-listed domains, provides DNS tunnelling detections, and the identification of Domain Generation Algorithms (DGA) traffic in conjunction with a cloud-based machine learning model.

IDP/IPS

IDP/IPS can protect client-side applications against known vulnerabilities using signatures and by detecting anomalies to thwart zero-day attacks. In addition, IPS can detect known malware communication based on signatures.

Encrypted Traffic Insights (ETI)

When it is not possible (or permitted) to use an SSL forward proxy to decrypt HTTPS traffic to the outside, Encrypted Traffic Insights (ETI) serves as a compromise between having no SSL security and maintaining security while analyzing communication using a cloud-based machine learning model for signs of malicious activity based on metadata and behavior.

SRX IoT detection

SRX IoT detection can categorize endpoints inside the data center based on their behavior by type of device, host operating systems, and version, with a capability to create IPFilter groups for applying in SRX firewall policy (e.g., Linux systems may have different protections than Windows).

Explicit Proxy

Explicit Proxy on the SRX can serve mostly all the above without the systems having a default gateway simply by having explicit proxy defined. Typically, Explicit Proxy is not used for caching purposes, but rather to isolate the systems from direct IP access.

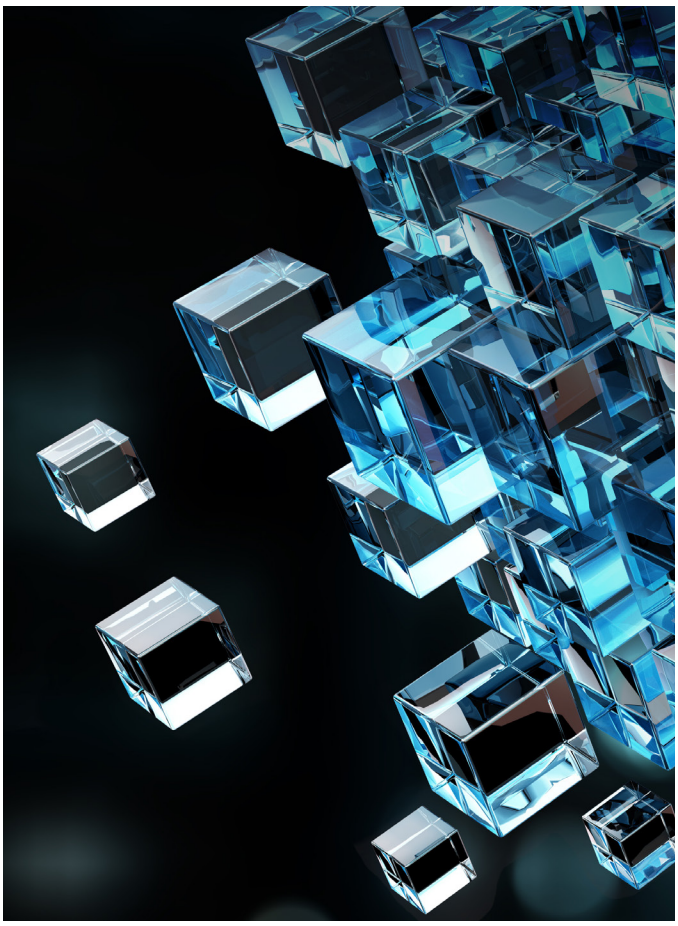
Adaptive Threat Profiling

Adaptive Threat Profiling, in conjunction with ATP Cloud, creates lists of endpoints based on certain events (e.g., triggering certain IPS rules, using certain L7 applications, or visiting specific web site categories). Lists can be used in policies to step up security controls.

Reverse shell detection

Part of the ATP Cloud offering is reverse shell detection, which identifies suspicious traffic flows hiding within interactive terminal sessions.

Please see the Data Center Next-Generation Firewall Use Case JVD for more information.



03 Juniper security tools for AI data centers

SRX security for incoming traffic

For incoming traffic security (traffic to protected AI data center assets), the following features can be used.

Firewall, AppID, GeoIP, and IPS

Firewall, AppID, GeoIP, and IPS serve similar roles as in the outgoing traffic protection, albeit in different modes. IPS focuses on server-side protection through signatures rather than client-side protection, while GeoIP is used to match and enforce the source of the traffic.

SecIntel

The SecIntel kit can also be used to match sources from the internet to protected assets, prohibiting traffic from known attackers.

Anti-malware

Anti-malware solutions can protect against threat vectors such as vendors uploading malicious system updates.

AI Predictive Threat Protection

The AI Predictive Threat Protection engine can complement the anti-malware for more thorough executable file inspection.

IPSEC

As outlined in the preface, IPSEC is a fundamental part of transit security for both user and system connections. Layer 7 security measures can also be used within the tunnels. This further reduces the attack surface only to trusted endpoints, rather than exposing a wide range of services to the outside.

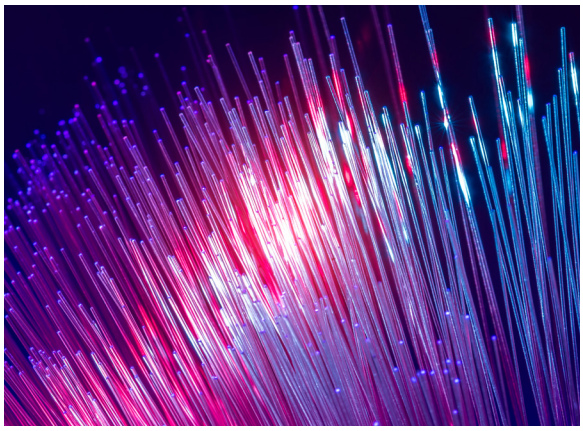
Reverse SSL proxy

Reverse SSL proxy enables the SRX firewall to decrypt, inspect, and reinitiate encryption toward backends. This dramatically increases the depth of possible inspections beyond SSL library attacks and IP stack protections available when there is not an option to decrypt SSL traffic.

Screening

Finally, the Screening feature set of the SRX firewall is a fundamental measure against network-level DoS/DDoS attacks. On high-end SRX platforms, screening is primarily implemented in hardware for enhanced performance and platform resiliency.

Please see the Data Center Next-Generation Firewall Use Case JVD for more information.



03 Juniper security tools for AI data centers

DDoS protection with Threat Defense Director

Threat Defense Director (TDD) can be used in conjunction with a Juniper MX or PTX router to provide volumetric DDoS mitigation for the AI data center (**Figure 11**). While the data center is likely to be multitenant (TDD offers specific multitenant portal functionality), the mitigation would be enforced at a global level (rather than per tenant) since the public IP presented to tenant customers to access their own backend resources would be shared.

While TDD provides first-line volumetric defense (**Figure 12**), it would not necessarily replace the Screen function on the SRX firewall behind it since Screens offer different (non-volumetric) defenses and mitigations (such as session limitations to prevent resource starvation attacks).

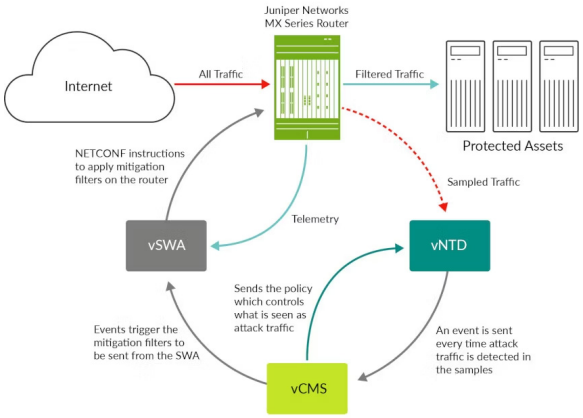


FIGURE 11

Juniper TDD DDoS protection solution

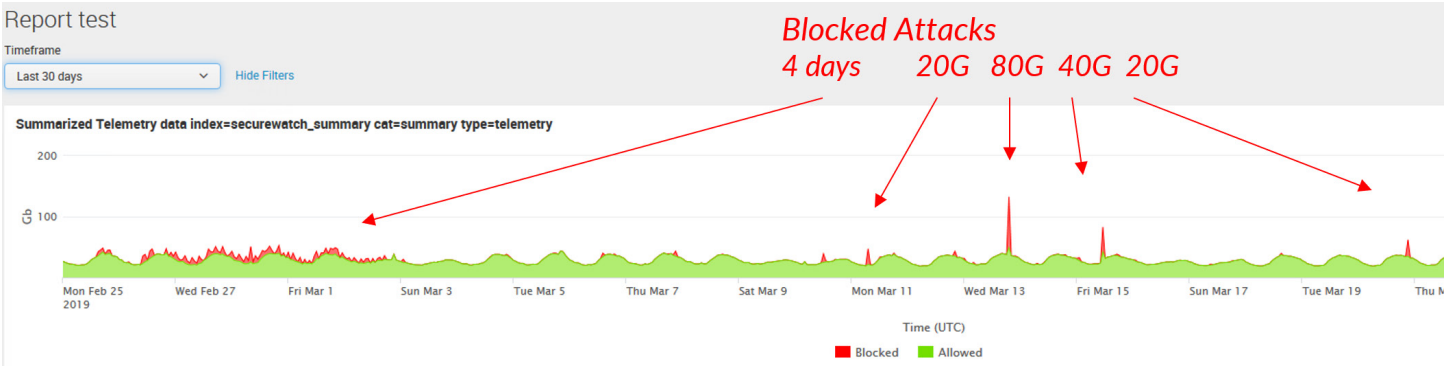


FIGURE 12

Juniper TDD in action blocking volumetric attacks

For more information, see the Juniper-Corero DDoS Protection Implementation Service datasheet.

03 Juniper security tools for AI data centers

Apstra Flow Data

Apstra Flow Data is a comprehensive, scalable, high-performance flow collector and analyzer.

This tool includes an array of network flow analytics that helps network operators understand application performance and usage across Apstra-managed networks. Flow data features out-of-the-box dashboards and advanced analytics capabilities, such as fine-grain filtering, customizable charts, and drill-down dashboards (**Figure 13**).

Apstra Flow Data provides insights into how applications utilize network resources and impact users. This tool also helps troubleshoot network issues, such as application performance issues, heavy traffic utilization, and anomalous behaviors. Flow Data works by collecting and analyzing flow data from network devices and provides rich visualizations of the network traffic. The flow data and visualizations provide a clearer understanding of what's happening on the network from an application and user perspective.

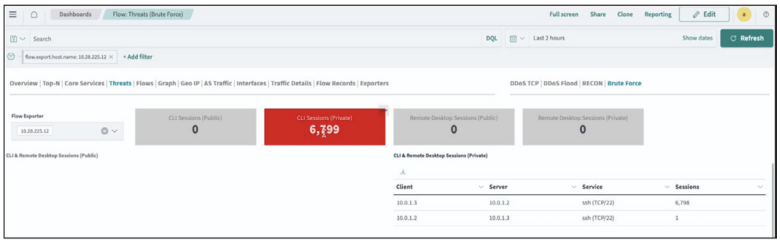


FIGURE 14

Apstra Flow Data threats dashboard

Apstra Flow Data can perform basic threat detection for flows. The threats dashboard shows any DDoS, port scans, and brute force attempts on your network. (**Figure 14**) shows an example of repeated SSH sessions that were sent between hosts. Here, Flow Data displays these sessions as brute-force attempts. Flow Data can also enrich the data that is displayed with DNS and IP geolocation.

For more information on Apstra Flow Data, see the Apstra datasheet.

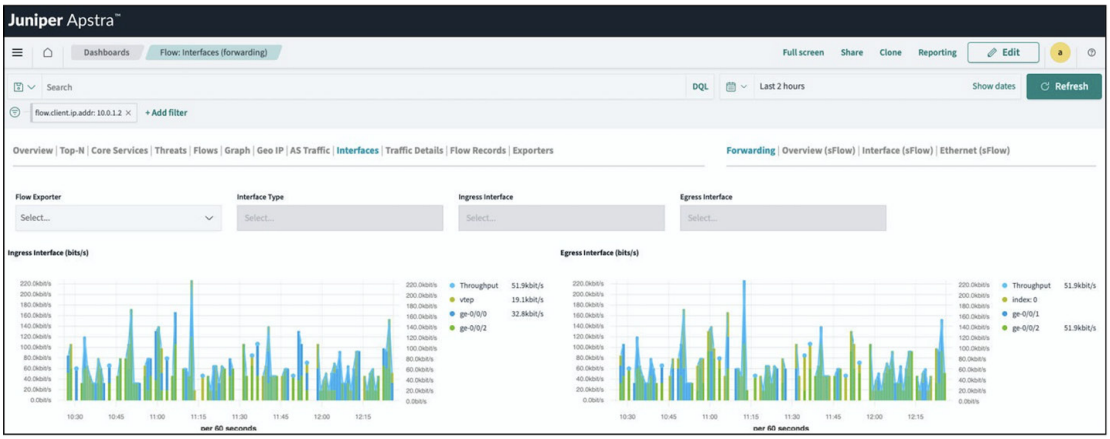


FIGURE 13

Apstra Flow Data interfaces dashboard

03 Juniper security tools for AI data centers

Juniper Secure Analytics Security Suite (JSASEC)

Juniper Secure Analytics (JSA) and the extended product suite (JSASEC) can offer SIEM, SOAR, and XDR functionality to detect possible threats and/or anomalous behavior.

These could be inbound threats from external networks, between tenants, or within a tenant—the latter two potentially acting as an indicator of a compromised host.

Not only can JSASEC detect threats (**Figure 15**), but its SOAR component can assist Security Operations Centers (SOCs) with incident forensics and mitigation by the use of response playbooks (among other tools), giving security operators a clear step-by-step path to resolution (**Figure 16**).

One aspect to consider is that communication in AI data centers can be of a more "static" form than in a traditional data center. That said, communication matrices, showing which other hosts any single host or network talks to, should be relatively fixed. Perhaps inbound connections to the exposed frontend, frontend to one of the hosts in the tenant inference cluster, and inference cluster to storage and/or RAG. This baseline matrix can make it very simple for JSA to detect anomalous flows, such as those generated when a compromised host attempts to propagate across a data center or communicate out to a Command and Control (C2) server.

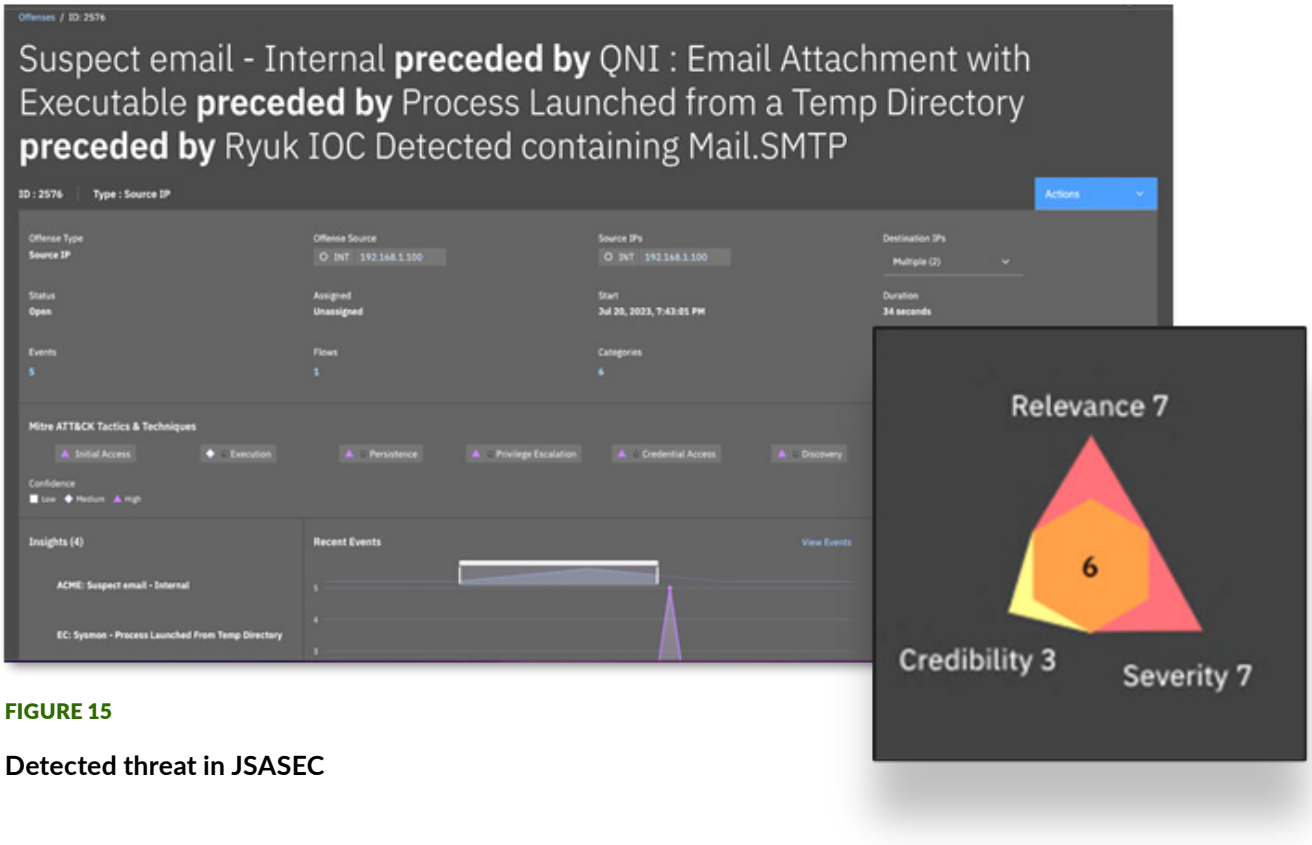


FIGURE 15

Detected threat in JSASEC

Some recommended uses for JSA are:

- Building “known” communication matrixes or enabling “new IP pair” correlation rules that would detect unusual or unexpected flows
- Detecting any outbound-initiated connections from internal hosts that would not normally communicate in this direction, such as inference clusters or management systems. Specific IP addresses could be whitelisted from the correlation rules used
- Network Behavior Anomaly Detection (NBAD) rules, based on network attributes such as throughput between hosts, subnets, or networks, which would profile traffic across the AI data center for a time and then produce alerts/offenses if expected thresholds are significantly exceeded
- Use of onboard JSA threat feeds, especially for outbound C2 detection
- Detection of more “traditional” threats, such as slow scans and sweeps, stealthy DDoS (resource starvation attacks), and correlation rules based on detections coming from the SRX firewalls running advanced services. For example, if an IDS log is received from an SRX firewall with target as host X, and that host then communicates to C2 server Y, a critical severity offense can be raised, as the likelihood of compromise is high

JSA is only as good as the data it ingests. So, if possible, it is recommended to receive both logs from the SRX Series Firewalls and flow data from switches and routers in the AI data center. Since JSA correlation rules are log-based, flow-based, and hybrid, collecting both allows the greatest possible breadth of detection.

One final recommendation would be to proactively scan the frontend systems, as well as the inference clusters for vulnerabilities (via a third-party scanner) and ingest the results in JSA. These could include web or general OS vulnerabilities. Not only can JSA incorporate vulnerability scan results into correlation rules (e.g., a ‘detected exploit’ && ‘target=vulnerable’ rule generates a critical severity offense), but it also has virtual patching functionality.

For more information, see the JSA datasheet.

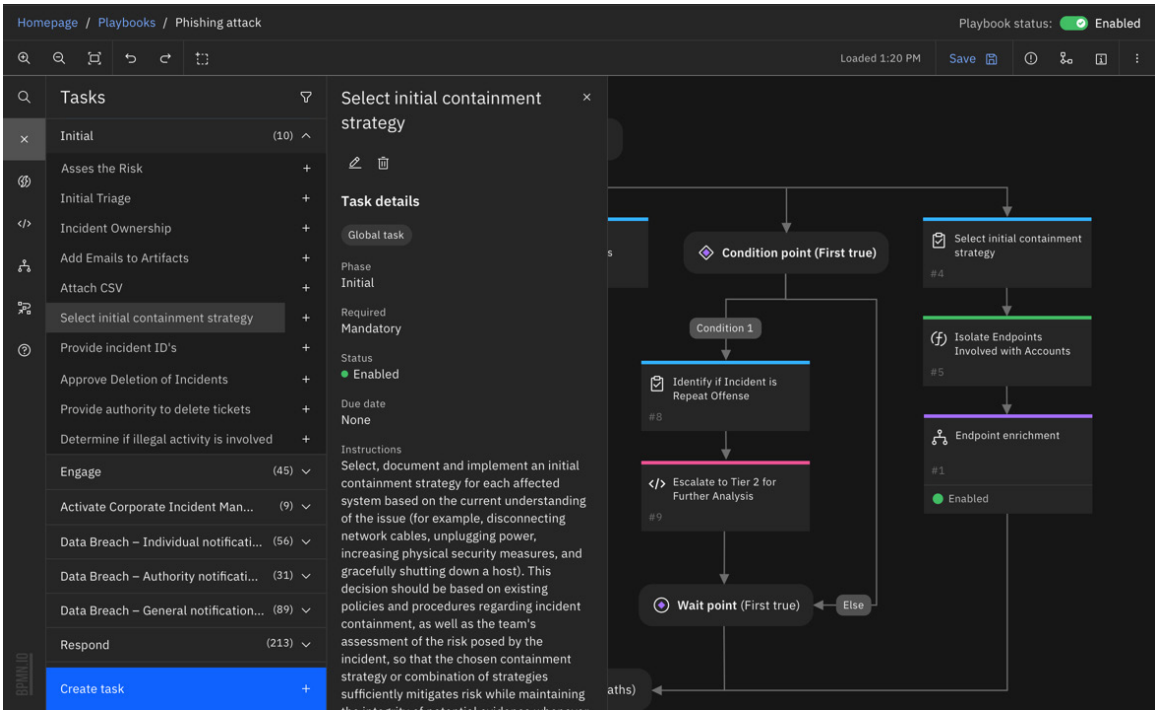


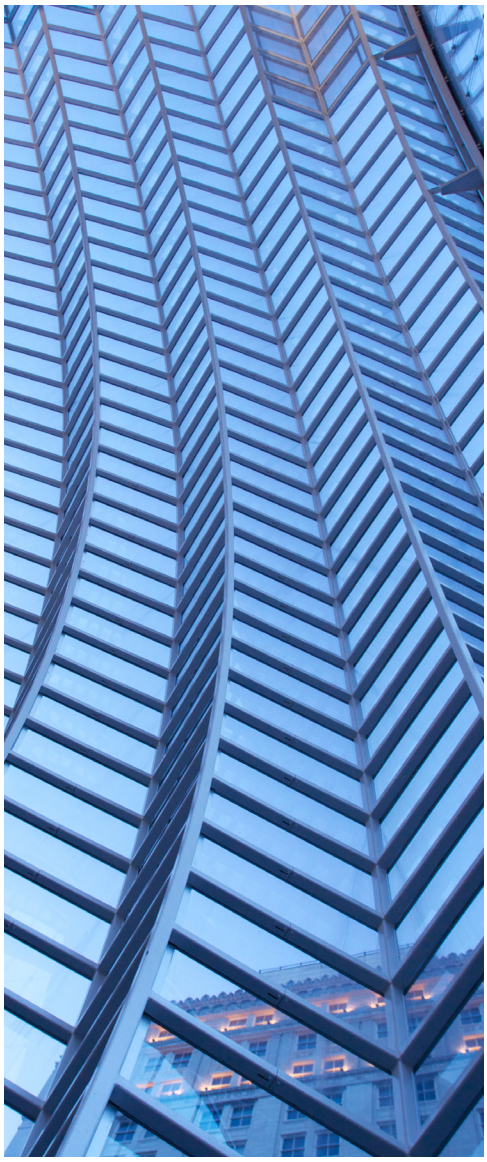
FIGURE 16

JSASEC: SOAR playbook in action

04 Conclusion

Conclusion

The leader in networking and security for AI data centers



With a fully integrated infrastructure, security, management, and automation stack, Juniper's AI service provider customers are enabled with the tools to build robust, performant, and secure AI data centers.

Juniper can provide a secure end-to-end AI data center solution comprised of the following components:

- **Routing:**

  MX and PTX routers for high-speed and large-scale BGP internetworking running on the solid JUNOS foundation used by Service Providers and the largest global networks for decades

- **Switching:**

  QFX switches automated with Apstra for ease of building and maintaining scalable switching fabrics. Juniper switching fabrics maintain the performance required for large-scale AI training GPU clusters for the best possible training performance. Apstra also provides enhanced visibility into AI data center traffic, uncovering potential security risks and scaling bottlenecks

- **Security:**

  SRX firewalls managed by Security Director and extended with APT Cloud provide industry-leading security efficacy levels at the highest performance in the industry. SRX firewalls also run on JUNOS, providing a clean and robust networking experience at scale

For additional information, please contact your Juniper sales representative who can bring in the security for AI subject matter experts to consult with your team.

05 Glossary of terms

Glossary of terms

AI: Artificial Intelligence

ATP Cloud: Juniper Advanced Threat Protection Cloud

C2: Command and Control

CDN: Content Delivery Network

DNS: Domain Name System

DOS: Denial of Service

DDOS: Distributed Denial of Service

ETI: Encrypted Traffic Insights

EVPN: Ethernet Virtual Private Network

GPU: Graphics Processing Unit

HA: High Availability

HTTP: Hypertext Transfer Protocol

HTTPS: Hypertext Transfer Protocol Secure

IDS: Intrusion Detection System

IOT: Internet of Things

IP: Internet Protocol

IPS: Intrusion Prevention System

IPSEC: Internet Protocol Security

JSA: Juniper Secure Analytics

JSASEC: Juniper Secure Analytics Security Suite

JUNOS: Juniper Operating System

JVD: Juniper Validated Design

LDAP: Lightweight Directory Access Protocol

MACSec: Media Access Control Security

MN-HA: Multi-Node High Availability

NAT: Network Address Translation

OOB: Out of Band

OS: Operating System

PKI: Public Key Infrastructure

RADIUS: Remote Authentication Dial-In User Service

RAG: Retrieval Augmented Generation

SAML: Security Assertion Markup Language

SIEM: Security Information and Event Management

SOAR: Security Orchestration, Automation, and Response

SSL: Secure Sockets Layer

TDD: Threat Defense Director

TLS: Transport Layer Security

URL: Uniform Resource Locator

VLAN: Virtual Local Area Network

VPN: Virtual Private Network

VNF: Virtual Network Functions

VNI: Virtual Network Identifier

VRF: Virtual Routing and Forwarding

VXLAN: Virtual Extensible Local Area Network

WAF: Web Application Firewall

XDR: Extended Detection and Response

Why Juniper

Juniper Networks believes that connectivity is not the same as experiencing a great connection. Mist®, Juniper's AI-native networking platform, is built from the ground up to leverage AI to deliver exceptional, highly secure, and sustainable user experiences, from the edge to the data center and cloud. Additional information can be found at Juniper Networks (www.juniper.net) or connect with Juniper on X (Twitter), LinkedIn, and Facebook.

About Author(s)

The majority of this paper was drafted by the Juniper CSEC SE Specialist team, including Karel Hendrych and Steven Jacques. Special thanks to the many contributors from various departments, including field SEs, data center SMEs, and PLM.

Useful links

- Juniper Validated Designs (JVDs) for Security
- AIDC Network with Juniper Apstra, NVIDIA GPUs, and WEKA Storage JVD
- Juniper SRX Series Firewalls
- Juniper vSRX Virtual Firewalls
- Juniper Security Director
- Juniper SecIntel
- Juniper Advanced Threat Protection
- AI-Predictive Threat Prevention
- Corero SmartWall Threat Defense Director (TDD)
- Juniper Secure Analytics

Take the next step

www.juniper.net

Juniper Networks Inc.
1133 Innovation Way
Sunnyvale, CA 94089

2000832-001-EN May 2025