SOLUTION ARCHITECTURE

The Zero Trust Evolution

A Practical Guide of the First Steps

Contents

Executive Summary

A new security model, Zero Trust, addresses the growing security challenges of the modern workplace. As the name would suggest, the principle underlying Zero Trust is that no single element across a data flow is inherently trusted all the time. Instead, a constant test of security posture and protection against attack is built into every workload, server, network connection, and endpoint for every user within a company.

We have architected and built out a working example of how to obtain a Zero Trust infrastructure—not only for endpoint users, but also in the data center. Using a unique combination of VMware products, this architecture provides a starting point on the path to understanding Zero Trust, and the concepts can scale out as a Zero Trust infrastructure expands.

To further simplify and deliver end-to-end Zero Trust architectures, we at VMware continue to evolve our cloud solutions, designing with Zero Trust in mind from the ground up.

Overview

All security models are built on assumptions. For example, traditional enterprise cyber security was built on the assumption that all systems within a corporation—from the workload to the end user—were known and trusted, while everything else outside the corporation was less known and less trusted.

A useful analogy is that of a medieval castle with high walls and a moat; everything inside the walls is trusted, and outside is less so. Similarly, traditional enterprise security once consisted of a single fenced-in perimeter comprising multiple layers, both physical—such as badges that controlled entry through the doors of an office building—and virtual—including firewalls and other cyber controls which housed and protected a private corporate computer network. This castle-and-moat, perimeter-based security model made sense when most computer systems remained on the corporate premises and connected only to the corporate network. But even before the explosive growth of mobile computing and cloud-hosted applications, such a model had clear flaws—ones which criminals and state-sponsored actors exploited all too frequently.

Cyber security has always evolved as new threats emerge. The COVID-19 pandemic, however, accelerated the need to address certain vulnerabilities, given the shift to employees working remotely. Moreover, businesses increasingly rely on multi-cloud workloads they can access on demand, such as infrastructure as a service, platform as a service, and software as a service (SaaS). Taken together, these factors create an urgent need for us to re-examine the basic assumptions underpinning the enterprise cyber security model. A new approach to both workplace and workload security is needed.

Enter Zero Trust.

What is Zero Trust?

In the simplest terms, Zero Trust means “trust no device and trust no user.” Access is constantly reevaluated for every user and system, and all devices and user identities undergo multi-factor verification. But Zero Trust goes even further, by also monitoring how a device and user act on a network. For instance, from what location is a user or device accessing data? And is accessing this data considered normal behavior for that user or device? Only when the answers to these questions are acceptable isl the minimum needed access granted for that specific activity.

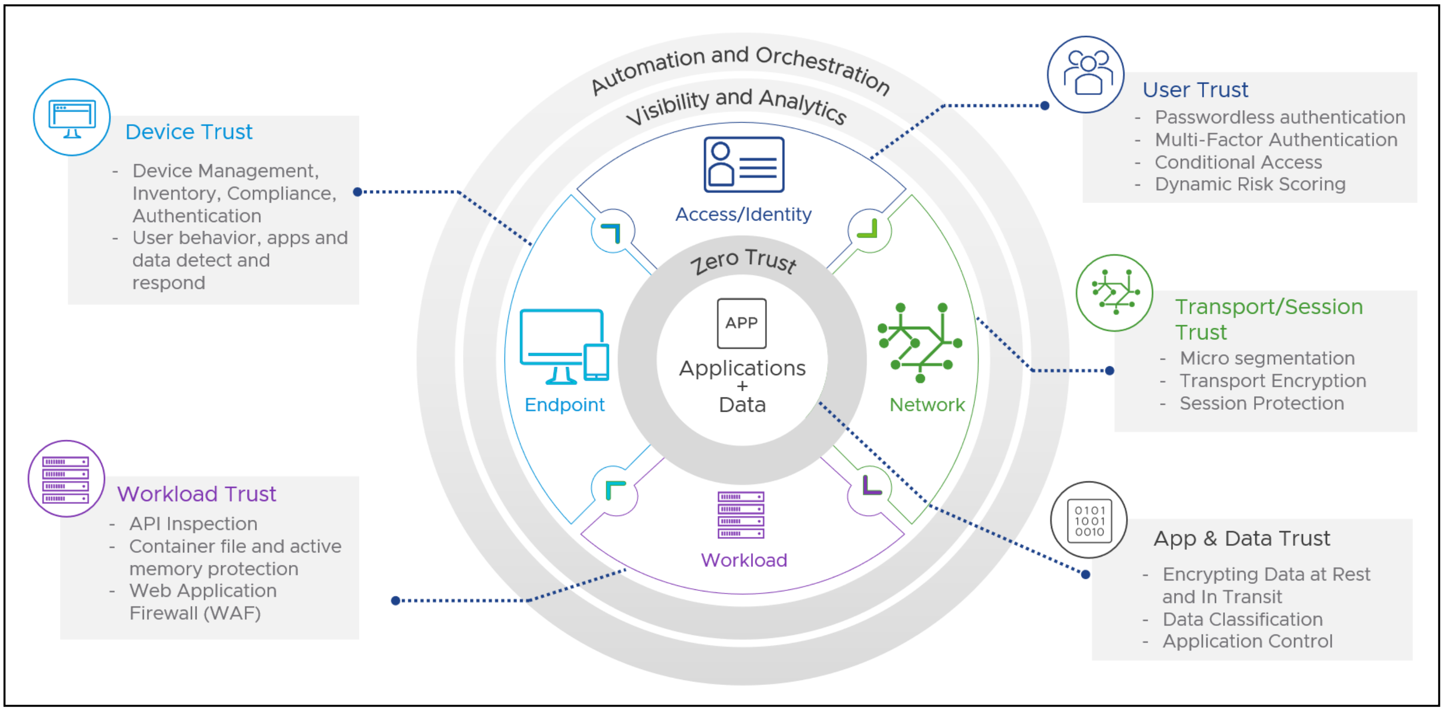
Traditional Perimeter Security

Traditional security models simply fence in devices and users. The model assumes that if you have been granted access inside the network then you can roam around freely. You may be locked out of certain sections, but for the most part, after you are checked at the door and allowed to enter, you have access to most systems on the network. Historically, administrators have tried to control access within networks in a variety of ways, using access control lists (ACLs), virtual local area networks (VLANs), private VLANs, and internal firewalls to create zones of separation from outside networks. These strategies, along with physical security to ensure that only authorized people can access areas housing physical devices and network ports, do increase internal network security, but they still leave many holes. Managing devices in a constantly changing environment with newly emergent threats puts a heavy burden on IT resources.

Zero Trust Security

With Zero Trust, we look at systems that we have allowed to enter through the perimeter door in exactly the same way we look at systems that are outside waiting to come in. Our assumption in both cases is the same: a system is inherently untrusted until proven otherwise. And, once people and devices have been granted access to the network, they no longer have the freedom to roam around and explore.

Even in the data center, workloads that interact with each other do not implicitly trust each other. Each transaction or work request goes through not only an authentication validation process, but also device validation, access management, and analysis of the behavior of the executing processes. Together, these evaluations provide a real-time examination of the security posture—that is, the trust state—of the system. If any one of these evaluations indicate a potential compromise or unacceptable risk level, then additional revalidation steps may trigger, and further access may be limited or denied depending on network policies. The more stringent the policy, the more stringent the validation action. In addition, all transactions undergo audit so that a clear trail of all activity on the network is kept.



Device Trust
- Device Management, Inventory, Compliance, Authentication
- User behavior, apps and data detect and respond

Workload Trust
- API Inspection
- Container file and active memory protection
- Web Application Firewall (WAF)

User Trust
- Passwordless authentication
- Multi-Factor Authentication
- Conditional Access
- Dynamic Risk Scoring

Transport/Session Trust
- Micro segmentation
- Transport Encryption
- Session Protection

App & Data Trust
- Encrypting Data at Rest and In Transit
- Data Classification
- Application Control

Architecting to Zero Trust Principles

VMware enables software-defined networking (SDN) with the NSX family of products, powerful endpoint management with Workspace ONE, and automation with the AI-driven endpoint and workload platform Carbon Black. These three solutions, combined with vSphere's abilities—to encrypt workload virtual machines (VMs), validate infrastructure against hardware with trusted platform modules (TPMs), and authorize workload trust through vSphere Trust Authority—set a strong foundation for building a modern infrastructure that aligns with Zero Trust principles.

We have architected and built out a working example of how to combine these products to obtain a Zero Trust infrastructure—not only for endpoint users, but also in the data center. This architecture provides a starting point on the path to understanding Zero Trust, and the concepts can scale out as a Zero Trust infrastructure expands.

Components

SOFTWARE

Software	Version
vSphere	7.0 U2
NSX-T	3.1
NSX Advanced Load Balancer (Avi)	20.1.5
Unified Access Gateway (UAG)	3.1.3
Workspace ONE Access Connector	20.10

HARDWARE

Server	Cluster
Dell vSAN Ready Nodes	Zero Trust Workload
Dell vSAN Ready Nodes	Control Plane
Dell vSAN Ready Nodes	“Trust Authority”

SAAS

VMware Cloud Services
VMware Workspace ONE UEM
VMware Workspace ONE Access
Carbon Black Cloud
VMware Horizon
VMware Unified Access Management

Architecture

Endpoint Zero Trust

A VMware Carbon Black Cloud sensor provides protection against known and unknown malware and living off the land attacks; visibility of software vulnerabilities; threat hunting and remote response to investigate potential attacks; and provision of additional threat telemetry to VMware Workspace ONE. VMware Workspace ONE Unified Endpoint Management (UEM) not only takes care of domain level authentication, but also checks endpoint compliance to software, network, and user access policies. Workspace ONE Access enforces what application a user can access within your organization, while Workspace ONE UEM makes sure the endpoint is compliant with corporate standards with regards to the software build and current patch level. Together, the two provide multi-factor authentication. Meanwhile, Workspace ONE Intelligence provides risk signals and automation of further response actions, and can leverage signals from Carbon Black Cloud as well as third party applications.

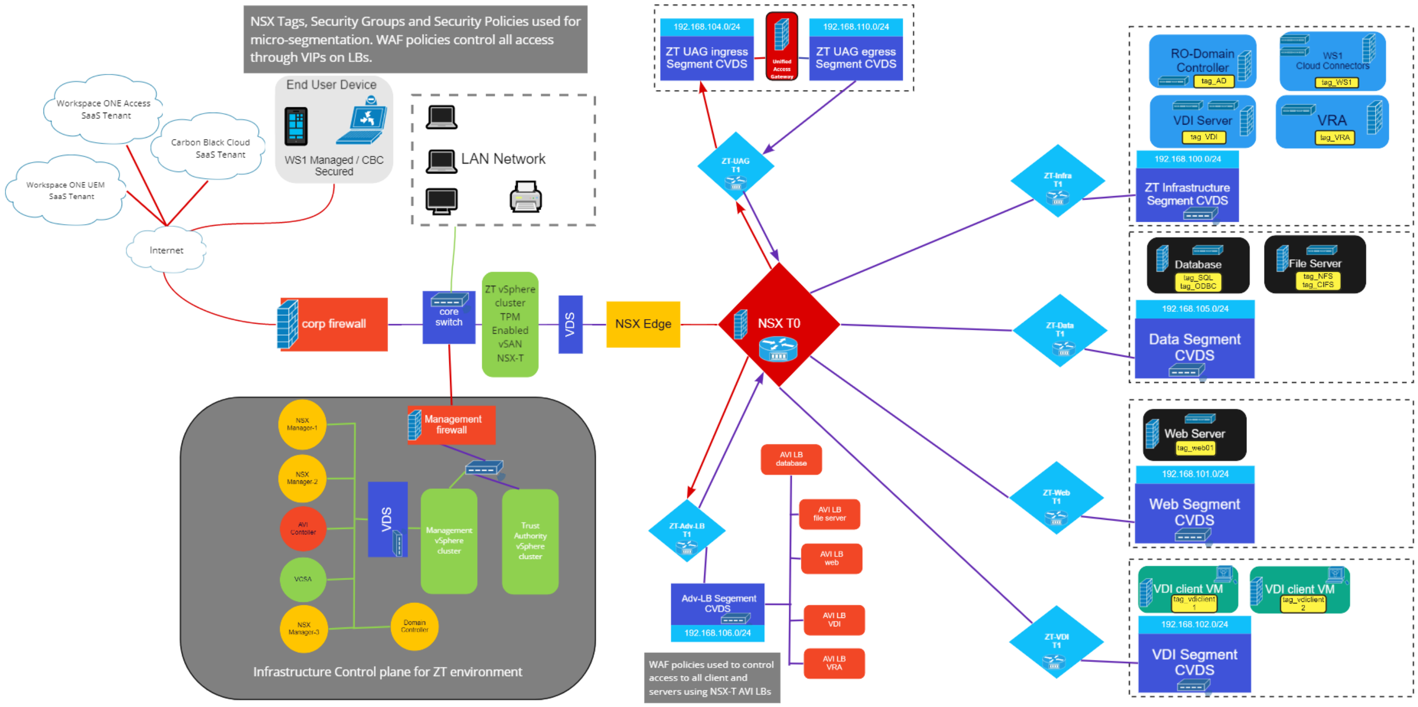
Data Center Zero Trust

To secure the data center with Zero Trust, we move to an SDN model. Using VMware NSX, we easily segment the data center and apply micro-segmentation rules down to the application level. VMware vSphere provides the cloud infrastructure and ensures hardware identity and compliance for workloads.

Zero Trust Control Plane

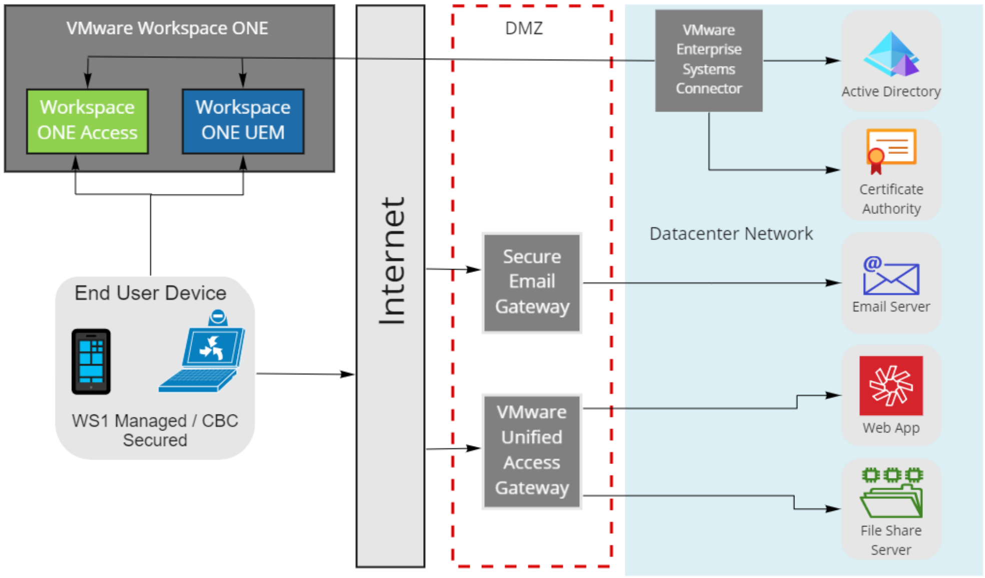
The Zero Trust control plane is separated from the Zero Trust workload environment. For the on-premises data center, we deploy a vSphere cluster hosting the separate control software stacks to manage the workload environment. The endpoints are managed using a SaaS instance of Workspace ONE, running in the public cloud with internal connectors for authentication with Active Directory (AD).

The following diagram offers an overview of the entire end-to-end Zero Trust workflow. We will discuss each section of this diagram in more detail and describe how each component joins in the workflow.



Secure User Access

The next diagram depicts a typical Workspace ONE environment. This represents how an endpoint is secured to the app using endpoint Zero Trust. While the user and endpoint are being checked for compliance, the internal data center in this example is still just fenced in—a more traditional network approach. This setup does provide a Zero Trust connection from the endpoint to the destination server app, but it does not apply Zero Trust principles among the back-end servers in the data center. While security using conditional access and compliance policies will address one of the most vulnerable connections in the workload chain, there is still a vulnerable security gap within the data center.



Secure Workload Access

In the data center, we apply NSX to create a micro-segmented network topology using security tags. These security tags are simply labels that we associate with particular virtual machines. Multiple security tags can identify workloads. The matching criteria of a Security Group can be a security tag, and a tagged workload can be automatically placed into a Security Group. We can add or remove a VM's security tags dynamically in response to various criteria, such as the results of malware or vulnerability scans, or alerts from intrusion prevention systems. ecurity tags can also be added and removed by VMware Carbon Black Workload Protection (For a deeper dive into security tags see: *https://docs.vmware.com/en/VMware-NSX-Data-Center-for-vSphere/6.4/com.vmware.nsx.admin.doc/GUID-0D5905B1-44EA-489D-8DF5-9208AB5E3754.html*)

By using NSX security tags, we filter traffic between hosts at the network layer—even on the same network segment. This is not a typical network ACL that decides access based on an IP, MAC filter to a port, or other IP address. Rather, we've created layer 7 policy-based rules on traffic from server A to server B, allowing or denying conditional access based on criteria of the packets being sent. NSX dynamically looks at this traffic and identifies its unique characteristics—a kind of fingerprint. Regardless of which port the traffic uses, its fingerprint determines which firewall rules apply to it.

The context-aware firewall in NSX allows rules to be based on specific characteristics, operating systems, and application identifiers. It also applies to the identity-based firewall filtering traffic based on a user's AD group, deployed in conjunction with VMware Horizon or an equivalent desktop VDI solution for virtual desktop infrastructure (VDI) connections.

Another layer of security can be added by deploying the NSX Advanced Load Balancer (formerly known as Avi networks) and forcing traffic through an intelligent web application firewall (iWAF) policy using the Avi distributed load balancer service engine. The iWAF policies offer web application security features and help achieve compliance with government laws and standards, such as the European Union's General Data Protection Regulation, the US Federal Health Insurance Portability and Accountability Act, and The Payment Card Industry Data Security Standard.

This setup is known as a positive security model, because no entities are trusted until they are specifically assigned a positive measure of trust. And, because the iWAF is an intelligent firewall, it gets smarter at security the more traffic it filters. While typical security policies are static, iWAF policies evolve as the firewall learns. With the iWAF constantly learning and training, it can protect applications from distributed denial-of-service attacks and other threats—including the top 10 security risks identified by the Open Web Application Security Project—in real time. Depending on policy settings, a positive security model helps to minimize false positives as well.

Combining the Unified Access Gateway (UAG), NSX-T micro-segmentation, and Avi advanced load balancer iWAF policies, all connection touchpoints are checked multiple times—from endpoint to application to data—throughout the connection lifecycle of the workload.

Hosting Zero Trust

Let us revisit our basic assumptions underlying Zero Trust principles. No single element across the data flow between the end user and the application workload, nor between workload elements that comprise an application, are inherently trusted all the time. Remembering this, we recognize that when building any secure environment, the security of the hardware, the hypervisor layer, and the data stores is just as important as the security of the application stack.

Using a TPM in the host servers provides a trust assurance tied to the hardware. A TPM 2.0 chip in the server hosting VMware's bare metal hypervisor ESXi will attest the host's identity. Using a Unified Extensible Firmware Interface secure boot will ensure that only signed software is loaded during the operating system boot.

Attestation is the process of authenticating and evaluating the state of the host's software at a given time. The high-level steps of the remote attestation process are:

1. *Establish the trustworthiness of the remote TPM and create an Attestation Key (AK) on it.*

   When an ESXi host is added to, rebooted from, or reconnected to a vCenter Server, the vCenter Server requests an AK from the host. Part of the AK creation process involves the verification of the TPM hardware itself, to ensure that a trusted vendor has produced it.

2. *Retrieve the Attestation Report from the host.*

   vCenter Server requests that the host sends an Attestation Report, which contains a quote of Platform Configuration Registers signed by the TPM, and other signed host binary metadata. By checking that the information corresponds to a trusted configuration, a vCenter Server identifies the platform on a previously untrusted host.

3. *Verify the host's authenticity*

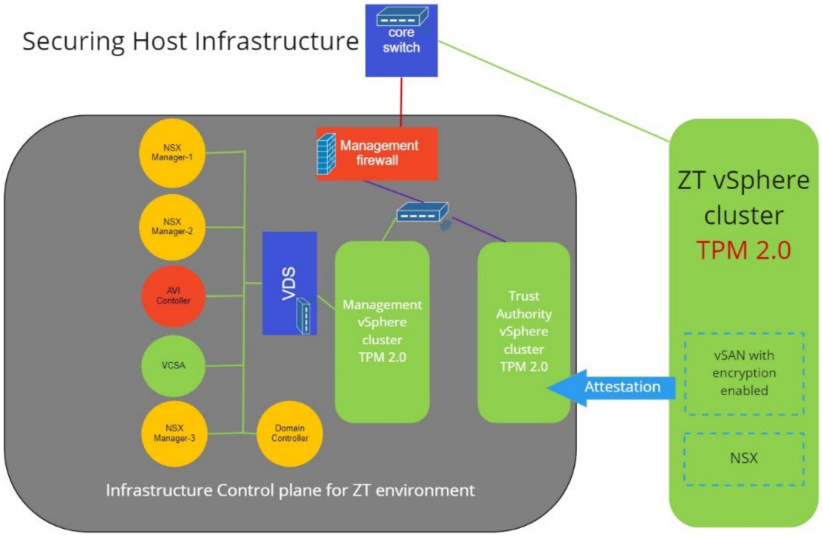
   A vCenter Server verifies the authenticity of the signed quote, infers the software versions, and determines the trustworthiness of said software versions. If a vCenter Server determines the signed quote is invalid, remote attestation fails and the host is not trusted.

For more information on Securing ESXi Hosts with Trusted Platform Modules see: *https://docs.vmware.com/en/VMware-vSphere/7.0/com.vmware.vsphere.security.doc/GUID-10F7022C-DBE1-47A2-BD86-3840C6955057.html*.

Any datastore used for hosting VMs or containers should use encryption to ensure the integrity of the Zero Trust environment. With the enterprise-class storage virtualization software vSAN, we can implement both data at rest and data in transit encryption on the datastore, thus protecting all virtual machines and data.

Adding a Trust Authority cluster in the control plane to manage attestation of the Zero Trust workload infrastructure separates the attestation of the overall environment from that of the workload. The vSphere Trust Authority Cluster acts as a centralized, secure management platform.

The Trust Authority Cluster attests the ESXi hosts in the trusted cluster remotely, releasing encryption keys only to attested ESXi hosts in the trusted cluster to encrypt virtual machines and virtual disks using trusted key providers.


Securing Host Infrastructure
core switch
Management firewall
NSX Manager-1
NSX Manager-2
AVI Contoller
VCSA
NSX Manager-3
Domain Controller
VDS
Management vSphere cluster TPM 2.0
Trust Authority vSphere cluster TPM 2.0
Attestation
ZT vSphere cluster
TPM 2.0
vSAN with encryption enabled
NSX
Infrastructure Control plane for ZT environment

Zero Trust Control Plane

Although our workload and endpoint environment follow Zero Trust principles, the control plane that manages it all cannot follow the same design principle. The problem: how can Zero Trust exist before the control plane for Zero Trust exists? And, if we stood up the control plane environment first and then pushed our Zero Trust policies to it, what happens if the control plane has an issue? We could be locked out of our own architecture. To return to our medieval castle analogy, the castle would raise the drawbridge—with us on the wrong side of the moat.

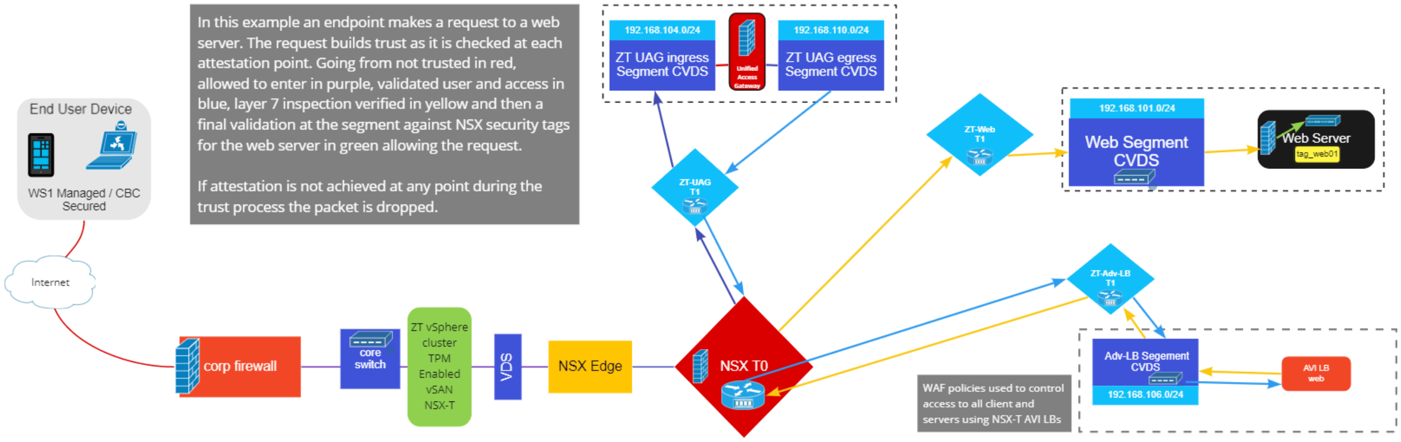
That's why we build a very secure fenced-in control plane design into this architecture. From there, we run the Zero Trust workload environment. The Control Plane is segmented from the rest of the workload and corporate network, and we apply very strict access policies to a very limited team. By using a different AD child domain for managing this infrastructure, user credentials are separated with no link or trust with the workload Zero Trust environment. User credentials must leverage multi-factor authentication, rather than rely solely on user IDs and passwords, and the management authority must be separate from that used to issue and manage credentials for the general user pool.

Another approach to further isolate the control plane—but not covered in this architecture—would be to manage the control plane as a Zero Trust environment using an external VMware Cloud environment. In that case, the external cloud control plane would manage the Zero Trust control plane, which in turn manages the workload Zero Trust environment. This scheme would force bad actors to compromise additional nested layers before they could try to break the security chain to or from the workload environment. The added complexity of management and added infrastructure cost likely makes this scheme prohibitive in many environments, however.

Network (North, South, East, West)

North to South

When an external client sends a request to the virtual IP address (VIP), it gets routed from the external router to the NSX tier-0 router, which then forwards it to the correct tier-1 for the NSX segment of the UAG appliance. The endpoint's request must first pass through the UAG before it can continue to the VIP. The UAG appliance is linked to separate segments—one segment for ingress to the UAG and one for egress from the UAG to the VIP on the Avi SE. Conditional access through the UAG appliance is granted based on rules and conditional checks from Workspace ONE Access cloud controller. The UAG checks that the client is meeting user authentication with proper device posture or compliance before allowing access to the data center application and forwarding the traffic to the Avi VIP. The traffic is then checked against the iWAF policy on the Avi load balancer for compliance before passing to the NSX firewall, which in turn validates whether the packets are compliant based on the security tags linked to security groups in the firewall policies.



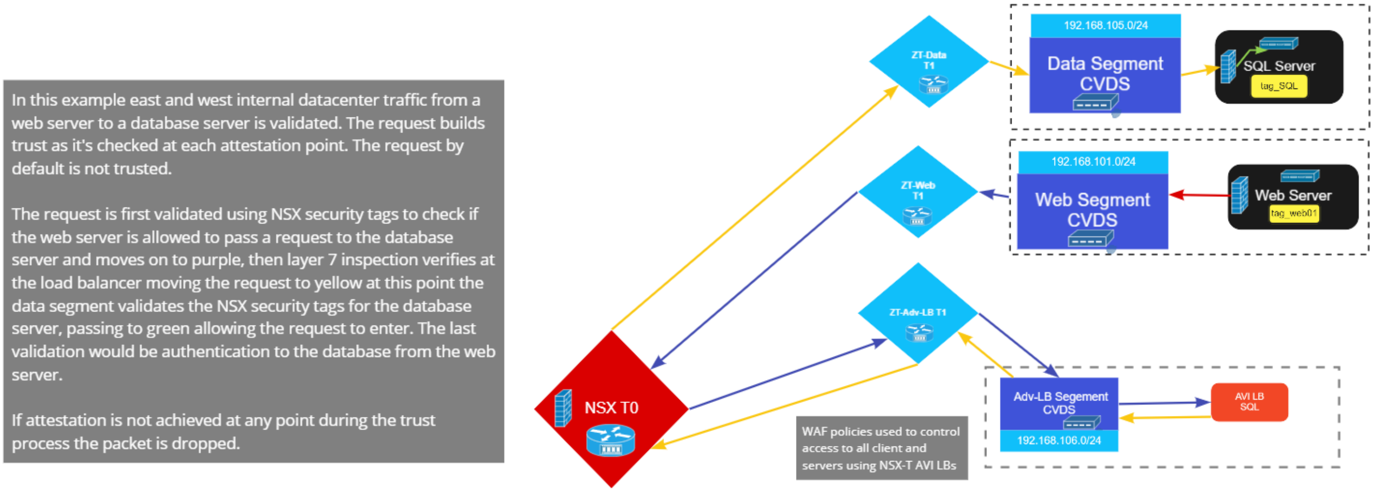
In this example an endpoint makes a request to a web server. The request builds trust as it is checked at each attestation point. Going from not trusted in red, allowed to enter in purple, validated user and access in blue, layer 7 inspection verified in yellow and then a final validation at the segment against NSX security tags for the web server in green allowing the request.

If attestation is not achieved at any point during the trust process the packet is dropped.

WAF policies used to control access to all client and servers using NSX-T AVI LBs

East to West

East-to-west traffic is handled in a similar method. Since this traffic is internal to servers in the data center, access is not tied to Workspace ONE, and so no UAG is used. All traffic is forwarded to the Avi VIP. Avi processes the traffic, checking against the iWAF policy for compliance before forwarding it on to the NSX firewall for compliance and validation. The security tag “fingerprint” of the traffic must be compliant for the network applications.



In this example east and west internal datacenter traffic from a web server to a database server is validated. The request builds trust as it's checked at each attestation point. The request by default is not trusted.

The request is first validated using NSX security tags to check if the web server is allowed to pass a request to the database server and moves on to purple, then layer 7 inspection verifies at the load balancer moving the request to yellow at this point the data segment validates the NSX security tags for the database server, passing to green allowing the request to enter. The last validation would be authentication to the database from the web server.

If attestation is not achieved at any point during the trust process the packet is dropped.

WAF policies used to control access to all client and servers using NSX-T AVI LBs

A Closer Look

Taking a closer look at each segment, we can see how data flows through the environment and is checked before moving on to the next connection in the workflow. In the following section, we walk through a basic workflow for an endpoint to connect to an application. This demonstrates how Zero Trust attestation can be implemented through the environment—not only for the endpoint connection but also internal data center connections.

Workflow

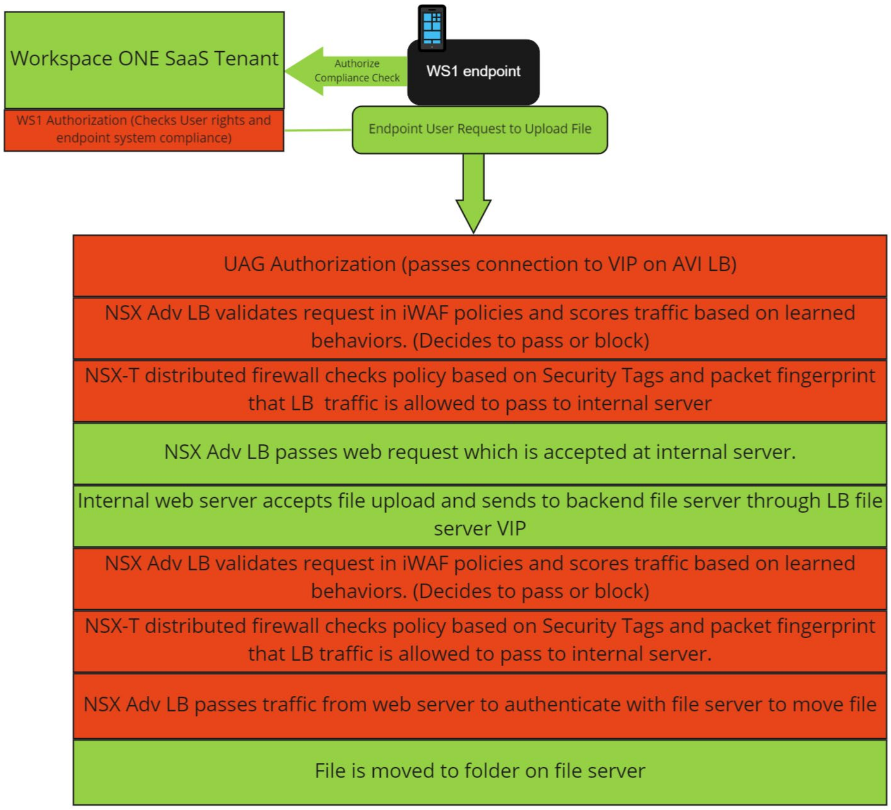
Client to Application

When users on an endpoint device initiate a connection to an application, they must first meet the policies defined in Workspace ONE for conditional access and compliance policies. This step checks that the endpoint has a compliant OS, proper patching, antivirus, and application versions—and meets other defined policies prior to being authorized to access the content.

The endpoint location proximity to the application and data, as well as user rights are also defined in Workspace One Access, to control the UAG policies to enter the network.

After the connection passes through the UAG, the packets are again checked as all data center connections terminate to VIPs on a service engine of the Avi. An iWAF checks behavior of the incoming traffic for compliance before passing it to the application server.

The graphic below illustrates the packet inspection points (highlighted in red) as the request moves through the policies and compliance rules.



WS1 Authorization (Checks User rights and endpoint system compliance)

Endpoint User Request to Upload File

UAG Authorization (passes connection to VIP on AVI LB)

NSX Adv LB validates request in iWAF policies and scores traffic based on learned behaviors. (Decides to pass or block)

NSX-T distributed firewall checks policy based on Security Tags and packet fingerprint that LB traffic is allowed to pass to internal server

NSX Adv LB passes web request which is accepted at internal server.

Internal web server accepts file upload and sends to backend file server through LB file server VIP

NSX Adv LB validates request in iWAF policies and scores traffic based on learned behaviors. (Decides to pass or block)

NSX-T distributed firewall checks policy based on Security Tags and packet fingerprint that LB traffic is allowed to pass to internal server.

NSX Adv LB passes traffic from web server to authenticate with file server to move file

File is moved to folder on file server

Application to Data

In the data center, many front-end hosted web applications will have back-end server connections to databases, files, or other content to serve information to a request. Under Zero Trust principles, we consider this type of connection just as much of a threat as external connections.

There are multiple ways to check internal server-to-server connections in order to meet our Zero Trust policies.

One method is to apply security tags to our servers in our SDN VMware NSX network. We can use security tags to define rules based on the server and type of connection. Since these are policy-based rules, a change of IP address or segment of the virtual server does not change the enforcement of the rule. Even if someone changes the IP address of the server, or moves the connection to a different segment, the enforcement policy will follow. This is very different from a traditional ACL, whereby a specific IP, VLAN or MAC address defines the policy, and changing those on a system can break the enforcement of the policy.

Another method for applying enforcement within our SDN is utilizing the Avi. By integrating with NSX, the Avi Service Engine can be used to inspect traffic using iWAF learning policies to ensure constant mitigation of potential threats inside the data center. By forcing the traffic through VIPs on the Avi load balancer, we inspect the packets before they are passed to the server. If the behavior of the packets appears out of the norm or otherwise suspect, the Avi will not allow the connection. Constant behavior analysis protects the data center from exploitation of application logic flaws that may otherwise remain undetected.

Lessons Learned: Past and Future Considerations

Past methods for securing networks and devices must adapt to the new landscapes of the modern hybrid cloud world. Yet, most organizations will find themselves using legacy applications much longer than they might like—whether due to perceived risks or the actual costs to upgrade or migrate to newer technologies. As IT teams adapt to new ways of working, they must still support these legacy applications. That’s why we will likely see organizations pursue a hybrid approach to Zero Trust principles in the future.

With this architecture, we leverage a set of capabilities in today’s VMware’s product groups that will enable enterprises to journey toward Zero Trust principles. We will continue to add additional capabilities and extend this architecture.

We note that our strategy involves applying some VMware product capabilities in a manner that may lie outside the scope of the products’ original documentation. We believe that, far from being a misuse of the products, our efforts actually offer a practical demonstration of how many end goals of Zero Trust principles may be achieved today, using readily available off-the-shelf and existing VMware solutions.

To further simplify and deliver end-to-end Zero Trust architectures, we at VMware continue to evolve our solutions, designing with Zero Trust in mind from the ground up.

Areas for Improvement

As the architecture evolves, we look to the following areas for future work:

- Centralized Control Plane: Having a single platform to link all components together for a Zero Trust architecture would be key to simplifying management and deployment. One example: the ability to push out a common policy and have it automatically update in each component of the infrastructure.
- Centralized Auditing: Aggregated audit logs of each component in the infrastructure would allow full reporting of potential security threat activities.
- Closer Product Integration: Avi and NSX could act together the same way they act independently. Avi distributed on NSX routers would better segment the traffic for east-to-west connection.
- Extending Workspace ONE: Adding data center integration for application and network compliance with Workspace ONE would extend these policies to the SDN at the micro-segmentation security level and align data centers with the endpoint.

Conclusion

Today’s companies have moved to a distributed and remote workforce model—just as significant portions of the production workload have moved out of the data center and into the cloud. As a result, IT no longer has complete control over all assets, networks, data, and physical locations.

Businesses can no longer rely on the castle-and-moat method to secure networks and managed assets.

Cyber threats have evolved from computer viruses—those which once caused costly down time, but could be cleaned and patched—to ransomware that can now hold company assets hostage or even ruin a business completely. Cyber threat actors continue to develop new tools and techniques to enable their craft—and they continue to gain sophistication and funding, all the way up to the nation-state level.

IT organizations must adopt Zero Trust models to protect business assets and reputation in the evolving technology landscape. VMware is committed to providing solutions to make Zero Trust a standard model in the emerging IT workflows of business.

Additional Resources/References

https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-207.pdf

https://www.cisa.gov/sites/default/files/publications/CISA%20Zero%20Trust%20Maturity%20Model_Draft.pdf

https://media.defense.gov/2021/Feb/25/2002588479/-1/-1/0/CSI_EMBRACING_ZT_SECURITY_MODEL_UOO115131-21.PDF

Author

Jerry Haskins is a Solution Architect in the VMware OCTO business unit, responsible for the design of various customer reference architecture solutions. Jerry's focus is on VMware Cloud, Hybrid Cloud, Security, Edge Computing, HPC and Kubernetes.

Acknowledgements

The author would like to thank Simon Phillips for editing, technical review, and additional content to this paper, and Dennis Moreau, Wayne Pauley, and Peter Bjork for their inputs, review, and feedback.

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