bpi
BITS

Adaptive Trust: Zero Trust Architecture in a Financial Services Environment

Executive Sponsors

Susan Koski, PNC

Gleb Reznik, Synchrony

Member Contributors

Vinicius Da Costa, Bank of America

John Ryan, Citizens Bank

Alemayehu Addis, KeyBank

KC Krebs, PNC

Bill Love, PNC

Joseph Frantz, Synovus

Anthony Grossi, TIAA

Michael Carroll, Truist Financial Corporation

Steve Harrington, Truist Financial Corporation

Andrew Strear, Truist Financial Corporation

Nevenko Zunic, Truist Financial Corporation

Noel Latsha, USAA

Randall Gamby, US Bank

BITS Staff

Andrew Kennedy

Table of Contents

About This Document

TRADEMARK INFO

Names, products and services referenced within this document may be the trade names, trademarks or service marks of their respective owners. References to commercial vendors and their products or services are provided strictly as a convenience to our readers and do not constitute or imply endorsement by the Financial Institutions, Bank Policy Institute (BPI) or the Authors' employers of any entity, event, product, service or enterprise.

ABOUT THE AUTHORS

BITS, the Bank Policy Institute's technology policy division, promotes current and emerging technology and fosters innovation to reduce fraud and improve cybersecurity, resilience and risk management practices for the nation's financial sector. BITS members include banks, insurance firms, asset managers, card companies, financial market utilities and technology providers. BITS also provides fora for board of directors, CEOs, CIOs, CISOs and senior executives of these organizations to have in-depth, informed, and insightful discussions on the use of technology in financial services. "Zero Trust" (also called Adaptive Trust) has been identified by the BITS member leadership as a significant technology challenge for their firms. As a result, subject matter experts from participating financial services organizations have joined efforts to create this paper.

Executive Summary

Legacy perimeter-based defense models used by financial institutions (FI) are insufficient to prevent malicious actors from causing financial, operational, reputational and client harm. The annual Verizon Cybersecurity Report shows that individuals are the weak link in security – whether clicking on malware or otherwise being socially engineered to provide credentials. This allows malicious actors to obtain access to the network, perform reconnaissance, identify high-value assets and move laterally within the organization to achieve their intent. Zero Trust Architecture (ZTA) is an approach that seeks to mitigate the risks associated with this scenario. It assumes that a malicious actor is already within an FI's network and uses a variety of strategies to reduce the likelihood that the threat actor (or careless insider) will move laterally through the network or elevate their privileges.

ZTA Objectives

- Assumes that all subjects are malicious until the subject can validate its identity and the policy engine concludes that the subject has authorization to access the resource.
- Reduces the likelihood and scope of unauthorized data access and potential exfiltration by implementing granular access controls at the data level.
- Decreases breach detection time by implementing greater visibility into the users, devices and network traffic on an FI's network to enable anomaly detection.
- Shrinks the threat surface by implementing micro-segmentation to protect resources and control access based on user identity, roles and context.
- Streamlines compliance reporting by increasing visibility into authentication and access events.
- Continuously and strongly authenticates a subject based on the risk of the data being accessed.

The Problem

In a traditional perimeter-based network, a user's identity is validated when they first come on a network. For organizations that have implemented single sign on (SSO), the cached credentials are merely presented to the next application that requires authentication from the subject. There is no defense around individual enterprise resources (e.g., apps and data), and no continuous assessment of whether a subject should be allowed access – based on both the identity of the individual and other real-time threat intelligence signals in the environment.

The Solution

ZTA evolves an FI from the traditional perimeter-based protection model to an Identity-centric protection model that is based on continuous validation of subjects that allows access to data and applications based on business models, real-time signals and risk scores. Rather than being a set of point solutions, ZTA covers several integrated capabilities and processes.

This BITS document will cover the following topics:

- Zero Trust - background information, strategy and benefits.
- Zero Trust Architecture – define the critical capabilities required in Identity & Access Management, Endpoint Device, Network, Infrastructure, Application, Data Security and Monitoring tools.
- Implementing a risk-based approach to accessing enterprise resources by leveraging contextual-based trust.

The capabilities and processes described within this document have been implemented in whole or in part across many FIs. Zero Trust is not about standing up several independent capabilities. Rather, it is about evolving into a new system of controls.

- The Zero Trust Architecture described in this paper is based upon the NIST Zero Trust Architecture (800-207).

After review, an enterprise should have the ability to review and document their current ZTA stance and then determine the path forward to mature their ZTA capabilities.

Purpose

The BITS Adaptive Trust Working Group was formed to address increasing public, industry and member organizations' concerns arising from more sophisticated and impactful cybersecurity threats. Such threats can best be met by evolving our cybersecurity strategy from the approach that builds a perimeter around computing resources to an approach that protects each individual enterprise resource. Companies should further invest in enabling the Zero Trust journey, to better protect clients, the individual institution and the US banking system which is essential for the overall health of the US economy. Additionally, this maturation in cybersecurity strategy is important to stakeholders maintaining confidence in the US banking system.

Why is Zero Trust needed? Today, most financial institutions employ a perimeter-based defense. After allowing a user to join their internal network, each user typically enjoys a broad range of lateral movement within the network; regardless of their role within the organization. Unfortunately, this approach allows each user far more access than they need to perform their job duties. A flat (non-segmented) network architecture allows a malicious actor, who gains access, the ability to move across a network and gain unauthorized access to sensitive data. Likewise, automated network-based malware or ransomware can spread laterally across an open network architecture damaging critical system functions and availability that could temporarily or permanently disable an organization's business operations. These are the types of risks that Zero Trust architectures are attempting to address.

What is involved with Zero Trust? Implementing Zero Trust can be a large and complex endeavor involving the deployment of layers of integrated technologies. Each layer is responsible for continually validating a subject's (for example, a user) permission to access an enterprise resource (for example, application, system or data). The validation is based upon a real-time risk (and policy) score.
How does this paper assist an FI? The goal of the Working Group was to develop a whitepaper which:

- Explains Zero Trust in a straightforward manner that helps financial institutions better understand what Zero Trust methodologies are and are not;
- Describes commonly used Zero Trust use cases, applicability and applicability benefits;
- Describes the high-level architectures and components; and discusses best practices to building a Zero Trust environment (e.g., stakeholder involvement, enabling technology capabilities, sequences, data, Identity, etc.).

What FI persons does this paper target? This paper is written for a broad range of financial services organizations across technical and business roles: technology strategists, cybersecurity specialists, architects, management and senior management who want a greater understanding of Zero Trust.

Evolving Landscape – The Deficiencies Of Today's Model and The Promise Of Zero Trust

The finance sector is a highly attractive segment for cyber attackers due to potential financial gains; hence, organizations need to be aware of the latest threats and attack vectors in order to minimize the effect of potential breaches. The financial industry is being targeted by cyber attackers and needs to constantly evolve and apply controls to minimize the impact of a breach. Limiting the impact of a breach and reducing cyber risks to an

acceptable level is the focus of this report. This report describes steps organizations can take to contain and limit disruptions to their business from a successful cyberattack by adopting and implementing "Zero Trust" principles.

For decades, the typical enterprise cybersecurity model was based on the concept of physical security and a trusted network with layers of security controls applied at the network perimeter. This is akin to a castle protected by a moat and with entry over a single, well-guarded drawbridge. To that end, an FI created a perimeter-based defense around its datacenter with firewalls and other inspection devices (analogous to drawbridge) to keep unauthorized actors out. Any visitor who was granted access was implicitly trusted and could roam the castle and its grounds. However, that model that served FIs well for decades is now antiquated.

- The hardened network perimeter that organizations have historically relied upon for protection is failing to prevent breaches. The tactics and techniques of bad actors follow a familiar pattern. It involves the bad actor establishing a foothold in the victim's network; then expanding their initial access by escalating privileges; finally, moving laterally to other systems until a desired target is compromised. The initial foothold is often the device of a user that fell prey to a social engineering technique, or an application compromised through a sophisticated supply chain attack. The damage resulting from a breach is directly related to the ability of the threat actor to move, unfettered within the network they've breached. This tactic, which is well documented in the MITRE ATT&CK Framework, has frequently shown that security controls implemented at the perimeter of the network are ineffective against these tactics and techniques. Instead, FIs must assume a bad actor has successfully entered the network. With a bad actor within the castle walls, the FI needs to protect individual resources (e.g., systems, applications and data) by adding real-time controls that restrict access to each enterprise resource.

The enterprise perimeter is no longer a well-defined boundary. The accelerating trend of remote work and mobile devices with access to enterprise resources means the threat actors are often on the same network as off-network enterprise endpoints. The threat is compounded because many organizations still rely on endpoint security tools and processes that are degraded when devices are off the enterprise network. This makes them a susceptible target for threat actors to compromise and ultimately, establish a foothold within the enterprise network.

The proliferation of public and virtual private cloud footprints (private cloud instances and SaaS offerings) of FIs makes the enterprise network perimeter less meaningful as a control point. Enterprise users are now dependent on applications and data outside of the traditional network perimeter, with their own set of security controls. While this presents an opportunity to apply more granular and risk-based security policies, it also creates at least two challenges. First, there is a risk of applying inconsistent security policies. A poorly secured public cloud resource is a ripe target for threat actors anywhere in the world. Second, a new security model is required that is focused on protecting data and applications regardless of the underlying system or network.

The combination of ineffective network perimeter controls and the porous enterprise perimeter means we must let go of the concept of a trusted network protected primarily at the perimeter. Instead, we must develop an approach to enforcing security based on the sensitivity of data, user and application identity and device security posture.

Shortcomings Of Current Models And Approaches

Over the years, threats and attacks have evolved and become more sophisticated. Cyber attackers are taking advantage of organizations that have not evolved their security controls to thwart new attack vectors.

Many organizations still have flat networks and are reliant on perimeter protection. Why is a flat network problematic? A flat network allows a bad actor, either external or internal, to easily move throughout a network and gain access to multiple resources beyond their initial compromised device. This increases the potential exposure and data loss for an organization. As a goal, lateral movements should be restricted and breaches contained to one area to minimize risks.

- Layered security controls have been inconsistently applied over the years and contribute to the ease of movement by an attacker with the lack of additional controls to overcome. Layered security controls can slow down an attacker and improve the ability to detect malicious activity.
- Use of weak authentication presents a risk to the organization and should be avoided. Passwords no longer provide sufficient protection to a network - many are weak, easy to hack or reused by employees or an FI's clients (so a breach at another company can make your FI vulnerable). Instead, stronger multi-factor authentication should be implemented at the perimeter to better keep bad actors out.
- Authentication is performed once, when a user enters the network. There is no consideration of whether a user is performing normal activities (versus attempting access to new enterprise resources). The ability to perform continual or risk-based authentication can limit the ability for a bad actor to move laterally or upward in privileges.

Initial Authentication may not consider the sensitivity of the data. The ability to perform risk-based authentication can limit the ability of a bad actor to move laterally or upward in privileges.

- The authentication request may originate from a compromised or unauthorized device. In today's environment, many FIs are limited in verifying the provenance of a device that connects to the network.

How an FI responds to incidents requires modernization. Security teams can no longer rely on manual reviews of potential incidents and lengthy timeframes to apply remedial controls. These processes need to be automated to quickly and efficiently react to potential breaches. Further, security controls need to be continually reviewed and enhanced to keep pace with new and advanced threats.

Establishing Trust

When an enterprise resource (system, application or data) receives an access request from a user or subject, how can it trust that the access is being initiated by the user whose identity is being asserted? How does the enterprise resource know it is not someone who has stolen the credentials of that user? Even if the access is by the asserted user, how can the application or system trust that the data or functionality being delivered to the user is not going to a personal client endpoint without enterprise security controls? Could this bad actor be using an enterprise-managed system that has been compromised and is being used for data exfiltration? Password-based authentication methods are insufficient to establish such trust and, in most cases, are insufficient for getting network connectivity to the application.

In a zero trust security model, access to the enterprise resource will not be granted unless the user, or subject and device are properly authenticated and establish a certain level of trust. User access to enterprise resources is defined by the sensitivity of the data or functionality being accessed and the permissions associated with the roles being granted. The strength of the authentication method should be commensurate with the level of access being requested. The subject and device attributes must also fall within predefined risk tolerances for the desired level of access. Strong user authentication entails the use of multiple authentication factors, such as one-time passwords, biometrics, tokens, location or knowledge-based answers. Options for device authentication include diagnostics and software versions that are often combined with enterprise-issued certificates.

Consideration of the authentication context such as the security posture of the device, time of day and geolocation can be used to establish the required degree of trust that the access attempt is from a legitimate and authorized user. Consideration of FI policies (e.g., separation of duties) can be used to establish the level of trust necessary to receive access to an enterprise resource.

The zero trust security model also applies to system-to-system communications. The source IP in a system-to-system communication is no longer a strong enough authenticator to allow connectivity. Access would require some combination of strong authentication credentials, consistency with predefined communication and application dependency maps and source system security posture checks. Such factors must be considered to determine if

communication is permitted and expected or if it is an indicator of potential lateral movement by a malicious actor that must be blocked or flagged for investigation.

Benefits Of A Zero Trust Architecture

Zero Trust provides technology, security and business benefits. The technology benefits arise from the ability to segment the network (e.g., reduces lateral movement and attack surface); Strengthen and shift identity to the center of security (e.g., which reduces the likelihood of bad actor co-opting legitimate credentials); focuses on risk-based authorization (e.g., authentication context, riskier transactions, etc.). The business benefits are associated with increased efficiency and reduced operational friction.

Regarding the security benefits of Zero Trust:

- Reduce attack surface - Through granular segmentation as well as by protecting individual enterprise resources, bad actors are limited in the damage they can do – whether it's impacting service availability or access and data for exfiltration. Additionally, bad actors have a much harder time "pivoting" from a compromised system or device to an uncompromised one. As a result, the "splash" associated with a breach is substantially smaller, and the "dwell time" that a malicious actor can remain undetected is reduced.
- Improved Application and Data Layer Protection - By restricting each enterprise resource to authorized users, with the proper context, device posture, etc., organizations can better secure client and business data – imposing greater controls on more sensitive data.
- Reduce the likelihood of malware and data exfiltration - Through monitoring the security posture and identity of devices, FIs can be more certain to avoid malware-infested devices
- Enhance the security of remote workers - With a dispersed workforce, making identity the center of security lessens the reliance on perimeter controls – especially when accessing data that is off premise
- Reduction of the time to detect a breach - Increased visibility of the network and real-time access control decisions will inevitably accelerate an FI's efforts to detect and respond to a breach. Zero Trust designs are more difficult for bad actors to traverse so bad actors will make much more "noise" in a breach scenario.

Regarding the business benefits of Zero Trust:

- Streamlining Access - By moving to Attribute Based Access Controls, in addition to RBAC, access to enterprise resources becomes dynamic, contextual and policy based.
- Reduce effort for periodic access reviews - Rather than performing reviews of every entitlement, attestations can be reduced to outliers – those not adhering to an archetype.
- Move from validating compliance to standards to continuous compliance - Policy Enforcement points in the Zero Trust network will ensure standards are followed – and audits merely look to central devices for policy adherence, rather than each device.
- Improve visibility of assets and how enterprise resources are used - The tools and techniques will help uncover what data an FI has and where.
- Support for Bring Your Own Device (BYOD) efforts - Once Zero Trust is in place, approved workforce should be able to securely access resources using personal devices including phones, laptops and tablets, not just their corporate-issued assets.
    1. Busy executives can use their personal iPad to access applications to approve transactions or collaborate with their peers
    2. Mobile developers can use Macs and personal devices to test their mobile applications

3. New hires can begin working in a limited capacity on personal devices before a corporate laptop is available

- Providing application-specific access - IT contractors and remote or mobile employees could have application-specific entitlements and can be an alternative to VPN-based access.
- Mergers and Acquisitions – Absorbing and integrating new companies, even those that have a traditional perimeter-centric network, may facilitate the integration process.

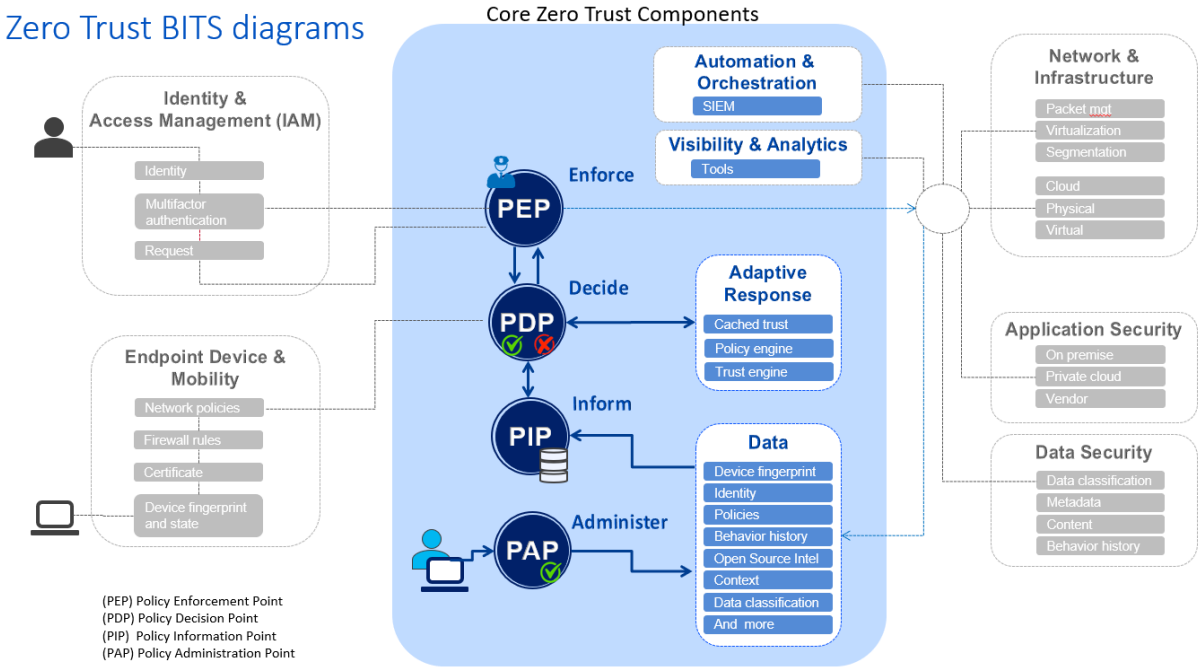
What is Zero Trust?

Unlike the historical perimeter security model that assumes that everything on the network is trustworthy, Zero Trust assumes the opposite. NIST identified several assumptions of a Zero Trust Network in publication SP800-207. The assumption that drives all other NIST assumptions is that no subject, regardless of location, is to be implicitly trusted. Because there is no implicit trust, each access request to an enterprise resource is to be validated and must include contextual data signals into the decision. As a result, Zero Trust grants Just-In-Time access; continually validates access; and, when trust is established – access is granted at a granular level.
The Zero Trust model (referred to in SP800-207) is based on a set of guiding principles: assuming a breach, use of least privilege and establishing a high degree of trust (based upon, roles, attributes, authentication context, sensitivity of target data, etc.) NIST outlines seven tenets of Zero Trust that should be viewed as goals of a Zero Trust architecture. These tenets include:

- All data sources and computing services are considered resources
- All communication is secured regardless of network location
- Access to individual enterprise resources is granted on a per-session basis
- Access to resources is determined by dynamic policy—including the observable state of client identity, application/service and the requesting asset—and may include other behavioral and environmental attributes
- The enterprise monitors and measures the integrity and security posture of all owned and associated assets
- All resource authentication and authorization are dynamic and strictly enforced before access is allowed
- The enterprise collects as much information as possible about the current state of assets, network infrastructure and communications and uses it to improve its security posture

Components of a Sustainable Adaptive Security Solution

Now that we have an understanding of the value and benefits of a zero trust architecture, we will discuss the technical components that are necessary to build such an environment. A zero trust implementation requires several key capabilities which will form the foundation for the solution. A description of the key components is provided in this section. This document divides them in two groups: Extended Ecosystem and Core Components. This is because Zero Trust Architecture influences, integrates and depends on most of an organization's infrastructure and security domains.



Extended Zero Trust Architecture Ecosystem Components and Role

Extended ecosystem includes Identity & Access Management, Device, Network, Infrastructure, Application and Data Security. These areas provide key telemetry to the core components and need to be adjusted to work and bring the most benefit of a Zero Trust Architecture implementation.

Identity & Access Management (IAM)

When users access a Zero Trust computing network, determining user identity beyond credentials is foundational in implementing this strategy. Only entitled and authentic subjects or users are allowed access to resources and the principle of least privilege is applied. Authentication and authorization approvals for subjects are not persistent and are required for each session. It is possible that some attributes may have changed that will influence subsequent decisions whether to grant or deny access.

The goal of Zero Trust IAM is to accurately and as completely as possible, identify authorized users from impostors by validating user and device context. Requests are directed to a policy decision point (PDP) that checks a combination of attributes associated with the requestor and their device and calculates a trust score to determine whether to grant access. The access decision is enforced by a Policy Enforcement Point (PEP).

Accomplishing this level of control requires supporting remote access capabilities in addition to traditional user ID/password combinations and multi-factor authentication. This value-added level of control is called Software Defined Access (SDA), which checks the cybersecurity and technology posture of the endpoint system requesting network access. SDA control checks may include a combination of device fingerprinting/profiling, geo-location identification, patch/anti-malware/anti-bot signature checking, Active Directory ID/password check, challenge questions, SMS code prompts and other attributes. SDA provides real-time visibility into all endpoints continuously as they interact with other resources, thereby mitigating risks posed by imposters trying to gain access even if they have obtained a valid user's access credentials.

Underlying Zero Trust IAM are two additional foundational principles. First, the entire process is based on the "Least Privilege" model, including control of ongoing privileged access. Attackers frequently target insiders with elevated access making use of a quality Privileged Access Management system a strategic imperative. Second is the need for

strong end-to-end monitoring of activity. In advanced Zero Trust IAM implementations, risk-aware security integration is in place where risk is dynamically monitored, evaluated and risk policy updates are propagated across control planes in a near real-time basis.

Finally, robust subject and user validation and remediation support services must be available to handle legitimate users who are rejected by this stringent process.

> **Crawl**: Organizations implement basic, central identity management capabilities such as: role-based access control; strong privileged access management capabilities; and ensure that shared services such are mature (e.g., improved the data quality - completeness and accuracy - of CMDB).
>
> **Walk**: Organizations implement fine grain identity capabilities such as: attribute-based access model that will serve as an input for the risk/policy engine; create an Access Model that ties Roles to application entitlements, and implement the role/group structure in the corporate Identity Services. Additionally, applications need to be integrated into or (through existing federation) into MFA.
>
> **Run**: Organizations have strong identity leveraging the risk components of Zero Trust. For examples, Risk Engine that receives inputs from real time sensing instrumentation; Applications are federated to the FI's identity store rather than having a standalone AuthN/AuthZ store; the Risk Engine engages with the Policy Enforcement Point. The Attribute Based Access Control (ABAC) model is used to define access policies across the enterprise using attributes instead of roles. This allows for granular, yet flexible access control policies.

Network & Infrastructure

Networks are the pipelines that allow subjects to access enterprise resources and networking controls can provide enhanced visibility and prevent threat actors from moving laterally. Macro and Micro segmentation is critical to preventing attackers from moving laterally after gaining a foothold in the environment. Historically, networks are divided into different and typically large trust zones – enforced by firewalls and guarded by other security monitoring and countermeasures devices such as WAFs, IDS/IDPs. Typically, the zones are internet (untrusted), DMZ and an internal zone (network, trusted). These internal networks could also contain one or more secure enclaves where high-value resources are placed such as an order routing application for a brokerage.

The concept of an internal network changes under a Zero Trust Architecture. Every resource is protected individually and not visible to the subject until authentication and authorization is completed. This process is continual and executed at the time an entity attempts to access a resource. Anyone trying to forcefully scan the network to identify potential targets will not be able to find them or even know they exist, simply because they will not respond. This can be achieved by employing mutual authentication where parties at each end of the network connection verify their identity.

The goal of Zero Trust Architecture is to maintain the smallest possible trust zones in a way that enforces least privilege network access for all resources. For example, an important outcome of reducing the size of trust zones is the ability to separate Information Technology (IT) from Operational Technology (OT). Companies need to run services such as ATMs that are critical. It is known that cyberattacks are also targeting critical infrastructure, not only information systems. In certain cases, OT systems are more vulnerable or could be more critical to the organization. The separation will prevent attackers from moving laterally and infecting OT infrastructure by attacking IT infrastructure first or, the other way around.

With Zero Trust Architecture, we change the focus from protecting the network perimeter to creating micro-perimeters around individual resources or a group of resources. Micro-perimeters or micro-segmentation should not be the only focus within the networking pillar of Zero Trust Architecture. End-to-end encryption and enhanced

monitoring should also be employed. Each organization will have a varying level of segmentation within their networks and should take a walk, crawl, run approach to achieving micro-segmentation. As an example:

> **Crawl**: Organizations can segment and define their network infrastructure using large perimeter and employ macro-segmentation. For example, segment your production network from your QA network. Organizations may leverage existing segmentation models such as HR or Legal to further segment their network.
>
> **Walk**: Organizations can segment their network infrastructure by configuring ingress and egress micro-perimeters and include some internal micro-segmentation. Granular policies can be defined at the perimeter.
>
> **Run**: Organizations have internal micro-segmentation based around all application flows and application traffic is encrypted end-to-end.

Endpoint Device & Mobility

A fundamental principle of Zero Trust is that endpoint/host systems connecting to the network are not implicitly "trusted" by default and those systems need to be validated against the enterprise inventory, evaluated for their technology/security status and appropriate access policies are dynamically assigned to these systems based on the results of the evaluation, i.e., their defined "context". As a result, untrustworthy systems may be rejected from joining the network, prevented from accessing protected resources or designated for remediation. The goals of this phase are to gain greater identification, visibility and overall control of all endpoint systems, including mobile devices.

To achieve this level of visibility and control requires the enterprise to have a complete and comprehensive inventory of all enterprise assets – infrastructure devices, applications, cloud components and endpoints, including mobile devices. Next, a broad series of tools must be in place which allow interrogation of each device, including Mobile Device Management (MDM), anti-malware, vulnerability/patch management, inventory checking, application/software management and other tools that define the technology and security context of each device. Finally, detailed policies must be specifically configured and managed to define authorized access based on attributes of the user's identity and security context of their device. In advanced implementations, access decisions of non-managed endpoints are made in real time to assign appropriate access levels or perform remediation activities.

Creating a detailed device fingerprint is critical to identify potential compromise of a valid endpoint, or a sophisticated impersonation of one. The fingerprint includes details about hardware components, TPM public key, firmware, OS, software, etc. By having a comprehensive picture of a device, the PDP will be able to identify potential anomalies and respond appropriately with a range of available enforcement policies.

> **Crawl**: Organizations should have central management of all end user and central devices; complete and accurate Central Management Database (CMDB) information of all end user and central devices; basic cyber capabilities such as Endpoint Protection (EP) or Endpoint Protection Platforms (EPP).
>
> **Walk**: Organizations can establish the identity of devices on the network, disallowing unknown devices from connecting; Zero Trust capabilities such as policy enforcement points are distributed to end user and central devices; application data flow is documented. Technologies are put in place to perform device posture assessments when connecting to the network from any location. Endpoint Detection & Response (EDR) solutions are employed to gain enhanced visibility into the endpoint and provide automated response to some alerts.

Run: Organizations allow or disallow devices to connect to enterprise resources based upon policy; enterprise resource groups are dynamically updated as they are added/removed from the inventory; PEP capabilities are in full use at a granular level; Device to enterprise resource policies are groomed and not enabled merely because they follow a historical pattern. Device posture assessments are performed continuously during the lifetime of a device connected to the network.

Application Security

The application layer plays a pivotal role in Zero Trust Architecture because it acts as the final data protection "gateway" or layer before data can be accessed. The application manages the security of data to and from the target data system/repository. The application is coded to act as the intermediary between continual checks from previous "identity awareness" Policy Decision Point (PDP) control layers on the front end and also provides the data control interface provided at the Policy Information Point (PIP) layer that manages appropriate access rights, permissions, privileges and data security controls to functions and data within the application. In addition to previously mentioned micro-segmentation technology, DevSecOps, Software Defined Computing (SDC) and software supply chain technologies and processes may come into play at this level of the Zero Trust methodology.

Crawl: Organizations control access to applications using local authorization. On-premises applications are accessed through the physical network or VPN. Application security testing is performed prior to deployment using manual, static and dynamic methods. Conduct penetration testing.

Walk: Centralized authentication and authorization technologies are employed to control access to applications. SSO technology is configured for the workforce to access on-premises applications. Organizations integrate application security testing into the development and deployment cycles using IAST, SCA and dynamic tools.

Run: Access to applications is authorized based on real-time risk analysis and authentication and authorization happens continuously throughout the connection. In-session monitoring and response is put in place for all applications. Application security testing is integrated throughout the development and deployment cycle and leverages regular automated testing of deployed apps. Use Breach and Attack Simulation Tools (BAS) to identify how attackers would laterally move within a network.

Data Security

The overriding focus of Zero Trust security models is to assume that an attacker is present in the environment at all times. This posture is assumed for one reason: to protect the organization's sensitive data assets from unauthorized access, misuse, alteration, theft, exfiltration or other methods of compromise. Data discovery, data flow mapping and data classification are key to establishing granular access controls where "data tagging" or labeling is used to ensure that individual users can only access data to which they are permissioned.

Zero Trust requires that data residing on any endpoint system and while it is in transit between systems must be protected. Data on endpoint systems must be encrypted when it is not in use, during transmission and when it is not in use on the target platform or database. Other data level security controls may be used as well that protect the data from compromise including, data rights management, software defined storage (SDS), encrypted databases, tokenization, data loss prevention, data masking, etc.

Crawl: Organization uses static controls for restricting access to data. Organization partially encrypts data at rest and leverages least privilege to control access to data. Manual categorization and inventorying of data results in poor or inconsistent categorization.

Walk: Organization encrypts all data at rest and leverages least privilege to control access to data. Device risk and other attributes are also considered when making access decisions to data. A combination of manual and automated analysis is used to classify data.

Run: Organizations encrypt all data at rest and in-motion. Technologies are employed to dynamically classify (tag) and categorize data. Access to data is controlled using dynamic and contextual risk-based decision and only allows just-in-time and just-enough access.

Visibility

Visibility, in a Zero Trust Architecture, is about having insight into the activities and events that are occurring on the network. In a perimeter-based defense, a subject is provided with Authorization and Authorizations when they authenticate and join the network. In a Zero Trust architecture, a subject is granted access to individual objects. Zero Trust visibility is about having visibility into:

- Network traffic and patterns of a subject. This requires integrating various cyber tools into a SIEM and a risk engine (discussed in later sections)
- Other real-time security events occurring in the network.
- The sensitivity of data stored in enterprise resources (e.g., where PII is stored in the environment)

Visibility & Analytics

Visibility into the above components across the ecosystem provides critical, contextual data to make trust decisions or act on potentially malicious activity. This will also improve detection of anomalous behavior and decrease the time it takes to identify and contain a threat. With a mature ZTA, organizations will log and inspect all traffic across all ZTA components to achieve enriched visibility and understand the behavior of the environment. Integrating visibility from multiple vantage points and adding telemetry data provides the network defenders an enriched picture of what's happening in the organization's environment and better protect sensitive applications and data. The visibility of all activity across all layers will also provide organizations the ability to make real-time decisions for updates, security policies and access requests.

A multitude of tools and technologies can be deployed to collect, track, baseline and monitor data activity to provide vital visibility and analytics to help an organization defend its network. Some of these tools and technologies include real-time monitoring, Endpoint Detection and Response (EDR), Network Detection and Response (NDR), Deep packet inspection (DPI), Encrypted Traffic Analysis (ETA) and Security Information & Event Management (SIEM) systems. Machine learning (ML) and Artificial intelligence (AI) will also need to be employed by organizations to augment their analysts by helping to identify patterns, drawing conclusions and making recommendations on potentially malicious activity.

Crawl: Organizations send all logs to a centralized system. Network visibility is achieved using network firewall logs. Periodic network discovery activities are performed to update device inventory systems.

Walk: Organizations send most of their logs to a centralized SIEM. Network Traffic Analysis tools are in place to analyze north-south traffic. IDS sensors are installed on external facing network connections and integrated with threat intelligence feeds.

Run: Organizations collect logs across all ZTA components to a central SIEM and integrate analysis across multiple sensor types to create automated alerts. Continuous device posture assessments are automated using technologies such as Endpoint Detection and Response (EDR) or Network Access Control (NAC). Encrypted traffic analysis technologies are employed and integrated with other sensor types to provide visibility across the enterprise.

Automation & Orchestration

Automation and orchestration provide the organization the ability to automate manual security processes and execute policy-based decisions at speed and scale. In a ZTA, security automation and orchestration integrate across all pillars of the ecosystem to provide automated responses to act on malicious behavior. Security

Orchestration, Automation & Response (SOAR) technologies integrate with an organization's SIEM to free up security resources by helping to manage disparate security tools and automating manual responses to threats. SOAR technologies require integrations through APIs and well-defined and tested processes to avoid any interruptions from legitimate business processes. Technology is at the foundation of automation and requires organizations to select heterogeneous vendor products that can be integrated with APIs to achieve automation and orchestration across the entire ZTA. This helps organizations achieve a proactive command and control posture and enforce a consistent security policy across the environment.
Automation and orchestration need to be intertwined with each ZTA component. As your Endpoint, Network and other components mature, so must the capabilities of automation. Maturity of automation and orchestration will increase as the maturity of other ZTA components progresses. Integration across all ZTA components is critical to fully realizing the benefits of automation and orchestration. We will use the Application component of ZTA to highlight how automation and orchestration can mature within a single pillar. The use of Security Orchestration & Response technologies will also be highlighted.

Crawl: Organizations determine hosting location and access of the application during provisioning. Network and infrastructure changes are manually initiated using change management workflows. SOAR tools are employed in the Security Operations Center (SOC) to automate workflows for security analysts.

Walk: Applications inform the infrastructure and network components of a changing state. Automated workflows are leveraged to manually initiate network and infrastructure changes. Organizations leverage SOAR technologies to automate response to alerts from the SIEM and take action to contain or remediate potential threats.

Run: Applications adapt to ongoing environmental changes for security and performance optimization. Organizations leverage Infrastructure as Code (IaC) to make network and infrastructure changes with pervasive automation. SOAR technologies are fully integrated with PDP components to update the policy engine if malicious activity is seen on the network. Integration with all PEP components is also leveraged to provide automated responses to alerts and incidents.

Core Zero Trust Architecture Components - Technical Support Framework

There are several core components that make up the heart and brain of Zero Trust Architecture. It receives requests, telemetry, context signals, data and policies from the extended ecosystem and orchestrates access based on ongoing evaluation of the combined information gathered. It never creates persistent trust. Instead, it will constantly interact with the extended components to continuously evaluate overall threat posture and respond appropriately. It works with a range of responses such as requiring an extra authentication factor, reduced functionality or denying access and locking an account. This improves user experience compared to a traditional Boolean response where a user is either granted or denied access.

Decision Layer

Policy Decision Point (PDP)

Zero Trust requires each user to be authenticated and continually authorized before granting or for maintaining access to a resource (e.g., data, application, system, etc.). The NIST Special Public Resource 800-207, “Zero Trust Architecture,” describes the framework of a Policy Decision Point, composed of a Policy Engine which makes and logs the decision whether to grant or deny a subject access to a resource and a Policy Enforcement Point, that enables, monitors and terminates sessions between a subject and a resource. The Policy Decision Point implies the concept of policy management which are the baseline rules the Policy Engine follows to grant or deny access. Policy Management covers the process of creating, implementing and managing the organization’s access rules – at a fairly granular level.
A request authorization Decision Layer in a Zero Trust Architecture implementation has three uniquely defining characteristics compared to traditional authorization models:

- **Authentication and Authorization** – Risk-based authentication and authorization will take place when the subject's access context changes. Fresh validation of user, device, security posture and context deliver higher assurance of the true identity of the requestor and validity of the request. To maintain usability and performance, the Decision Layer can be configured to determine the risk of a new transaction and apply stepped-up authentication.
- **Comprehensive Telemetry, Context and Intent** – Zero Trust Decision will consider a much larger data set for authorization. That's why it is critical that organizations standardize telemetry, signals and logs to provide a simple, efficient and more comprehensive view of what is behind an access request or action. In addition to traditional information such as policies, credentials and roles, ZT Decision will incorporate contextual information such as location, behavior and device —among others — before it defines what the best response is. As an example, an associate's access may change based on whether she is in a public or private location.
- **Dynamic and adaptive** – An authorization decision in Zero Trust will increase from a binary response (yes or no) to a range that best reflects the reality of security and maintains a positive user experience. This creates the ability to provide access to low-risk public information in certain conditions but block access to confidential information at the same time for the same person. The decision reflects the context around the request and dynamically adapts the response accordingly. Also, new signals, telemetry and context information will immediately change the response without intervention from an administrator.

The idea behind this more sophisticated response is to better understand the risk context associated with a particular request and react appropriately. In a basic authorization process, the system will check whether the credential is valid, check what roles/privileges the credential is assigned, match with pre-defined policies and make a binary decision.

A binary authentication/authorization decision leaves the organization open to insider attacks and malicious actors that compromise credentials (identity theft). Zero Trust incorporates other factors that could enable an organization to thwart a malicious use of credentials. For example, UEBA might sense a series of transactions that align with the steps threat actors take to exfiltrate data then immediately reduce the trust score of the user, such that valuable resources are not accessible. In turn, this could prevent the insider or threat actor from achieving their goal.

The challenge of such a model are false positives and negatives based on more subjective and automated decision making. However, this model will never grant access to illegitimate activities. It will never increase undue access as foundational policies remain in place. It will provide an added layer of risk identification.

Implementing a Zero Trust Decision layer can be performed incrementally; however, the requirement for continual authentication is foundational. It is critical to change the authentication paradigm from the edge of the typical company "secure" network to a continual model independently of where the request is coming from.

Policy Engine

This component is ultimately responsible for the decision to allow access from a subject to a resource. Security policies and external sources that can provide contextual information as input to the decision process need to be integrated with the policy engine. The adaptive Policy Engine is a core feature of the Zero Trust Decision layer. It is the component where the information is aggregated from policy, context, telemetry and signals, and where the final decision is made. There are different ways the engine can be designed:

- **Risk scoring** – some Zero Trust implementations adopted a numerical risk that is associated to each input and its thresholds. A calculation outputs a final risk score that is used to determine the response depending on ranges.
- **Criteria-based** - A defined set of criteria must be met before a subject is given authorization to access an enterprise resource. Each enterprise resource or group of resources should have its own criteria defined.
- **Count and Sum** – Simple binary result for each attribute with zero (0) assigned to normal and one (1) assigned to abnormal. Sum up results and compare to a simple risk scale.

Creating a working decision engine is one of the most important factors of a Zero Trust implementation. At the time of this publication, we could not find a product that could fulfill all requirements. Several companies are developing their own engines and vendors are investing in the concept as well. Selecting technologies that have robust APIs and integration capabilities should be favored.

The Policy Engine needs to be aligned to the enterprise's security policy and integrated so updates to the policy are automatically applied. Enterprises will need to convert their natural language policies (NLP) to digital policies (DP) for the policy engine to interpret and enforce organizational security policies.

Contextual awareness information should also be sent to the policy engine. There are many data points that can be used for the decision engine context awareness. Identifying the list of available data sets, signals and telemetry is an important part of your company's journey for implementing this feature. This allows for a progressive integration of new data and rules into the decision engine over time depending on security value and effort. To better illustrate the concept, consider the following data point examples:

- **Location** – can be used as part of behavior analysis in case a credential is used from an unusual location. It can also be used to determine privacy provided by a particular location, i.e. home, office, public area, etc.
- **Travel time** – derived from the location it provides the ability to determine whether consecutive accesses from different locations would be physically possible.
- **Time, day** – understanding what time of the day, or day of the week access is requested can also determine authorization results.

It is important to understand that individual context data may not be sufficient to determine an attack. It is also important to consider that information can be manipulated by attacks, such as the use of a virtual private network or proxy technologies to mask the location of the requester. These technologies are simple to use and can be deceiving but also detectable.

In summary: the more comprehensive the set of context data, the better the decision output. If many of those are signaling anomalies, there will be a higher chance of identifying potential malicious intent early. It's also important to understand that this data can be used to adjust access -- for instance, requesting an extra alternative authentication or reducing access to less risky assets.

Administration Layer

Policy Administration Point

Management of the policies that the PDP uses to decide an access request is critical to ensure a successful implementation of the decision layer of ZTA. This incorporates the creation, implementation, fine-tuning and governance of the organization's digital policies (DP).

Policy Creation

To create Access Policies, an organization must have a comprehensive understanding of its subjects and resources.

- **Subjects** must be gathered into logical collections and then divided into sub-collections. For example, business users and third parties may be large collections which are each then divided into sub-collections such as wealth management, corporate fixed income, government bonds, trading and third parties into call centers, business service providers, technical services providers, supplemental staffing, business process outsourcing, etc. Of course, these groups can be further divided until there is a diminishing return in further perceived risk reduction.
- **Resources** must be gathered into logical collections, perhaps based on user groups or business functions. This is dependent on having a comprehensive and accurate system of record that holds an organization's resource inventory as well as the ability to update that inventory in near real-time as the resource footprint changes.
- **Data Flows** between Assets and Resources must be established, from a policy perspective. Once established, actual data flows can be analyzed to identify gaps in the policy. This can be an iterative process to balance security and operation risk as the organization advances the data flow policy.
- **Application Architecture** is also considered as there may be reasons to consider which services within an application should be able to establish connections with other services.
- **Attributes and Organizational Roles** of both the subject and resources need to be considered to implement fine-grained access control. For example, should Call Center QA Managers be able to access call recordings from a particular call center, or do we want to restrict that based on geography?
- **Additional Data Feeds**, such as those described by NIST as inputs (e.g., alert management systems, UEBA, etc.), can be used to factor into the policy decisions. For example, an alert from an endpoint security agent that discovered malware can be used to deny access to critical company resources but allow access to the IT resources required to remediate the malware.

In addition to the traditional identity-based methods of defining access controls, Policy Management should consider more modern methods to determine whether access should be granted to a subject. Behavioral patterns, such as typing speed, time of day or historical access patterns, can be included in an Access Policy.

Further risk-based considerations, such as the sensitivity of the data being accessed (type of data or amount of data), as well as type of access (read, update, configuration change, ability to decrypt, etc.) will need to be considered in access policies.

Implementing the Policies

Once the policies are created, the organization must determine a method to implement the policies seamlessly and consistently.

Organizations need to understand where all the data points, required to make a policy decision, reside. Many organizations may face an interim step where they build enablers before implementing Policy Management. Organizations need to consider the capabilities it has on hand as well as the capabilities it will need to instantiate to implement the policies. For example, understanding the directory services and their interaction with human resource applications, sensing capabilities such as User Behavior Analytics (UEBA) and alerts from elastic log monitors can be used to implement effective access policies.

In addition, consider whether these data points can be retrieved in real time by the PDP to make access decisions or should they be consolidated into a system that can service the PDP in accordance with the access management requirements. Policies should be defined and refined in an incremental fashion with an eye on automation and incorporation into the Identity Life Cycle Management process.

Response Tuning
Response tuning allows for the optimization (e.g., performance) of an existing policy set, the fine tuning of policy outcome (e.g., allow or block), and the pre-flighting of new rules.

Optimization is about ensuring that the Policy Rule set can be efficiently and rapidly (sub-second determination of access / no access) traversed by the Policy Engine. This will be a bit of science where each vendor can provide guidance on optimizations for their rule's engine. Additionally, there will be a bit of art, where the senior engineers will leverage heuristics from the past as well as common sense (if you have a rule that denies 80% of access requests, make that the first rule you test).

Fine Tuning is about getting the precise results. The level of rule granularity will be determined by the various tools in use. While one can create very precise and arcane rules, consideration must be given to the level of administrative burden these rules will impose on the team. Additionally, is the greater precision actually resulting in lower risk to your environment? The search for balance between precision and administrative load will likely result in rules that get more specific for higher value resources in the environment.

Pre-flighting rules are about understanding the impact of changes in rule sets before they are deployed or enforced. Some organizations follow an SDLC / Change Control process whereby rules are implemented in lesser environments before they are implemented in the production environment. While that is necessary, it is only as reliable as the lesser environments reflect the production environments. Also, from an infrastructure environment – consider what borders actually exist between non-production and production environments. Other organizations may implement a new policy in monitor mode for a pre-defined duration before enabling enforcement. Pre-flight is an analysis of the likely impact of rules changes. Some tools may allow you to backtest the proposed rule changes by going through historic data, or looking at current network traffic, and displaying the differences in traffic flow pre and post rule changes. Ancillary to Pre-Flighting is to ensure that the organization establishes a comprehensive set of use cases with which to test.

Governance
Zero Trust Policy Governance is inextricable from both Identity Management and Risk Management. Identity governance includes areas such as:

Authoritative Identity Sources: A well-curated (read: strong life cycle management and process) data source for authoritative Identity. The identity store, which may be federated information (read: identity mesh) rather than a monolithic data store, will have attribute data about Subjects (e.g., roles, locations, etc.) that can be used in making Policy Decisions.

Identity Proofing and Identity Validation is about ensuring that, from the beginning of the lifecycle, the organization has assurance that the Subject is who they claim to be. That suggests a stronger identity proofing capability that originates in the Human Resources systems and flows into an integrated Identify Store. Identity Validation is making certain that the organization has various, more challenging methods to validate that the claimed Subject is the actual Subject. One method of validating identities is by leveraging MFA capabilities. Companies are beginning to look at behavioral biometric data as a way to strengthen Identity Validation capabilities.

Privileged Access and Entitlements Organizations need a mechanism to differentiate between standard and privileged accounts so that access decisions can be made in that light. Additionally, application entitlements need to be exposed to the Zero Trust Policy engine – so that the engine can make access decisions based upon differing application entitlements.

Threat and Risk Inputs such as data from external applications, UEBA SIEMs, UEBA and threat intelligence feeds.

Risk Information will come from a collection of resources – data sensitivity, application sensitivity, business impact analysis.

Risk Calculus – is the collection of business rules that will allow access or cause access to be terminated. An organization's risk appetite must also be included in the access decisions.

Management – is the establishment of a documented process about the lifecycle of rules – where they originate from, how they factor into risk scores, how changes are tested and promoted into production, how they are periodically reviewed for efficiency and efficacy, how rules are groomed for efficiency and how Cyber Operations / Fusion Centers handle rule exceptions and rule alerts (e.g., for terminated sessions, or a series of access denials from a single source that may be an IOC).

Enforcement Layer

Policy Enforcement Point

A Policy Enforcement Point (PEP) in a Zero Trust Architecture implementation layer that sits within the control plane that "enables, monitors and terminates connections between subjects and enterprise resources".[1] That is, a PEP intercepts the request for access from a subject to an enterprise resource; it sends the request to the Policy Decision Point (PDP) to be analyzed; it receives the response received from the PDP (allow, disallow, terminate); and then implemented the PDP decision and creates the connection or drops the connection request or terminates the connection. The PEP is a core logical component but can be composed of multiple components depending on the architecture deployed. For example, a PEP can be installed on client and in front of enterprise resources.

A key piece of Zero Trust Architecture in the enforcement layer is the ability to intercept and validate the security context of the subject for all access requests. This is in contrast with occasionally re-authenticating the subject's access based on a refresh token the subject has been granted. In this scheme, re-authentication does not validate the security context of the subject and serves only to extend the authentication session of the subject. Where re-authentication is utilized, the use of short-lived refresh tokens is strongly recommended (e.g., idle timeout for employee sessions). An FI needs to strike a balance between token lifetime and impact on subjects. Re-authentication contrasts with continuous authentication which is a tenet of Zero Trust Architecture.

Approaches to Zero Trust Architecture

The different styles of Zero Trust implementation will determine which PEP capabilities are needed. NIST identifies three architectural approaches: software defined perimeters, micros segmentation and enhanced identity governance

- Network Segmentation –this layer 7 approach (although can be implemented lower in the stack) leverages a software agent on the client and interacts with network infrastructure to create a connection between Subject and enterprise resource.
- Micro-Segmentation – which involves placing both subjects and enterprise resources into unique network segments protected by a gateway (software or network). This can leverage a gateway to allow or deny traffic between subjects and resources; or leverage a software agent on subject devices to work in conjunction with the network gateways.
    - In these enclave-like deployments, the gateway component can reside at the boundary of the resource collections
    - Alternatively, a resource-based portal could exist to serve as the enforcement/entry point to the collection of resources

- Identity Governance – leverages policies that rely on Identity and attributes assigned to an Identity and can include real-time device status to determine whether access is to be allowed or disallowed.

Variations of Zero Trust Architecture and Deployment
In practice, the PEP has a few deployment patterns and organizations can choose to deploy a combination of these architectural patterns. Client-side agent and a resource-side gateway. In this pattern, the device from which the subject is initiating a connection connects to a gateway on the enterprise resource, provided the Policy Engine allows it.

- Client-side agent and a gateway in front of an enclave holding a collection of enterprise resources. In this patter, the device from which the subject is initiating a connection connects to a gateway on the enterprise resource, provided the Policy Engine allows it
- Portal that acts as a gateway for the resources in a zone that sites behind said gateway.
- Sandboxing – where applications run compartmentalized in a sandbox and accept requests from the PEP but not each other.

Maturing from Macro to Micro Segments
In a Zero Trust Architecture, we look to improve security posture by moving layered controls closer to the resources instead of relying on the network perimeter itself. Rather than building a perimeter wall to keep the bad actors out, we move the controls closer to the resources we are protecting. This can introduce overhead, and the need arises for grouping resources into collections based upon several criteria.

- **Grouping Subjects**: Consideration should be given as to how to group Subjects. There are several methods to use such as grouping human users by job role/function or internal vs external and further demised by another attribute such as geography, customer base supported, business line, etc. These considerations also apply to groupings of services/APIs and assets. There are a variety of different ways to accomplish this, but the methods are highly reliant on data quality (completeness and accuracy) as well as life cycle management processes to continually maintain the groupings.
- **Grouping Resources**: Consideration should also be given to how resources (data, applications, systems, printers) can be grouped into collections. Throughout this document, we state that a Zero Trust deployment is about protecting each resource. This can seem daunting to organizations beginning on their zero-trust journey, so organizations can begin by grouping applications or resources. As an organization validates its data and processes, gradually it can increase the granularity of the resource groups and hit the ideal Zero Trust state of protecting each resource individually.

Information Layer

An information layer in a Zero Trust Architecture implementation needs to be a robust analytics layer allowing the Policy Enforcement Point (PEP) to make decisions based on information presented by a resource. This information usually contains information on what we know about the requestor, their session or the requestor's behaviors. All the data required to make quality and timely decisions need to be entered into the data layer.
Creating an information layer in can be done incrementally depending on the level of integration and the current phase in your Zero Trust architecture. There are multiple potential uses of the information layer:

- Fast-Check or rules-based decisions – Fast-check decisions are fast table-mapped decisions that allow for instant response to the PEP for allow/deny operations. This can include but isn't limited to disallowing access from certain geographic areas, IP networks or without the proper software installations. Rules based decisions can be implemented in many ways, but the result is a real-time allow/deny based on system policy rules.

- Thorough-Check or Modeling based scoring outcomes – Thorough-check decisions are longer-running and potential modeled decisions where more than one piece of data is collected, analyzed and a resulting "score" is generated. This can include but isn't limited to more complex AND/OR/XOR algorithms such as: allow geographic areas unless coming from a certain type of network AND certain software signatures are detected (foreign keyboard layouts, unusual time zones, etc.) AND a session parameter has just changed. The definition and creation of these algorithms are outside the scope of this document but can be categorized as thorough-check outcomes. Decision scientists and security analysts should be employed to help determine the right thorough-check decisions for your deployment.

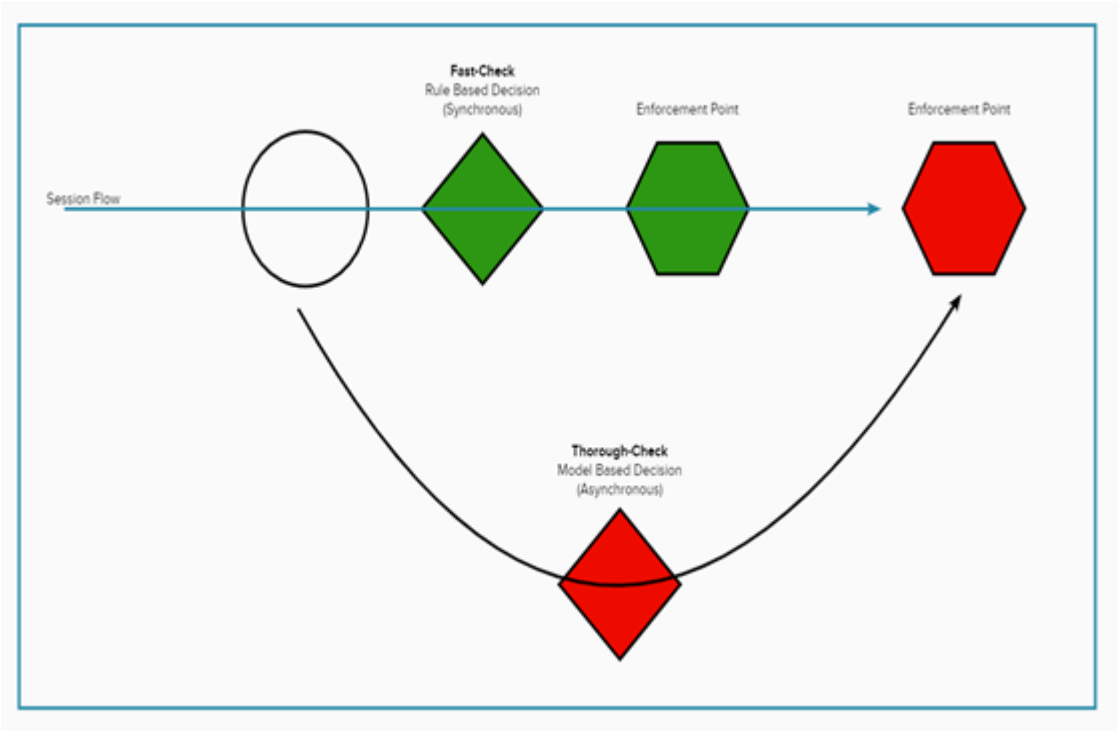


Figure 1.1: Session flow with synchronous and asynchronous decisions

To satisfy these uses, the information layer will need multiple things:

- Data warehousing and retrieval - the continual storage of information as it relates to a requestor's overt actions, covert actions, session information and the behaviors of the requestor. This information could be pulled both for fast-check decisions and for thorough-check decisions.
- Data analytics – a layer to perform mathematical calculations using the data within warehouse capabilities. This layer should have enough computing power to meet the requirements for your implementation. On-premise analytics engines and/or cloud alternatives can be employed based on your needs. The data analytics engines will also feed your enforcement policies and rules-based decisioning.
- Data Protection – this layer will include sensitive information that is used to directly feed the decision engine. This information must be protected to ensure the decision engine is making the correct decisions at the correct time.
- Threat intelligence feeds – These feeds (internal and external) will help inform the policy engine and threat algorithms. This information could come via multiple sources and will need to be available to the data analytics engine to produce the correct data outputs.

- Industry compliance systems (NIST SP 800-207) – Compliance considerations should be made for your industry. This will inform policy decisions and policy rules that are based on regulations your company must follow.

The following are core technologies or capabilities to create a robust information layer and better inform the PDP if the subject requesting access to the resource should be allowed. These include, but are not limited to:

User & Entity Behavior Analytics (UEBA) – Perhaps one of the most powerful and promising features of the Zero Trust decision layer is the incorporation of UEBA. The idea of mapping individual behavior patterns over time to compare to new activities of a credential and/or device can be very powerful. Human and machine patterns are very predictable for most cases. The more data points consumed by the UEBA component of the engine the better are the chances to identify anomalous behavior.

Like the context information described above, UEBA becomes stronger when you have more attributes to work with. There is a possibility of an attacker introducing new behaviors over time to make attack actions look like part of the norm. This is the risk of UEBA with continuous learning. However, it becomes more difficult to trick all different possible attributes consumed for behavior pattern creation.

UEBA can be implemented in simpler statistical form as discussed above in the decision engine models, or can be implemented using Machine Learning solutions, which are described below.

Lateral Movement Detection – Advanced Persistent Threats (APT) are a type of threat actor that is difficult to prevent and detect. These actors are well funded, technically savvy, innovative and have the patience and time required to explore and exploit resources.

This feature of the decision layer is designed to identify the correlation of actions between different accounts and servers that may indicate a malicious escalation of privilege. By coalescing telemetry from the network, endpoints and centralized controls, such as the zero-trust enforcement point(s), the decision layer can identify malicious actions or intent and deny access accordingly.

This is a process that will be part of the asynchronous decision process. The concept of synchronous and asynchronous decision is discussed in more detail in the architecture section. It refers to the fact that the engine will be doing real-time (synch) checks and some near real-time checks (asynchronous). This architectural approach balances the user experience with the ability to stop an attack as early as possible. Some data points must be validated in full before access is granted. Other data points may require more time to collect, such that waiting would significantly impact the user experience.

Lateral movement analysis aligns with the asynchronous decision process. This feature will be running continuously, like UEBA, and as they identify potential risk, they will share with the decision engine to take the necessary action.

Machine Learning (ML) – One of the most widely used buzzwords in the last several years and which became “the solution for all problems”. Considering that context, the authors still believe there is a strong value from ML for Zero Trust.

With ML, the output quality is directly related to data input quality supported by a robust information layer (described in the previous section). More importantly, the ability to have good training data and an ongoing feedback loop with false positives and false negatives to ensure quality output. Most ML algorithms are statistical models adapted to learn from a dataset and provide a statistical response to a new input. Hence the need for quality training data and fresh validated input.

The belief in the importance of ML for Zero Trust comes from the volume of information available in this complex battle between defense and attack. The ability to scale defense in context will depend on good ML algorithms.

All the information referenced above in the other features becomes the input to the ML model. By nature, a complete ML model will be tuned to provide output data in near real-time and for that reason be part of the asynchronous decision process. This is because of the computing power required to provide a meaningful answer. However, there is a possibility to implement a simplified model or rules-based analytics that could provide instantaneous decisions in real time.

Summary

- Implementing a Decision Layer is a long journey
- Zero Trust requires continual re-authentication
- Firewall or perimeter authentication on their own do not provide sufficient security guarantees
- Unless explicitly authorized, the decision is deny
- If authorized, the decision will be adjusted based on data, context, intent (behavior)
- Machine Learning and User & Entity Behavior Analytics could be powerful tools for continuously enhancing the Zero Trust methodology

Conclusion

Recent experience with bad actors and insider threats has made it clear that traditional perimeter protection is inadequate to the task of protecting client and business data and services. Bad actors in a Financial Institution's network should be assumed. Consequently, a more mature and sophisticated approach is needed to:

- Reduce the likelihood of lateral migration;
- Reduce the likelihood of privilege escalation; and
- Implementing finer grain access to enterprise resources.

Zero Trust Architecture – which puts *identity* at the center of security – helps FIs move in that direction by implementing:

- Stronger, continual authentication;
- Risk-based authentication to enterprise resources that includes attributes and real-time signals from the environment; and
- Fine grain access to enterprise resources.

Guidance from NIST as well as experienced cybersecurity practitioners in the financial services industry suggests that these controls and capabilities play a critical role in reducing malicious actors and insider threats. More complete control of which Subjects can connect to which enterprise resources and under what conditions will reduce risk to client and business data and facilitate the protection of enterprise applications from events that can impact the availability or integrity of services. Evolving from perimeter-based protection to Zero Trust will reduce reputation and financial risk to the institution and US banking system – and better protect our clients' private data.

In all Zero Trust frameworks, an approach to build an implementation plan starts with:

- An inventory and maturity rating of current capabilities;
- The identified capability gaps are specific to an enterprise and create the roadmap of technology and infrastructure changes required within an enterprise to move closer to Zero Trust.

Zero Trust is comprised of many different Capabilities that work together to continually challenge the authentication and authorization of communications within an enterprise. Zero Trust is not "one size fits all." Instead, it is a unique journey and depends on the implementing organization's specific risk appetite. An enterprise needs to define its Zero Trust strategy to adapt to the modern environment and current threat landscape while protecting data and devices.