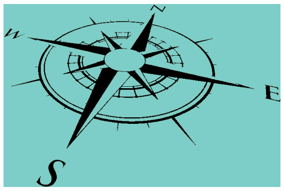
Zero Trust Architecture

Zero trust (ZT) is the term for an evolving set of cybersecurity paradigms that move network defenses from static, network-based perimeters to focus on users, assets, and resources. A zero trust architecture (ZTA) uses zero trust principles to plan enterprise infrastructure and workflows. Zero trust assumes there is no implicit trust granted to assets or user accounts based solely on their physical or network location (i.e., local area networks versus the internet). Authentication and authorization (both user and device) are discrete functions performed before a session to an enterprise resource is established. Zero trust is a response to enterprise network trends that include remote users and cloud-based assets that are not located within an enterprise owned network boundary. Zero trust focus on protecting resources, not network segments, as the network location is no longer seen as the prime component to the security posture of the resource...

Logical Components of Zero Trust Architecture — There are numerous logical components that make up a ZTA deployment in an enterprise. These components may be operated as an on-premises service or through a cloud-based service... Note that this is an ideal model showing logical components and their interactions... the policy decision point (PDP) is broken down into two logical components: the policy engine and policy administrator ... The ZTA logical components use a separate control plane to communicate, while application data is communicated on a data plane... [In this StratML rendition each of the two planes is documented as a goal, with the components of each documented as objectives.]

Contents

Vision....................3
Mission....................3
Values....................3
1. Networks....................6
1.1. Policy Engine....................6
1.2. Policy Administration....................6
1.3. Policy Enforcement....................6
2. Data....................7
2.1. Diagnostics & Mitigation....................7
2.2. Compliance....................7
2.3. Threat Intelligence....................7
2.4. Access....................7
2.4.1. Micro-Segmentation....................8
2.4.2. Software Defined Perimeters....................8
2.5. PKI....................8
2.6. IDs....................8
2.6.1. Policy Creation....................9
2.7. Logs....................9
2.8. SIEM....................9
Administrative Information....................10



National Institute of Standards and Technology (NIST)

Stakeholder(s):

Wilbur L. Ross, Jr. :
Secretary, U.S. Department of Commerce

Walter Copan :
NIST Director and Under Secretary of Commerce for Standards and Technology

Scott Rose :
Co-Author — Advanced Network Technologies Division, Information Technology Laboratory

Oliver Borchert :
Co-Author — Advanced Network Technologies Division, Information Technology Laboratory

Stu Mitchell :
Co-Author — Stu2Labs, Stafford, VA

Sean Connelly :
Co-Author — Cybersecurity & Infrastructure Security Agency, Department of Homeland Security

Federal CIO Council :
This document is the product of a collaboration between multiple federal agencies and is overseen by the Federal CIO Council. The architecture subgroup is responsible for development of this document, but there are specific individuals who deserve recognition. These include Greg Holden, Alper Kerman, and Douglas Montgomery.

Greg Holden :
project manager of the Federal CIO Council ZTA project

Alper Kerman :
project manager for the NIST/National Cybersecurity Center of Excellence ZTA effort

Douglas Montgomery

Enterprise Security Architects :
This document is intended to describe zero trust for enterprise security architects. It is meant to aid understanding of zero trust for civilian unclassified systems and provide a road map to migrate and deploy zero trust security concepts to an enterprise environment.

Cybersecurity Managers :
Agency cybersecurity managers, network administrators, and managers may also gain insight into zero trust and ZTA from this document.

Network Administrators

Enterprises :
It is not intended to be a single deployment plan for ZTA as an enterprise will have unique business use cases and data assets that require protection. Starting with a solid understanding of the organization's business and data will result in a strong approach to zero trust.

Vision

Resources are protected regardless of networks

Mission

To protect resources

Values

Tenets: Tenets of Zero Trust -- Many definitions and discussions of ZT stress the concept of removing wide-area perimeter defenses (e.g., enterprise firewalls) as a factor. However, most of these definitions continue to define themselves in relation to perimeters in some way (such as micro-segmentation or micro perimeters ...) as part of the functional capabilities of a ZTA. The following is an attempt to define ZT and ZTA in terms of basic tenets that should be involved rather than what is excluded. These tenets are the ideal goal, though it must be acknowledged that not all tenets may be fully implemented in their purest form for a given strategy. A zero trust architecture is designed and deployed with adherence to the following zero trust basic tenets:

Resources: 1. All data sources and computing services are considered resources. A network may be composed of several different classes of devices. A network may also have small footprint devices that send data to aggregators/ storage, software as a service (SaaS), systems sending instructions to actuators, and other functions. Also, an enterprise may decide to classify personally owned devices as resources if they can access enterprise owned resources.

Communication: 2. All communication is secured regardless of network location. Network location does not imply trust. Access requests from assets located on enterprise-owned network infrastructure (e.g., inside a traditional network perimeter) must meet the same security requirements as access requests and communication from any other nonenterprise-owned network. In other words, trust should not be automatically granted based on the device being on enterprise network infrastructure. All communication should be done in the most secure manner available, protect confidentiality and integrity, and provide source authentication.

Confidentiality

Integrity

Authentication

Trust: 3. Access to individual enterprise resources is granted on a per-session basis. Trust in the requester is evaluated before the access is granted. This could mean only “sometime previously” for this particular transaction and may not occur directly before initiating a session or performing a transaction with a resource. However, authentication and authorization to one resource will not automatically grant access to a different resource.

Access: 4. Access to resources is determined by dynamic policy—including the observable state of client identity, application, and the requesting asset—and may include other behavioral attributes. An organization protects resources by defining what resources it has, who its members are (or ability to authenticate users from a federated community), and what access to resources those members need. For zero trust, client identity includes the user account and any associated attributes assigned by the enterprise to that account or artifacts to authenticate automated tasks. Requesting asset state includes device characteristics such as software versions installed, network location, time/date of request, previously observed behavior, and installed credentials. Behavioral attributes include automated user analytics, device analytics, and measured deviations from observed usage patterns. Policy is the set of access rules based on attributes that an organization assigns to a user, data asset, or application. These rules and attributes are based on the needs of the business process and acceptable level of risk. Resource access and action permission policies can vary based on the sensitivity of the resource/data. Least privilege principles are applied to restrict both visibility and accessibility

Policy

Security: 5. The enterprise ensures that all owned and associated devices are in the most secure state possible and monitors assets to ensure that they remain in the most secure state possible. No device is inherently trusted. Here, “most secure state possible” means that the device is in the most practicable secure state and still performs the actions required for the mission. An enterprise implementing a ZTA should establish a CDM or similar system to monitor the state of devices and applications and should apply patches/fixes as needed. Devices that are discovered to be subverted, have known vulnerabilities, and/or are not managed by the enterprise may be treated differently (including denial of all connections to enterprise resources) than devices owned by or associated with the enterprise that are deemed to be in their most secure state. This may also apply to associated devices (e.g., personally owned devices) that may be allowed to access some resources but not others. This, too, requires a robust monitoring and reporting system in place to provide actionable data about the current state of enterprise resources.

Monitoring

Dynamics: 6. All resource authentication and authorization are dynamic and strictly enforced before access is allowed. This is a constant cycle of obtaining access, scanning and assessing threats, adapting, and continually reevaluating trust in ongoing communication. An enterprise implementing a ZTA would be expected to have Identity, Credential, and Access Management (ICAM) and asset management systems in place. This includes the use of multifactor authentication (MFA) for access to some or all enterprise resources. Continuous monitoring with possible reauthentication and reauthorization occurs throughout user interaction, as defined and enforced by policy (e.g., time-based, new resource requested, resource modification, anomalous user activity detected) that strives to achieve a balance of security, availability, usability, and cost-efficiency.

Enforcement

Information: 7. The enterprise collects as much information as possible about the current state of network infrastructure and communications and uses it to improve its security posture. An enterprise should collect data about network traffic and access requests, which is then used to improve policy creation and enforcement. This data can also be used to provide context for access requests from subjects (see Section 3.3.1).

Technology Agnosticism: The above tenets attempt to be technology agnostic. For example, “user identification (ID)” could include several factors such as username/password, certificates, and onetime password. These tenets apply to work done within an organization or in collaboration with one or more partner organizations and not to public or consumer-facing business processes. An organization cannot impose internal policies on external actors (e.g., customers or general internet users).

1. Networks

Control network communications

1.1. Policy Engine

Determine whether to grant access to resource for given subjects.

Policy engine (PE): This component is responsible for the ultimate decision to grant access to a resource for a given subject. The PE uses enterprise policy as well as input from external sources (e.g., CDM systems, threat intelligence services described below) as input to a trust algorithm (see Section 3.3 for more details) to grant, deny, or revoke access to the resource. The PE is paired with the policy administrator component. The policy engine makes and logs the decision, and the policy administrator executes the decision.

1.2. Policy Administration

Establish and/or shut down the communication path between a subject and a resource.

Policy administrator (PA): This component is responsible for establishing and/or shutting down the communication path between a subject and a resource. It would generate any authentication and authentication token or credential used by a client to access an enterprise resource. It is closely tied to the PE and relies on its decision to ultimately allow or deny a session. Some implementations may treat the PE and PA as a single service; here, it is divided into its two logical components. The PA communicates with the PEP when creating the communication path. This communication is done via the control plane.

1.3. Policy Enforcement

Enable, monitor, and terminating connections between subjects and enterprise resources.

Policy enforcement point (PEP): This system is responsible for enabling, monitoring, and eventually terminating connections between a subject and an enterprise resource. This is a single logical component in ZTA but may be broken into two different components: the client (e.g., agent on user’s laptop) and resource side (e.g., gateway component in front of resource that controls access) or a single portal component that acts as a gatekeeper for communication paths. Beyond the PEP is the implicit trust zone (see Section 2) hosting the enterprise resource.

2. Data

Provide input and policy rules used by the policy engine when making access decisions.

In addition to the core components in an enterprise implementing a ZTA, several data sources provide input and policy rules used by the policy engine when making access decisions. These include local data sources as well as external (i.e. , nonenterprise-controlled or -created) data sources. These include:

2.1. Diagnostics & Mitigation

Gather information about the enterprise asset's current state and applies updates to configuration and software components.

Continuous diagnostics and mitigation (CDM) system: This gathers information about the enterprise asset's current state and applies updates to configuration and software components. An enterprise CDM system provides the policy engine with the information about the asset making an access request, such as whether it is running the appropriate patched operating system (OS) and applications or whether the asset has any known vulnerabilities.

2.2. Compliance

Comply with relevant regulatory regimes.

Industry compliance system: This ensures that the enterprise remains compliant with any regulatory regime that it may fall under (e.g., FISMA, healthcare or financial industry information security requirements). This includes all the policy rules that an enterprise develops to ensure compliance.

2.3. Threat Intelligence

Provide information to help the policy engine make access decisions.

Threat intelligence feed(s): This provides information from internal or external sources that help the policy engine make access decisions. These could be multiple services that take data from internal and/or multiple external sources and provide information about newly discovered attacks or vulnerabilities. This also includes blacklists, newly identified malware, and reported attacks to other assets that the policy engine will want to deny access to from enterprise assets.

2.4. Access

Define the attributes, rules, and policies about access to enterprise resources.

Data access policies: These are the attributes, rules, and policies about access to enterprise resources. This set of rules could be encoded in or dynamically generated by the policy engine. These policies are the starting point for authorizing access to a resource as they provide the basic access privileges for accounts and applications in the enterprise. These policies should be based on the defined mission roles and needs of the organization.

2.4.1. Micro-Segmentation

Place individual or groups of resources on their own network segments protected by a gateway security components.

ZTA Using Micro-Segmentation — An enterprise may choose to implement a ZTA based on placing individual or groups of resources on its own network segment protected by a gateway security component. In this approach, the enterprise places NGFWs or gateway devices to act as PEPs protecting each resource or group of resources. These gateway devices dynamically grant access to individual requests from a client asset. Depending on the model, the gateway may be the sole PEP component or part of a multipart PEP consisting of the gateway and client-side agent (see Section 3.2.1). This approach applies to a variety of use cases and deployment models as the protecting device acts as the PEP, with management of said devices acting as the PE/PA component. This approach requires an identity governance program to fully function but relies on the gateway components to act as the PEP that shields resources from unauthorized access and/or discovery. The key necessity to this approach is that the PEP components are managed and should be able to react and reconfigure as needed to respond to threats or change in the workflow. It is possible to implement some features of a micro-segmented enterprise by using less advanced gateway devices and even stateless firewalls, but the administration cost and difficulty to quickly adapt to changes make this a very poor choice.

2.4.2. Software Defined Perimeters

Use the network infrastructure to implement a ZTA.

ZTA Using Network Infrastructure and Software Defined Perimeters — The third approach uses the network infrastructure to implement a ZTA. The ZT implementation could be achieved by using an overlay network (i.e., layer 7 but also could be set up lower of the ISO network stack). These approaches are sometimes referred to as software defined perimeter (SDP) approaches and frequently include concepts from SDN [SDNBOOK] and intent-based networking (IBN) [IBNVN]. In this approach, the PA acts as the network controller that sets up and reconfigures the network based on the decisions made by the PE. The clients continue to request access via PEPs, which are managed by the PA component. When the approach is implemented at the application network layer (i.e., layer 7), the most common deployment model is the agent/gateway (see Section 3.2.1). In this implementation, the agent and resource gateway (acting as the single PEP and configured by the PA) establish a secure channel used for communication between the client and resource.

2.5. PKI

Generate and log certificates issued to resources, subjects, and applications.

Enterprise public key infrastructure (PKI): This system is responsible for generating and logging certificates issued by the enterprise to resources, subjects, and applications. This also includes the global certificate authority ecosystem and the Federal PKI, which may or may not be integrated with the enterprise PKI. This could also be a PKI that is not built upon X.509 certificates.

2.6. IDs

Create, store, and manage user accounts and identity records.

ID management system: This is responsible for creating, storing, and managing enterprise user accounts and identity records (e.g., lightweight directory access protocol (LDAP) server). This system contains the necessary user information (e.g., name, email address, certificates) and other enterprise characteristics such as role, access attributes, and assigned assets. This system often utilizes other systems (such as a PKI) for artifacts associated with user accounts. This system may be part of a larger federated community and may include nonenterprise employees or links to nonenterprise assets for collaboration.

2.6.1. Policy Creation

Use the identity of actors as the key component of policy creation.

ZTA Using Enhanced Identity Governance — The enhanced identity governance approach to developing a ZTA uses the identity of actors as the key component of policy creation. If it were not for subjects requesting access to enterprise resources, there would be no need to create access polices. For this approach, enterprise resource access policies are based on identity and assigned attributes. The primary requirement for resource access is based on the access privileges granted to the given subject. Other factors such as device used, asset status, and environmental factors may alter the final confidence level calculation (and ultimate access authorization) or tailor the result in some way, such as granting only partial access to a given data source based on network location. Individual resources or PEP components protecting the resource must have a way to forward requests to a policy engine service or authenticate the subject and approve the request before granting access. Enhanced identity governance-based approaches for enterprises are often found using an open network model or an enterprise network with visitor access or frequent nonenterprise devices on the network (such as with the use case in Section 4.3 below). Network access is initially granted to all assets with access to resources that are restricted to identities with the appropriate access privileges. The identity-driven approach works well with the resource portal model since device identity and status provide secondary support data to access decisions. Other models work as well, depending on policies in place.

2.7. Logs

Aggregate asset logs, network traffic, resource access actions, and other events that provide feedback.

Network and system activity logs: This is the enterprise system that aggregates asset logs, network traffic, resource access actions, and other events that provide real-time (or near-real-time) feedback on the security posture of enterprise information systems.

2.8. SIEM

Collect security centric information for analysis.

Security information and event management (SIEM) system: This collects security centric information for later analysis. This data is then used to refine policies and warn of possible attacks against enterprise assets.

Administrative Information

Start Date:
End Date:

Publication Date: **2020-05-24**
Source: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-207-draft2.pdf

Submitter:

Given Name: Owen
Surname: Ambur
Email: Owen.Ambur@verizon.net
Phone: