## Draft NIST Special Publication 800-207

# Zero Trust Architecture

Scott Rose
Oliver Borchert
Stu Mitchell
Sean Connelly

This publication is available free of charge from:
https://doi.org/10.6028/NIST.SP.800-207-draft

COMPUTER SECURITY

National Institute of
Standards and Technology
U.S. Department of Commerce

## Draft NIST Special Publication 800-207

# Zero Trust Architecture

Scott Rose
Oliver Borchert
*Advanced Network Technologies Division*
*Information Technology Laboratory*

Stu Mitchell
*Stu2Labs*
*Stafford, VA*

Sean Connelly
*Department of Homeland Security*

This publication is available free of charge from:
https://doi.org/10.6028/NIST.SP.800-207-draft

September 2019

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

National Institute of Standards and Technology
*Walter Copan, NIST Director and Under Secretary of Commerce for Standards and Technology*

## Authority

This publication has been developed by NIST in accordance with its statutory responsibilities under the Federal Information Security Modernization Act (FISMA) of 2014, 44 U.S.C. § 3551 *et seq.*, Public Law (P.L.) 113-283. NIST is responsible for developing information security standards and guidelines, including minimum requirements for federal information systems, but such standards and guidelines shall not apply to national security systems without the express approval of appropriate federal officials exercising policy authority over such systems. This guideline is consistent with the requirements of the Office of Management and Budget (OMB) Circular A-130.

Nothing in this publication should be taken to contradict the standards and guidelines made mandatory and binding on federal agencies by the Secretary of Commerce under statutory authority. Nor should these guidelines be interpreted as altering or superseding the existing authorities of the Secretary of Commerce, Director of the OMB, or any other federal official. This publication may be used by nongovernmental organizations on a voluntary basis and is not subject to copyright in the United States. Attribution would, however, be appreciated by NIST.

National Institute of Standards and Technology Special Publication 800-207
Natl. Inst. Stand. Technol. Spec. Publ. 800-207, 49 pages (September 2019)
CODEN: NSPUE2

This publication is available free of charge from:
https://doi.org/10.6028/NIST.SP.800-207-draft

Certain commercial entities, equipment, or materials may be identified in this document in order to describe an experimental procedure or concept adequately. Such identification is not intended to imply recommendation or endorsement by NIST, nor is it intended to imply that the entities, materials, or equipment are necessarily the best available for the purpose.

There may be references in this publication to other publications currently under development by NIST in accordance with its assigned statutory responsibilities. The information in this publication, including concepts and methodologies, may be used by federal agencies even before the completion of such companion publications. Thus, until each publication is completed, current requirements, guidelines, and procedures, where they exist, remain operative. For planning and transition purposes, federal agencies may wish to closely follow the development of these new publications by NIST.

Organizations are encouraged to review all draft publications during public comment periods and provide feedback to NIST. Many NIST cybersecurity publications, other than the ones noted above, are available at https://csrc.nist.gov/publications.

**Public comment period: *September 23, 2019* through *November 22, 2019***

National Institute of Standards and Technology
Attn: Advanced Network Technologies Division, Information Technology Laboratory
100 Bureau Drive (Mail Stop 8920) Gaithersburg, MD 20899-8920
Email: zerotrust-arch@nist.gov

All comments are subject to release under the Freedom of Information Act (FOIA).

## Reports on Computer Systems Technology

The Information Technology Laboratory (ITL) at the National Institute of Standards and Technology (NIST) promotes the U.S. economy and public welfare by providing technical leadership for the Nation's measurement and standards infrastructure. ITL develops tests, test methods, reference data, proof of concept implementations, and technical analyses to advance the development and productive use of information technology. ITL's responsibilities include the development of management, administrative, technical, and physical standards and guidelines for the cost-effective security and privacy of other than national security-related information in federal information systems. The Special Publication 800-series reports on ITL's research, guidelines, and outreach efforts in information system security, and its collaborative activities with industry, government, and academic organizations.

## Abstract

Zero Trust is the term for an evolving set of network security paradigms that move network defenses from wide network perimeters to narrowly focusing on individual or small groups of resources. A Zero Trust Architecture (ZTA) strategy is one where there is no implicit trust granted to systems based on their physical or network location (i.e., local area networks vs. the Internet). Access to data resources is granted when the resource is required, and authentication (both user and device) is performed before the connection is established. ZTA 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. ZTA focuses 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. This document contains an abstract definition of ZTA and gives general deployment models and use cases where ZTA could improve an enterprise's overall IT security posture.

## Keywords

architecture; cybersecurity; enterprise; network security; zero trust.

## Acknowledgments

This document is the product of a collaboration of multiple federal agencies and overseen by the Federal CIO Council. The Architecture sub-group is responsible for the development of this document, but there are specific individuals who deserve recognition. These include Greg Holden, the project manager of the Federal CIO Council ZTA project, and Alper Kerman, the project manager for the NIST/NCCoE ZTA effort and Douglas Montgomery.

## Audience

This document is intended to be a description of ZTA strategies for enterprise network architects. The document is meant to aid understanding of ZTA for civilian unclassified systems and provide a roadmap to migrate and deploy ZTA concepts to an enterprise network. Agency cybersecurity managers, network administrators, and managers may also gain insight into ZTA from this document. This document 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 your organization's business and data will result in a strong approach to zero trust.

## Note to Reviewers

The purpose of this Special Publication is to develop a technology-neutral set of terms, definitions, and logical components of network infrastructure using a ZTA strategy. This document does not give specific guidance or recommendations on how to deploy zero trust components in an enterprise. Reviewers are asked to tailor their comments based on the stated purpose of the document.

## Trademark Information

## Call for Patent Claims

This public review includes a call for information on essential patent claims (claims whose use would be required for compliance with the guidance or requirements in this Information Technology Laboratory (ITL) draft publication). Such guidance and/or requirements may be directly stated in this ITL Publication or by reference to another publication. This call also includes disclosure, where known, of the existence of pending U.S. or foreign patent applications relating to this ITL draft publication and of any relevant unexpired U.S. or foreign patents.

ITL may require from the patent holder, or a party authorized to make assurances on its behalf, in written or electronic form, either:

a) assurance in the form of a general disclaimer to the effect that such party does not hold and does not currently intend holding any essential patent claim(s); or

b) assurance that a license to such essential patent claim(s) will be made available to applicants desiring to utilize the license for the purpose of complying with the guidance or requirements in this ITL draft publication either:

i) under reasonable terms and conditions that are demonstrably free of any unfair discrimination; or

ii) without compensation and under reasonable terms and conditions that are demonstrably free of any unfair discrimination.

Such assurance shall indicate that the patent holder (or third party authorized to make assurances on its behalf) will include in any documents transferring ownership of patents subject to the assurance, provisions sufficient to ensure that the commitments in the assurance are binding on the transferee, and that the transferee will similarly include appropriate provisions in the event of future transfers with the goal of binding each successor-in-interest.

The assurance shall also indicate that it is intended to be binding on successors-in-interest regardless of whether such provisions are included in the relevant transfer documents.

Such statements should be addressed to: zerotrust-arch@nist.gov

## Table of Contents

## List of Appendices

## List of Figures

## List of Tables

# 1 Introduction

A typical enterprise's network infrastructure has grown increasingly complex. A single enterprise may operate several internal networks, remote offices with their own local infrastructure, remote and/or mobile individuals, and cloud services. This complexity has outstripped traditional methods of perimeter-based network security as there is no single, easily identified perimeter for the enterprise.

This complex enterprise has led to a new way to plan enterprise network security known as Zero Trust Architecture (ZTA). A ZTA approach is primarily focused on data protection but can be expanded to include all enterprise assets. ZTA assumes the network is hostile and that an enterprise-owned network infrastructure is no different—or no more secure—than any non-enterprise owned network. In this new paradigm, an enterprise must continuously analyze and evaluate the risks to their internal assets and business functions and then enact protections to mitigate these risks. In ZTA, these protections usually involve minimizing access to resources to only those who are validated as needing access and continuously authenticating the identity and security posture of each access request.

This publication provides a definition of ZTA, its logical components, possible deployment scenarios, and threats. It also presents a general roadmap for organizations wishing to migrate to a ZTA-centered network infrastructure and discusses relevant federal policies that may impact or influence a zero trust architecture.

ZTA is not a single network architecture but a set of guiding principles in network infrastructure design and operation that can be used to improve the security posture of any classification or sensitivity level. Transitioning to ZTA is a journey and cannot be accomplished without a wholesale replacement of technology. That said, many organizations already have elements of a ZTA in their enterprise infrastructure today. Organizations should seek to incrementally implement zero trust principles, process changes, and technology solutions that protect its data assets and business functions. Most enterprise infrastructures will operate in a hybrid Zero Trust/Legacy mode during this time while continuing to invest in ongoing IT modernization initiatives and improving organization business processes.

Organizations need to implement effective information security and resiliency practices for zero trust to be effective. When complemented with existing cybersecurity policies and guidance, identity and access management, continuous monitoring, and general cybersecurity, ZTA can reinforce an organization's security posture using a managed risk approach and protect against common threats.

## 1.1 Background

The concept of zero trust has been present in cybersecurity since before the actual term "zero trust" was coined. The work of the Jericho Forum publicized the idea of limiting implicit trust based on network location and the limitations of relying on static defenses [JERICHO]. The concepts in de-perimeterization evolved and improved into a larger concept that became known as zero trust. Later, Jon Kindervag coined the term “Zero Trust”[1] while at Forrester[2] (now at Palo Alto Networks). This work included key concepts and a zero trust network architecture model that improved upon the concepts discussed in the Jericho Forum.

In many ways, federal agencies have been moving to network security based on zero trust principles for over a decade. Federal agencies have been building capabilities and policies starting with the Federal Information Security Management Act (FISMA) followed by the Risk Management Framework (RMF); Federal Identity, Credential, and Access Management (FICAM); Trusted Internet Connection (TIC); and Continuous Diagnostics and Mitigation (CDM) programs. All of these programs aim to restrict data and resource access to authorized parties. When these programs were started, they were limited by the technical capabilities of information systems. Security policies were largely static and were enforced at large “choke points” that an enterprise could control to get the largest effect for the effort. As technology matures, it is becoming possible to continuously analyze and evaluate access requests in a dynamic and granular fashion.

## 1.2 Structure of this Document

The rest of the document is broken down as follows:

- **Section 2**: Defines ZTA and lists some network assumptions when designing ZTA enterprise networks. This section also includes a list of the tenets of ZTA design.
- **Section 3**: Documents the logical components, or building blocks, of a ZTA. It is possible that unique implementations compose ZTA components differently yet serve the same logical functionality.
- **Section 4**: Lists some possible use cases where ZTA may make enterprise networks more secure and less prone to successful exploitation. These include enterprises with remote employees, cloud services, guest networks, etc.
- **Section 5**: Discusses the threats to an enterprise using a ZTA strategy. Many of these threats are similar to more traditionally architected networks but may require different mitigation techniques.
- **Section 6**: Discusses how ZTA tenets fit into and/or complement existing guidance for federal agencies.
- **Section 7**: Presents the starting point for transitioning an enterprise (such as a federal agency) to a ZTA. This includes a description of the general steps needed to plan and deploy applications and network infrastructure that are guided by ZTA tenets.

[1] https://www.paloaltonetworks.com/resources/videos/zero-trust

[2] Any mention of commercial products or services within NIST documents is for information only; it does not imply recommendation or endorsement by NIST.

## 2 Zero Trust Network Architecture

Zero Trust Architecture is an end-to-end approach to network/data security that encompasses identity, credentials, access management, operations, endpoints, hosting environments, and the interconnecting infrastructure. Zero trust is an architectural approach that is focused on data protection. The initial focus should be on restricting resource access to those with a "need to know." Traditionally, agencies (and enterprise networks in general) have focused on perimeter defense, and authorized users are given broad access to resources. As a result, unauthorized lateral movement within a network has been one of the biggest challenges for federal agencies. The Trusted Internet Connections (TIC) and agency perimeter firewalls provide strong Internet gateways. This helps block attackers from the Internet, but the TICs and perimeter firewalls are less useful for detecting and blocking attacks from inside the network.

An operative definition of ZTA is as follows:

> *Zero Trust Architecture* (ZTA) provides a collection of concepts, ideas, and component relationships (architectures) designed to eliminate the uncertainty in enforcing accurate access decisions in information systems and services.

This definition focuses on the crux of the issue, which is to *eliminate unauthorized access to data and services*, coupled with making the *access control enforcement as granular a*s *possible*. That is, authorized and approved subjects (user/machine) can access the data to the exclusion of all other subjects (i.e., attackers). To take this one step further, the word "resource" can be substituted for "data" so that ZTA is about resource access (e.g., printers, compute resources, IoT actuators, etc.) and not just data access.

In order to lessen uncertainties (as they cannot be totally eliminated), the focus is on authentication, authorization, and shrinking implied trust zones while minimizing temporal delays in network authentication mechanisms. Access rules are restricted to least privilege and made as granular as possible.

In Figure 1, a user or machine needs access to an enterprise resource. Access is granted through a Policy Decision Point (PDP) and corresponding Policy Enforcement Point (PEP).

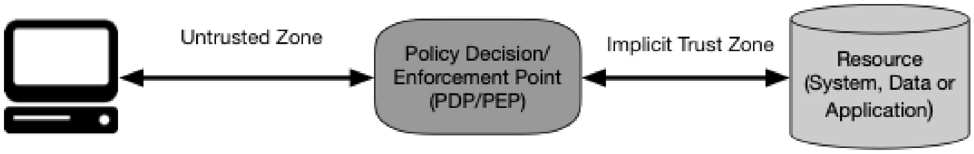
Untrusted Zone
Policy Decision/ Enforcement Point (PDP/PEP)
Implicit Trust Zone
Resource (System, Data or Application)

**Figure 1: Zero Trust Access**

The system must ensure the user is "trustworthy" and the request is valid. The PDP/PEP passes proper judgment to allow the subject to access the resource. This implies that Zero Trust applies for two basic areas: authentication and authorization. Can the system remove sufficient doubt about the user's true identity? Is the user justified in their access request? Is the device used for the request trustworthy? Overall, enterprises need to develop risk-based policies for resource access and set up a system to ensure that these policies are executed correctly. This means that an enterprise should not rely on implied trustworthiness, wherein if the user has met a base authentication level (i.e., logging into a system), all resource requests are assumed to be equally valid.

The "Implied Trust Zone" represents an area where all the entities are trusted to at least the level of the last PDP/PEP gateway. For example, consider the passenger screening model in an airport. All passengers pass through the airport security check point (PDP/PEP) to access the boarding gates. The passengers mill about in the terminal area and all the cleared passengers have a common trust level. In this model, the implied trust zone is the boarding area.

The PDP/PEP applies a common set of controls such that all traffic beyond the checkpoint has a common level of trust. The PDP/PEP cannot apply policy beyond its location in the flow of traffic. In order to allow the PDP/PEP to be as specific as possible, the Implied Trust Zone has to be as small as possible.

Zero Trust Architecture provides technology and capabilities to allow the PDP/PEPs to move closer to the resource. The idea is to authenticate and authorize every single flow in the network from actor (or application) to data.

## 2.1 Tenets of Zero Trust Architecture

Many definitions and discussions of ZTN/ZTA stress the concept of removing perimeter defenses (e.g., firewalls, etc.) from the equation. However, most continue to define themselves in relation to perimeters in some way (such as micro-segmentation or micro-perimeters). The following is an attempt to define ZTA in terms of basic tenets that should be involved, not what is excluded.

A *Zero Trust Architecture* is designed and deployed adhering to the following basic tenets:

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, systems sending instructions to actuators, etc. Also, an enterprise may decide to classify personally-owned devices as resources if they are allowed to access enterprise-owned resources.
2. **All communication is secure regardless of network location**. Network location does not imply trust. Access requests from systems located on enterprise-owned network infrastructure (e.g., inside a legacy network perimeter) must meet the same security requirements as access requests and communication from any other non-enterprise owned network. In other words, there should not be any trust automatically granted based on the device being on enterprise network infrastructure. All communication should be done in a secure manner (i.e., encrypted and authenticated).
3. **Access to individual enterprise resources is granted on a per-connection 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 connection with a resource. However, authentication to one resource will not automatically grant access to a different resource.

4. **Access to resources is determined by policy, including the observable state of user identity and the requesting system, and may include other behavioral attributes.** An organization protects resources by defining what resources it has, who its members are, and what access to resources those members need. User identity includes the network account used and any associated attributes assigned by the enterprise to that account. Requesting system state includes device characteristics such as software versions installed, network location, previously observed behavior, installed credentials, etc. Behavioral attributes include automated user analytics, device analytics, and measured deviations from observed usage patterns. Policy is the set of attributes an organization assigns to a user, data asset, or application. These attributes are based on the needs of the business process and acceptable level of risk. Resource access policies can vary based upon the sensitivity of the resource/data. Least privilege principles are applied in order to restrict both visibility and accessibility.
5. **The enterprise ensures all owned and associated systems are in the most secure state possible and monitors systems to ensure that they remain in the most secure state possible.** An enterprise implementing a ZTA strategy should establish a Continuing Diagnostics and Mitigation (CDM) program to monitor the state of systems and apply patches/fixes as needed. Systems that are discovered to be subverted, vulnerable, and/or non-enterprise-owned may be treated differently (including denial of all connections to enterprise resources) than systems owned by or associated with the enterprise that are deemed to be in their most secure state.
6. **User authentication is dynamic and strictly enforced before access is allowed.** This is a constant cycle of access, scanning and assessing threats, adapting, and continuously authenticating. An enterprise implementing a ZTA strategy has a user provisioning system in place and uses the system to authorize access to resources. This includes the use of multi-factor authentication (MFA) for access to some (or all) enterprise resources. Continuous monitoring and re-authentication occur throughout user interaction, as defined and enforced by policy (e.g., time-based, new resource requested, resource modification, etc.) that strives to achieve a balance of security, availability, usability, and cost-efficiency.

The above tenets attempt to be as technology-agnostic as possible. For example, "network ID" could include several factors such as username/password, certificates, one-time password, or some other identification.

## 2.2 A Zero Trust View of a Network

There are some basic assumptions for network connectivity for any organization that utilizes ZTA in network planning and deployment. Some of these assumptions apply to enterprise-owned network infrastructure, and some apply to enterprise-owned resources used on non-enterprise-owned network infrastructure (e.g., public WiFi). The network in an enterprise implementing a ZTA strategy should be developed with the ZTA tenets outlined above and with the following assumptions.

### 2.2.1 Assumptions for Enterprise-Owned Network Infrastructure

1. **The enterprise private network is not trustworthy.** Systems should always act as if an attacker is present on the enterprise network, and communication should be done in a

secure manner (see Tenet 2 above). This entails actions such as authenticating all connections and encrypting all traffic.

2. **Devices on the network may not be owned or configurable by the enterprise.** Visitors and/or contracted services may include non-enterprise-owned systems that need network access in order to perform their role. This also includes bring-your-own-device (BYOD) policies that allow enterprise users to use non-enterprise-owned devices to access enterprise resources.
3. **No device is inherently trusted.** Every device must authenticate itself (either to resource or PEP) before connecting to an enterprise-owned resource (see Tenet 6 above). Enterprise-owned devices can have artifacts that enable authentication and provide a higher trust score (see Section 3.2) than the same request coming from non-enterprise-owned devices. User credentials are insufficient for device authentication to an enterprise resource.

### 2.2.2 Assumptions for Non-Enterprise-Owned Network Infrastructure

1. **Not all enterprise resources are on enterprise-owned infrastructure.** This includes remote users as well as cloud services. The enterprise must be able to monitor, configure, and patch any system, but any system may rely on the local (i.e., non-enterprise) network for basic connectivity and network services (e.g., DNS, etc.).
2. **Remote enterprise users cannot trust the local network connection.** Remote users should assume the local (i.e., non-enterprise-owned) network is hostile. Systems should assume all traffic is being monitored and potentially modified. All connection requests should be authenticated, and all traffic should be encrypted (see the Tenets of ZTA above).

# 3 Zero Trust Architecture Logical Components

There are numerous logical components that make up a ZTA network deployment in an enterprise. These components may be operated as an on-premises service or through a cloud-based service. The conceptual framework model in Figure 2 shows the basic relationship of the components and their interactions. Note that this is an ideal model showing logical components and their interactions. From Figure 1, the Policy Decision Point (PDP) is broken down into two logical components: the Policy Engine and Policy Administrator (defined below).

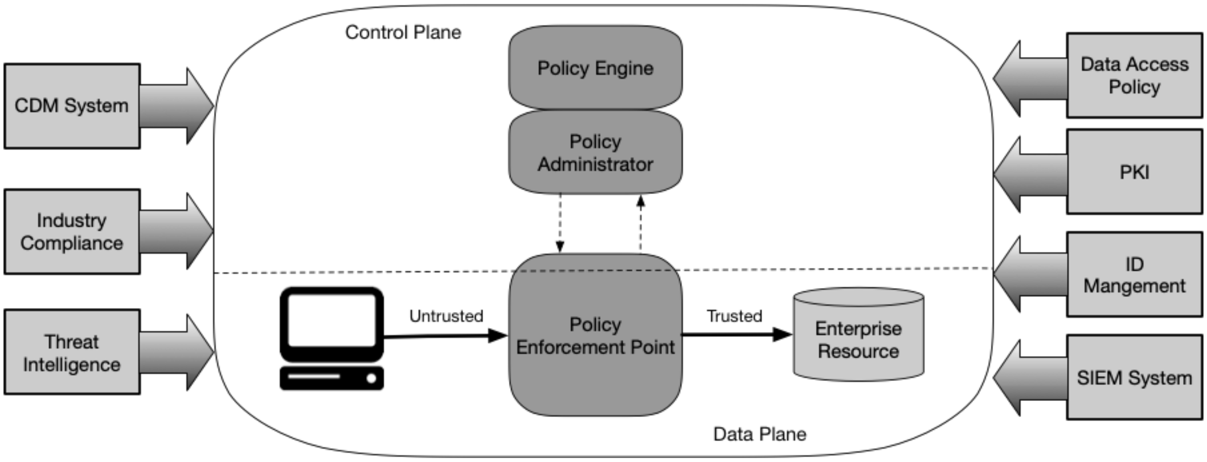

Figure 2: Core Zero Trust Logical Components

The component descriptions:

- **Policy Engine (PE):** This component is responsible for the ultimate decision to grant access to a resource for a given client or subject. The Policy Engine uses enterprise policy as well as input from external sources (e.g., IP blacklists, threat intelligence services) as input to a "trust algorithm" to decide to grant or deny access to the resource. The Policy Engine is paired with the Policy Administrator component. The Policy Engine makes (and logs) the decision, and the Policy Administrator executes the decision (approval or denial).
- **Policy Administrator (PA):** This component is responsible for establishing the connection between a client and a resource. It would generate any authentication token or credential used by a client to access an enterprise resource. It is closely tied to the Policy Engine and relies on its decision to ultimately allow or deny the connection. Implementations may treat the Policy Engine and Policy Administrator as a single service; here, it is divided into its two logical components. The PA communicates with the Policy Enforcement Point (PEP) when creating the connection. This communication is done via the control plane.
- **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 up 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 connections.

In addition to the core components in an enterprise implementing a ZTA strategy, there are several data sources that 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., non-enterprise controlled or created) data sources. These include:

- **Continuous Diagnostics and Mitigation (CDM) System(s):** This system(s) gathers information about the enterprise system's current state and applies updates to configuration and software components. An enterprise CDM system provides the Policy Engine with the information about the system making an access request, such as whether it is running the appropriate patched OS and applications or whether the system has any known vulnerabilities.
- **Industry Compliance System:** This system ensures that the enterprise remains compliant with any regulatory regime they may fall under (e.g. FISMA, HIPAA, PCI-DSS, etc.). This includes all the policy rules an enterprise develops to ensure compliance.
- **Threat Intelligence Feed(s):** This system provides information from outside sources that help the Policy Engine make access decisions. These could be multiple services that take data from multiple external sources and provide information about newly discovered attacks or vulnerabilities. This also includes DNS blacklists, discovered malware, or command and control systems that the Policy Engine will want to deny access to from enterprise systems.
- **Data Access Policies:** This is the set of attributes, rules, and policies about data access created by the enterprise around enterprise resources. This set of rules could be encoded in the Policy Engine or dynamically generated by the PE. These policies are the starting point for granting access to a resource as they provide the basic access privileges for actors and applications in the enterprise. These roles and access rules should be based on user roles and the mission needs of the organization.
- **Enterprise Public Key Infrastructure (PKI):** This system is responsible for generating and logging certificates issued by the enterprise to resources, actors, and applications. This also includes the global CA ecosystem and the Federal PKI[3], which may or may not be integrated with the enterprise PKI.
- **ID Management System:** This system is responsible for creating, storing, and managing enterprise user accounts and identity records. This system contains the necessary user information (e.g., name, email address, certificates, etc.) and other enterprise characteristics such as role, access attributes, or assigned systems. This system often utilizes other systems (such as a PKI above) for artifacts associated with user accounts.
- **Security Incident and Event Management (SIEM) System:** The enterprise system that aggregates system logs, network traffic, resource entitlements, and other events that

[3] https://www.idmanagement.gov/topics/fpki/

provide feedback on the security posture of enterprise information systems. This data is then used to refine policies and warn of possible active attacks against enterprise systems.

## 3.1 Deployed Variations of the Abstract Architecture

All of these components are logical components. They do not necessarily need to be unique systems. A single system may perform the duties of multiple logical components, and likewise, a logical component may consist of multiple hardware or software elements to perform the tasks. For example, an enterprise PKI may consist of one component responsible for issuing certificates for devices and another used for issuing certificates to end users, but both use intermediate certificates issued from the same enterprise root certificate authority. In many ZTA network offerings currently available on the market, the PE and PA components are combined in a single service.

There are several variations on the deployment of selected components of the architecture that are outlined in the sections below. Depending on how an enterprise network is set up, multiple ZTA deployment models may be in use for different business processes in one enterprise.

### 3.1.1 Device Agent/Gateway-Based Deployment

In this deployment model, the PEP is divided into two components that reside on the resource or as a component directly in front of a resource. For example, each enterprise-issued system has an installed device agent that coordinates connections, and each resource has a component (i.e., gateway) that is placed directly in front so that the resource only communicates with the gateway, essentially serving as a reverse proxy for the resource. The gateway is responsible for connecting to the Policy Administrator and only allows approved connections configured by the Policy Administrator (see Figure 3).

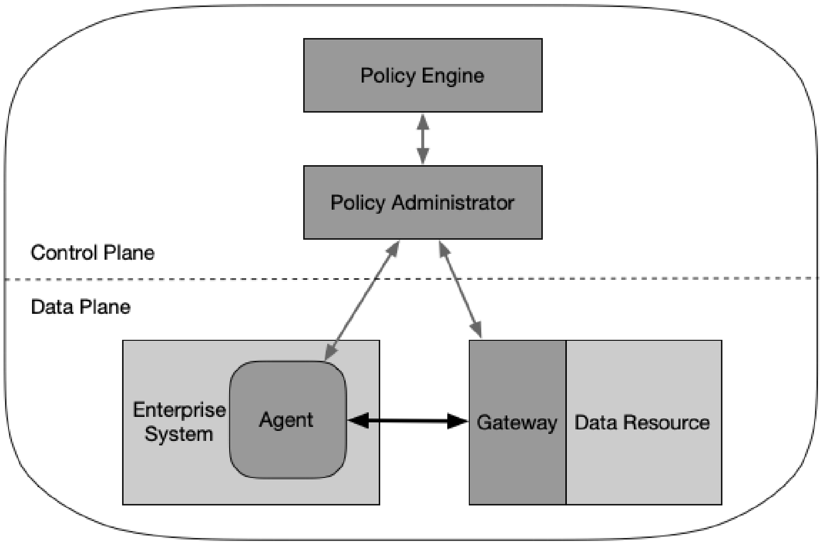

**Figure 3: Device Agent/Gateway Model**

In a typical connection scenario, a user with an enterprise-issued laptop wishes to connect to an enterprise resource (e.g., HR application/database). The connection request is taken by the local agent, and a connection request is sent to the Policy Administrator. The Policy Administrator (and Policy Engine) could be an enterprise local system or a cloud-hosted service. The Policy

Administrator forwards the request to the Policy Engine for evaluation. If the request is authorized, the Policy Administrator configures a communication channel between the device agent and the relevant resource gateway (via the control plane). This may include IP address/port information, session key, or similar security artifacts. The device agent and gateway then connect, and encrypted application data flows begin. The connection between the device agent and resource gateway is terminated when the workflow is completed or when triggered by the Policy Administrator due to a security event (e.g., session timeout, failure to re-authenticate, etc.).

This model is best utilized for enterprises that have a robust device management program in place and discrete resources that can communicate with the gateway. For enterprises that heavily utilize cloud services, this is a client-server implementation of the Cloud Security Alliance (CSA) Software Defined Perimeter (SDP) [CSA-SDP]. This model is also good for enterprises that do not want to have a bring-your-own-device (BYOD) policy in place. Access is only granted via the device agent, which can be placed on enterprise-owned systems.

### 3.1.2 Microperimeter-Based Deployment

This deployment model is a variation of the device agent/gateway model above. In this model, the gateway components may not reside on systems or in front of individual resources but instead reside at the boundary of a resource enclave (e.g., on-location data center) as shown in Figure 4. Usually, these resources serve a single business function or may not be able to communicate directly to a gateway (e.g., legacy database system that does not have an API that cannot be used to communicate with a gateway). This deployment model may also be useful for enterprises that use cloud-based microservices for business processes (e.g., user notification, database lookup, or salary disbursement). In this model, the entire private cloud is located behind a gateway.

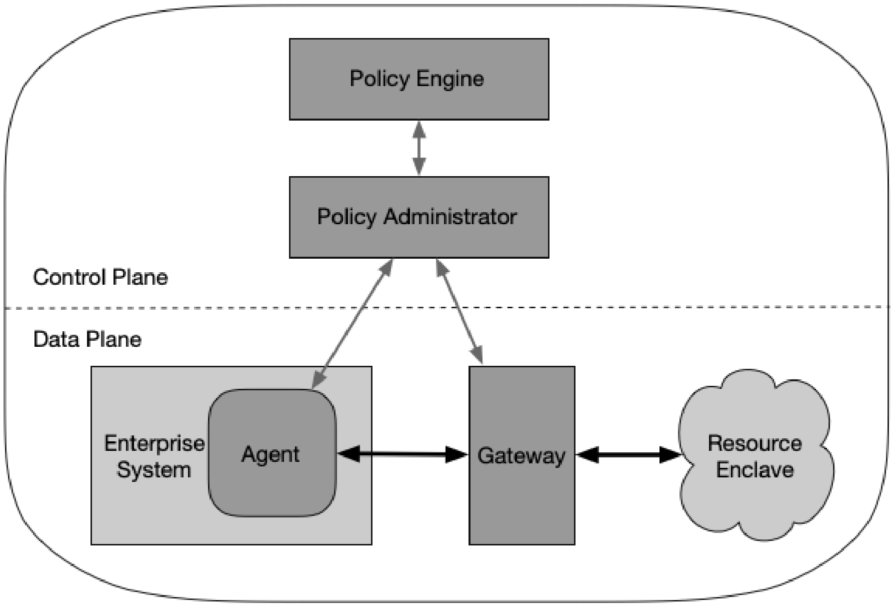

**Figure 4: Enclave Gateway Model**

It is possible for this model to be a hybrid with the device agent/gateway model. In this model, enterprise systems have a device agent that is used to connect to microperimeter gateways, but these connections are created using the same process as the basic device agent/gateway model.

This model is useful for enterprises that have legacy applications or on-premises data centers that cannot have individual gateways in place. The enterprise needs to have a robust device management program in place to install/configure the device agents. The downside is that the gateway protects a collection of resources and not each resource individually. This is a relaxation of the ZTA tenet that each resource should have its own PEP protecting it. This may also allow for clients to see resources for which they do not have privileges to access.

### 3.1.3 Resource Portal-Based Deployment

In this deployment model, the PEP is a single component, which acts as a gateway for user requests. The gateway portal can be for an individual resource or a microperimeter for a collection of resources used for a single business function. One example would be a gateway portal into a private cloud or data center containing legacy applications as shown in Figure 5.

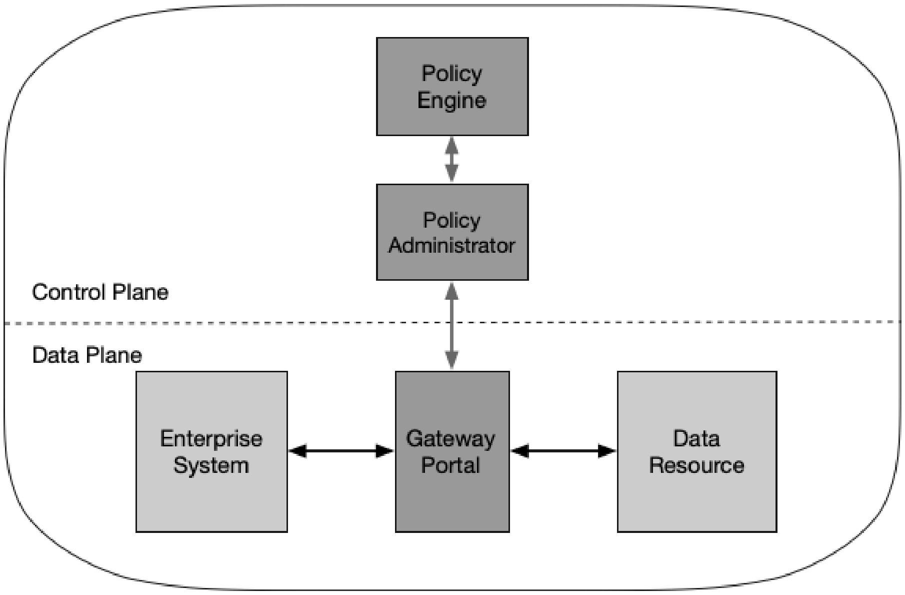

**Figure 5: Resource Portal Model**

The main benefit of this model over the others is that there does not need to be a software component installed on all enterprise systems. This model is also more flexible for BYOD policies and inter-organization collaboration projects. Enterprise administrators do not need to ensure that each device has the appropriate device agent before use. However, limited information can be inferred from devices requesting access. It can only scan and analyze systems and devices once they connect to the PEP portal and may not be able to continuously monitor them for malware and appropriate configuration.

The main difference with this model is that there is no local agent that handles requests. This model allows for more flexibility in client systems and BYOD policies and may make it easier to grant resource access to non-enterprise collaborators. The disadvantage is that the enterprise may not have full visibility or control over enterprise-owned systems as they can only see/scan them when they connect to a portal. These systems may be invisible to the enterprise between these connection sessions. This model also allows for attackers to discover and attempt to access the portal or attempt a denial-of-service (DoS) attack against the portal.

### 3.1.4 System Application Sandboxing

Another variation of the agent/gateway deployment model is having trusted applications run compartmentalized on systems. These compartments could be VMs, containers, or some other implementation, but the goal is the same: to protect the application from the host and other applications running on the system.

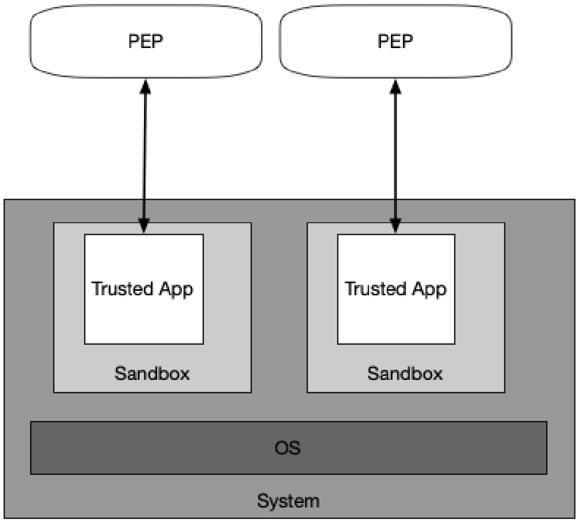

**Figure 6: Application Sandboxes**

In Figure 6 above, the user system runs trusted applications in a sandbox. The trusted application can communicate with the PEP to request access to resources, but the PEP will refuse connections from other (non-trusted) applications on the system. The PEP could be an enterprise local service or a cloud service in this model.

The main advantage of this model variant is that individual applications are segmented away from the rest of the system. If the system cannot be scanned for vulnerabilities, these individual sandboxed applications may be protected from a potential malware infection on the host system. One of the disadvantages to this model is that enterprises must maintain these sandboxed apps for all systems and may not have full visibility into client systems.

## 3.2 Trust Algorithm

For an enterprise with a ZTA deployment, the Policy Engine can be thought of as the brain and the PE's trust algorithm its primary thought process. The trust algorithm is the process used by the Policy Engine to ultimately grant or deny access to a resource. The Policy Engine takes input from multiple sources: the policy database with information about users, user attributes and roles, historic user behavior patterns, threat intelligence sources, and other metadata sources. The process can be visualized in Figure 7.

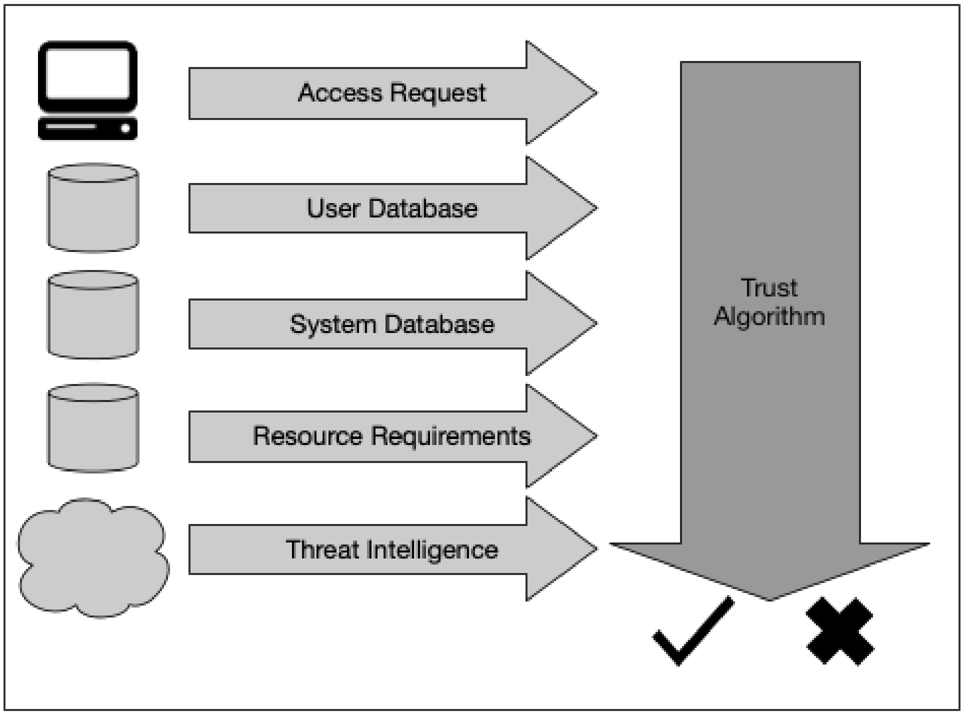

**Figure 7: Trust Algorithm Input**

In the figure, the inputs can be broken down into categories based on what they provide to the trust algorithm.

- **Access request:** The actual request from the application. The resource requested is the primary information used, but information about the requester is also used. This can include OS version, application used, and patch level. Depending on the system state, access to assets might be restricted or denied.
- **User identification, attributes, and privileges:** This is the "who" that is requesting access to a resource. This is the set of users (human and processes) of the enterprise and a collection of user attributes developed by the enterprise. These users and attributes form the basis for policies for resource access [SP800-162][NISTIR 7987]. User identities can include a mix of logical identity (e.g., account ID/password), biometric data (e.g., fingerprints, facial recognition, iris recognition, retina, and odor/scent), and behavior characteristics (e.g., typing rhythm, gait, and voice). Attributes of identity that should be factored into deriving trust scores include time and geolocation. A collection of privileges given to multiple users could be thought of as a role, but privileges should be assigned to a user on an individual basis and not simply because they may fit into a particular role. This should be encoded and stored in an ID management system and policy database.
- **System database and observable status:** This is the database containing the known status of each enterprise-owned system (physical and virtual, to some extent). This is compared to the observable status of the system making the request. This can include OS version, application used, location (network location and geolocation), Trusted Platform

Module (TPM), and patch level. Depending on the system state, access to assets might be restricted or denied.

- **Resource access requirements:** This is the complementary set of policies to the user ID and attributes database. This defines the minimal requirements for access to the resource. Requirements may include authenticator assurance levels, such as multifactor authentication (MFA) and network location (e.g., deny access from overseas IP addresses) or requests for system configuration. These requirements should be developed by both the data custodian (i.e., those responsible for the data) and those responsible for the business processes that utilize the data (i.e., those responsible for the mission).
- **Threat intelligence:** This is an information feed (or feeds) about general threats and active malware operating on the Internet. This can include attack signatures and mitigations. This is the only component that will rarely be under control of the enterprise but most likely a service.

The weight of importance for each data source may be a proprietary algorithm or may be configured by the enterprise. These weight values can be used to reflect the importance of the data source to an enterprise.

The final determination is then passed to the PA for enforcement. The PA's job is to configure the necessary PEPs to enable the connection. Depending on how the ZTA is deployed, this may involve sending authentication results and connection configuration information to gateways and agents or resource portals. The PA is also responsible for terminating the connection based on policy (e.g., after a timeout, when the workflow has been completed, or due to a security alert).

### 3.2.1 Trust Algorithm Variations

There are different ways to implement a ZTA Trust Algorithm (TA). Different implementors may wish to weigh the above factors differently, according to their perceived importance. There are two other major characteristics that can be used to differentiate TAs. The first is how the factors are evaluated, either as binary decisions or weighted parts of a whole "score." The second is how they evaluate requests in relation to other requests by the same user (or application) ID.

- **Criteria vs. Score-based:** A criteria-based TA assumes a set of qualified attributes that must be met before access is granted to a resource. These criteria are configured by the enterprise and should be independently configured for every resource. Access is granted to a resource only if all the criteria are met. A score-based TA computes a "score" based on values for every data source and enterprise-configured weights. If the score is greater than the configured threshold value for the resource, access is granted. Otherwise, the access is denied.

- **Singular vs. Contextual:** A singular TA treats each request individually and does not take the user/application history into consideration when making its evaluation. This can allow for faster evaluations, but there is a risk that an attack can go undetected if it stays within a user's allowed role. A contextual TA takes a user's (or network agent's) recent history into consideration when evaluating access requests. This means the PE must maintain some state information of all users and applications but may be more likely to detect an attacker using subverted credentials to access information in a pattern that is atypical of what the PE sees for the given user/agent.

The two factors are not dependent on each other. It is possible to have a TA that assigns trust scores to every user and/or device and still considers every access request independently (i.e., singular). Likewise, a different TA may be score-based but be contextual in that every successful and failed access request could be used to change the ultimate trust score value.

Ideally, a ZTA Trust Algorithm should be contextual, but this may not always be possible. This can mitigate threats where an attacker stays close to a “normal” set of access requests for a compromised user account (or insider attack). It is important to balance security, usability, and cost effectiveness when defining and implementing Trust Algorithms. Continually prompting a user for re-authentication against behavior that is consistent with historical trends and norms for their mission function and role within the organization can lead to usability issues. For example, if an employee in the HR department of an agency normally accesses 20-30 employee records in a typical workday, a contextual TA may send an alert if the access requests suddenly exceed 100 records in a day as this could be an attacker exfiltrating records using a compromised HR account. This is an example where a contextual TA can detect an attack whereas a singular TA may fail to detect the new behavior. Another example is an accountant who typically accesses the financial system during normal business hours and is now trying to access the system in the middle of the night from an unrecognizable location. A contextual TA may trigger an alert and require the user to satisfy a more stringent score or other criteria as outlined in NIST SP 800-63a [SP800-63A].

Developing a set of criteria or weights/threshold values for each resource requires planning and testing. Enterprise administrators may encounter issues during the initial deployment of ZTA where access requests that should be approved are denied due to misconfiguration. This will result in an initial “tuning” phase of deployment. Criteria or scoring weights may need to be adjusted to ensure that the policies are enforced while still allowing the enterprise’s business processes to function.

## 3.3 Network Components

In a ZTA network, there should be a separation (logical or possibly physical) between communication flows used to control and configure the network and application communication flows used to perform the actual work of the organization. This is often broken down to a *control plane* for network control communication and a *data plane* for application communication flows [Gilman].

The control plane is used by the various infrastructure components for maintaining systems; judging, granting, or denying access to resources; and performing any necessary operations to set up connections between resources. The data plane is used for the actual communication between applications. This communication channel may not be possible prior to the connection being established via the control plane. For example, the control plane could be used by the PA and PEP to set up the connection between the user and the enterprise resource. The application workload would then use the data plane connection that was established.

### 3.3.1 Network Requirements to Support ZTA

1. **Enterprise systems have basic network connectivity.** The local network provides basic routing and infrastructure (e.g., DNS, etc.). The remote enterprise system may not necessarily use all infrastructure services.

2. **The enterprise must be able to determine which systems are owned or managed by the enterprise and which devices are not owned or managed.** This is determined by enterprise-issued credentials and not unauthenticated information (e.g., MAC addresses, etc.).

3. **The enterprise can capture all network traffic.** The enterprise can record packets seen on the data plane but may not be able to perform Deep Packet Inspection (DPI) on all packets. The enterprise can filter out metadata about the connection (e.g., destination, time, device identity, etc.).

4. **Enterprise resources should not be discoverable without accessing a PEP.** Enterprise resources do not accept arbitrary incoming connections from the Internet. Resources only accept custom configured connections after a client has been authenticated. These connections are set up by the PEP. This prevents attackers from scanning the network to identify targets and launching DoS attacks against resources.

5. **The Data Plane and Control Plane are logically separate**. The Policy Engine, Policy Administrator, and PEPs communicate on a network that is logically separate and inaccessible by enterprise systems and resources. Enterprise systems use the data plane when performing network tasks. The Policy Engine, Policy Administrator, and PEPs use the control plane to communicate and manage connections between systems. The PEPs must be able to send and receive messages from both the data and control planes.

6. **Enterprise systems can reach the PEP component.** Enterprise users must be able to access the PEP component on their enterprise ZTA network in order to gain access to resources. This could take the form of a web portal or software agent on the enterprise system that enables the connection.

7. **The PEP is the only component that can access the Policy Administrator and Policy Engine.** Each PEP operating on the enterprise network has a connection to the Policy Administrator in order to establish connections from clients. The PA may be discoverable, but only PEPs are allowed to connect.

8. **Remote enterprise systems should be able to access enterprise resources without needing to traverse through enterprise infrastructure**. For example, a remote user should not be required to use a secure link back to the enterprise network (i.e., VPN) in order to access services utilized by the enterprise and hosted by a public cloud provider (e.g., email).

9. **Enterprise systems may not be able to reach certain PEPs due to observable factors**. For example, mobile systems may not be able to reach certain resources unless they are using enterprise network infrastructure. These factors could be based on location (geolocation or network location), device type, etc.

# 4 Deployment Scenarios/Use Cases

Any enterprise network can be designed with zero trust tenets in mind. Most organizations already have some elements of zero trust in their enterprise infrastructure today or are on their way through implementation of information security and resiliency policies and best practices. There are several deployment scenarios and use cases that lend themselves more readily to a zero trust architecture. For instance, ZTA has its roots in organizations that are geographically distributed and/or have a highly mobile workforce. That said, any organization that has a sizable network with multiple resources can benefit from a zero trust architecture.

In the use cases below, ZTA is not explicitly indicated, as the enterprise likely has both legacy and (possibly) ZTA infrastructures. As discussed in Section 7.2, there will likely be a period of time when ZTA components and legacy network infrastructure are concurrently in operation in an enterprise.

## 4.1 Enterprise with Satellite Facilities

The most common scenario is an enterprise with a single headquarters and one or more geographically dispersed locations that are not joined by an enterprise-owned physical network connection (see Figure 8). Employees at the remote location may not have a full enterprise-owned local network but still need to access enterprise resources in order to perform their tasks. Likewise, employees may be teleworking or in a remote location using enterprise-owned or personally-owned devices. In such cases, an enterprise may wish to grant access to some resources (e.g., employee calendar, email) but deny access to more sensitive resources (e.g., HR database).

In this use case, the PE/PA is best hosted as a cloud service with end systems having a connection agent (see Section 3.1.1) or accessing a resource portal (see Section 3.1.3). It may not be most responsive to have the PE/PA hosted on the enterprise local network as remote offices and workers must send all traffic back to the enterprise network in order to reach cloud services.

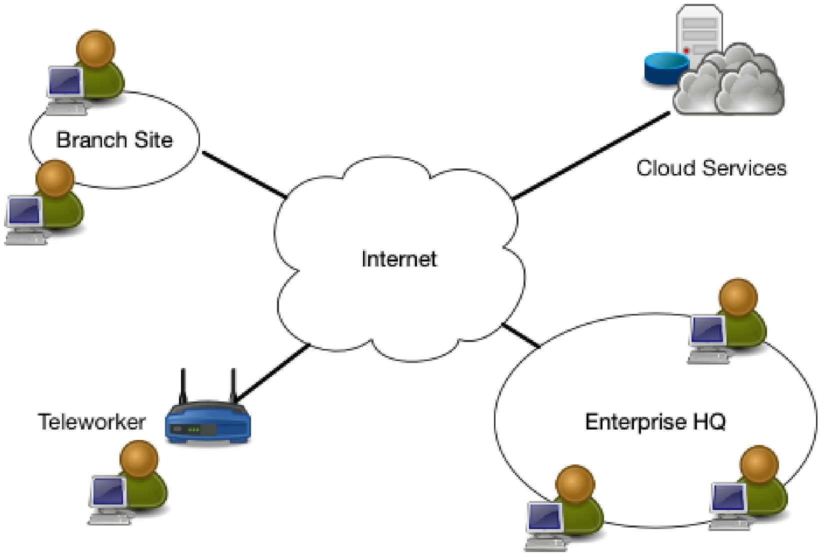

**Figure 8: Enterprise with Remote Employees**

### 4.2 Multi-Cloud Enterprise

One increasingly common use case for deploying a ZTA strategy is an enterprise utilizing multiple cloud providers (see Figure 9). In this use case, the enterprise has a local network but uses two (or more) cloud service providers to host applications and data. Sometimes, the application is hosted on a separate cloud service than the data source. For performance and ease of management, the application hosted in Cloud Provider A should be able to connect directly to the data source hosted in Cloud Provider B rather than force the application to tunnel back through the enterprise network.

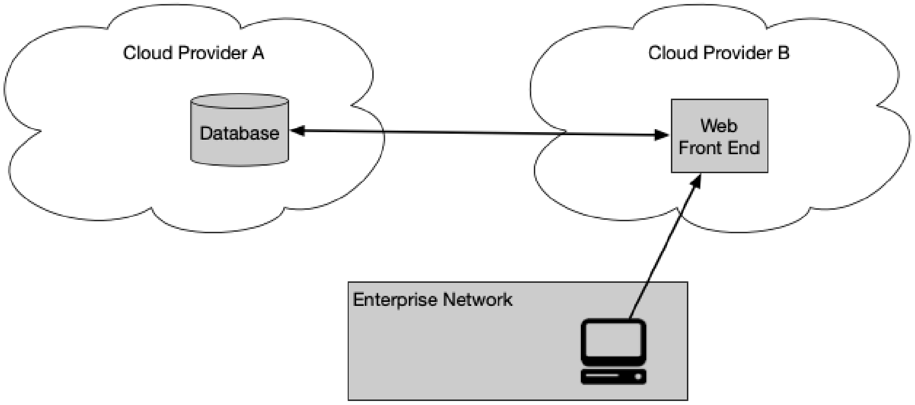

**Figure 9: Multi-Cloud Use Case**

This multi-cloud use case is one of the main drivers of ZTA adoption. It is the server-server implementation of the CSA's SDP specification [CSA-SDP]. As enterprises move to more cloud-hosted applications and services, it becomes apparent that relying on the enterprise perimeter for security becomes a liability. As discussed in Section 2.2, ZTA takes the view that there should be no difference between enterprise-owned and operated network infrastructure and infrastructure owned by any other service provider. The zero trust approach to multi-cloud use is to place PEPs at the access points of each application and data sources. The PE and PA may be a service located in either cloud or even on a third cloud provider. The client (via a portal or local installed agent) then accesses the PEPs directly. That way, the enterprise can still manage access to resources even when hosted outside of the enterprise.

### 4.3 Enterprise with Contracted Services and/or Non-Employee Access

Another common scenario is an enterprise that includes on-site visitors and/or contracted service providers that require limited access to enterprise resources to do their work (see Figure 10). For example, an enterprise has its own internal applications, databases, and employee work systems. These include services contracted out to providers who may occasionally be on-site to provide maintenance tasks (e.g., a smart HVAC system and lighting system that is owned and managed by external providers). These visitors and service providers will need network connectivity to perform their tasks. A ZTA network could facilitate this by allowing these devices (and any visiting service technician) access to the Internet while obscuring enterprise resources.

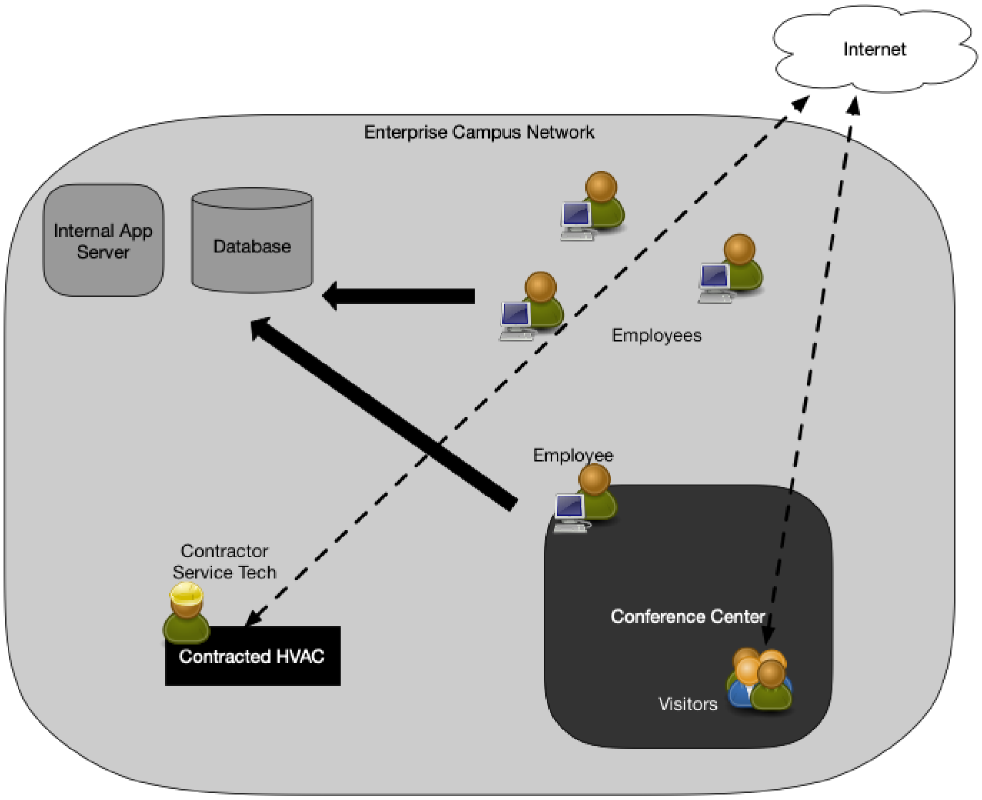

**Figure 10: Enterprise with Non-Employee Access**

In this example the organization also has a conference center where visitors interact with employees. Again, with a ZTA strategy of SDPs, employee devices and users are differentiated and may be able to access appropriate enterprise resources. Visitors to the campus can have Internet access but cannot access enterprise resources. They cannot even conduct network scans to look for enterprise services that may be visible (i.e., prevent active network reconnaissance).

In this use case, the PE and PA could be hosted as a cloud service or on the LAN (assuming little or no use of cloud-hosted services). The enterprise systems could have an installed agent or access resources via a portal. The PA ensures all non-enterprise systems (those that do not have installed agents or cannot connect to a portal) cannot access local resources but may access the Internet.

## 4.4 Collaboration Across Enterprise Boundaries

A fourth use case is cross-enterprise collaboration. For example, there is a project involving employees from Enterprise A and Enterprise B (see Figure 11). The two enterprises may be separate federal agencies (G2G) or even a federal agency and a private enterprise (G2B). Enterprise A operates the database used for the project but must allow access to the data for certain members of Enterprise B. Enterprise A can set up specialized accounts for the employees of Enterprise B to access the required data and deny access to all other resources.

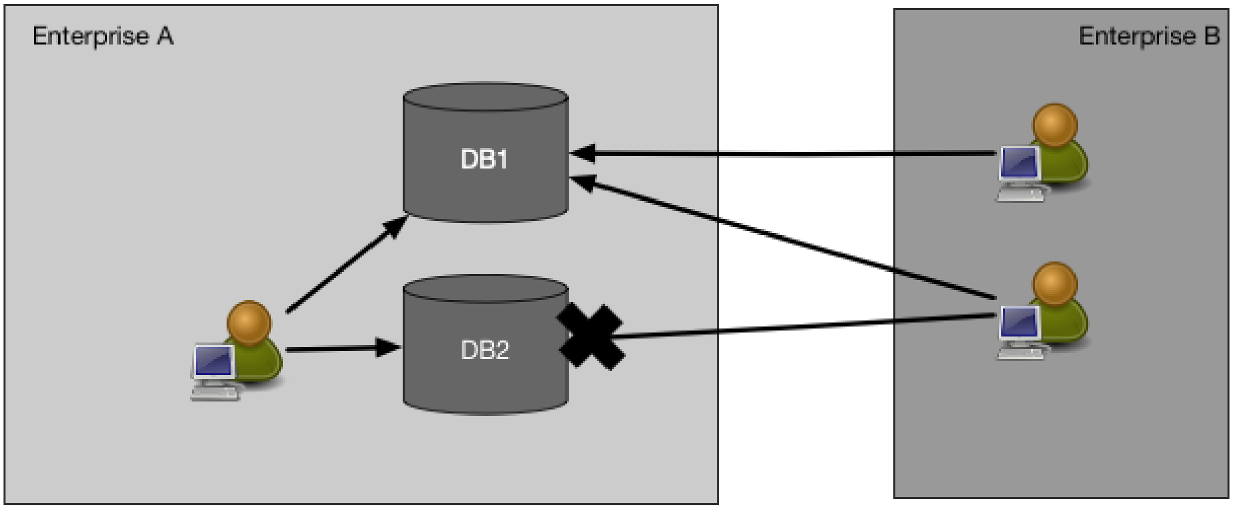

**Figure 11: Cross-Enterprise Collaboration**

This scenario can be similar to use case 1 above as employees of both enterprises may not be located on their organization's network infrastructure, and the resource they need to access may be within one enterprise network or hosted in the cloud. This means that there do not need to be complex firewall rules or enterprise-wide ACLs allowing certain IP addresses belonging to Enterprise B to access resources in Enterprise A. How this access is accomplished depends on the technology in use. Similar to use case 1, the PE and PA would ideally be hosted as a cloud service. The employees of Enterprise B may be asked to install a software agent on their system or access the necessary data resources through a web proxy gateway (see Section 3.1.3).

# 5 Threats Associated with Zero Trust Architecture

No enterprise can completely eliminate cybersecurity risk. When complemented with existing cybersecurity policies and guidance, identity and access management, continuous monitoring, and general cyber hygiene, ZTA can reduce overall risk exposure and protect against common threats. However, there are some threat risks unique to ZTA.

## 5.1 Subversion of ZTA Decision Process

In ZTA, the Policy Engine and Policy Administrator components are the key components of the entire enterprise. No connection between enterprise resources occurs unless it is approved, and possibly configured, by the PE and PA. This means that these components must be properly configured and maintained. Any enterprise administrator with configuration access to the PE's rules may be able to perform unapproved changes (or make mistakes) that can disrupt enterprise operations. Likewise, a compromised PA could allow access to resources that would otherwise not be approved (e.g., to a subverted personally-owned device). Mitigating associated risks means that the PE and PA components must be properly configured and monitored, and any configuration changes must be logged and subject to audit.

## 5.2 Denial-of-Service or Network Disruption

In ZTA, the Policy Administrator (PA) is the key component for resource access. Enterprise resources cannot connect to each other without the PA's permission and, possibly, configuration action. If an attacker disrupts or denies access to the PEP or PA (i.e., denial-of-service (DoS) attack), it can adversely impact enterprise operations. It is assumed that most enterprises will mitigate this threat by having the policy enforcement reside in a cloud or be replicated in several locations following guidance on cyber resiliency [SP800-160].

This mitigates the risk but does not completely eliminate it. Botnets such as Mirai produce massive DoS attacks against key Internet service providers and disrupt service to millions of Internet users. It is also possible that an attacker could intercept and block traffic to a PEP or PA from a portion (or all) of the user accounts within an enterprise (e.g., a branch office or even a single remote employee). In such cases, only a portion of enterprise users are affected. This is also possible in traditional VPN-based access, as well, and is not unique to ZTA.

There is also the possibility of the hosting provider accidently taking the Policy Administrator offline. Similar to the Amazon S3 outage in February 2017[4] that prevented access to customers, an operational error could prevent an entire enterprise from functioning if the policy enforcement component becomes inaccessible from the network.

There is also the risk that enterprise resources may not be reachable from the PA, so even if access is granted to a user, the PA cannot configure the access connection from the network. This is similar to any other network disruption in that some or all enterprise users cannot access a particular resource due to that resource not being available for some reason.

[4] https://aws.amazon.com/message/41926/

### 5.3 Insider Threat

Properly implemented ZTA strategies, information security and resiliency policies, and best practices reduce the risk of an insider attack. ZTA does prevent a compromised account or system from accessing resources outside of its normal purview or normal access patterns. Implementation of MFA for network access may also reduce the risk of access from a compromised account. However, just like traditional enterprises, an attacker with valid credentials (or a malicious insider) may still be able to access resources for which the account has been granted access. For example, an attacker (or compromised employee) who has the credentials and enterprise-owned system of a valid human resources employee may still be able to access an employee database.

ZTA increases resistance to this attack and prevents any compromised accounts or systems from moving laterally throughout the network. In addition, a contextual Trust Algorithm (see Section 3.2.1) is more likely to detect and respond quickly to this attack than in a legacy network. The contextual TA can detect access patterns that are out of normal behavior and deny the compromised account (or insider threat) access to sensitive resources.

### 5.4 Visibility on the Network

As mentioned in Section 3.3.1, all traffic is inspected and logged on the network and analyzed to identify and react to potential attacks against the enterprise. However, as also mentioned, some (likely the majority) of the traffic on the enterprise network may be opaque to network analysis tools. This traffic may be from non-enterprise-owned systems (e.g., contracted services that use the enterprise infrastructure to access the Internet) or applications that are resistant to passive monitoring. The enterprise cannot perform DPI or examine the encrypted traffic and must use other methods to assess for a possible attacker on the network.

That does not mean that the enterprise is unable to analyze encrypted traffic that it sees on the network. The enterprise can collect metadata about the encrypted traffic and use that to detect possible malware communicating on the network or an active attacker. Machine learning techniques [Anderson] can be used to analyze traffic that cannot be decrypted and examined. Employing this type of machine learning would allow the enterprise to categorize traffic as valid or possibly malicious and subject to remediation. In a ZTA deployment, only the traffic from non-enterprise-owned systems would need to be examined as all enterprise traffic is subject to analysis by the Policy Administrator (via the PEPs).

### 5.5 Storage of Network Information

A related threat to enterprise analysis of network traffic is the analysis component itself. If network traffic and metadata are being stored for further analysis, that data becomes a target for attackers. Just like network diagrams, configuration files, and other assorted network architecture documents, these resources should be protected. If an attacker can successfully gain access to stored traffic information, they may be able to gain insight into the network architecture and identify assets for further reconnaissance and attack.

Another source of reconnaissance information for an attacker on a ZT network is the management tool used to encode access policies. Like stored traffic, this component contains access policies to resources and can give an attacker information on which accounts are most valuable to compromise (e.g., the ones that have access to the desired data resources).

Like all valuable enterprise data, adequate protections should be in place to prevent unauthorized access and access attempts. As these resources are vital to security, they should have the most restrictive access policies and only be accessible from designated (or dedicated) administrator accounts.

### 5.6 Reliance on Proprietary Data Formats

ZTA relies on several different data sources in order to make access decisions, including information about the requesting user, system used, enterprise and external intelligence, threat analysis, etc. Often, the systems used to store and process this information do not have a common, open standard on how to interact and exchange information. This can lead to instances where an enterprise is locked into a subset of providers due to interoperability issues. If one provider has a security issue or disruption, an enterprise may not be able to migrate to a new provider without extreme cost (e.g., replacing several systems) or going through a long transition program (e.g., translating policy rules from one proprietary format to another). Like DoS attacks, this risk is not unique to ZTA, but since ZTA is heavily dependent on the dynamic access of information (both enterprise and service providers), disruption can affect the core business functions of an enterprise. To mitigate associated risks, enterprises should evaluate service providers on a holistic basis by taking into consideration factors such as vendor security controls, enterprise switching costs, and supply chain risk management.

### 5.7 Use of Non-Person Entities (NPE) in ZTA administration.

Artificial Intelligence (AI) and other software-based agents are being deployed to manage security issues on enterprise networks. These components need to interact with the management components of ZTA (e.g., Policy Engine, Policy Administrator, etc.), sometimes in lieu of a human administrator. How these components authenticate themselves in an enterprise implementing a ZTA strategy is an open issue. It is assumed most automated technology systems will use some means to authenticate when using an API to resource components.

The associated risk is that an attacker will be able to induce or coerce an NPE agent to perform some task that the attacker does not have privilege to perform. The software agent may have a lower bar for authentication (e.g., API key vs. MFA) to perform administration or security-related tasks compared to a human user. If an attacker can interact with the agent, they could theoretically trick the agent into allowing the attacker greater access or to perform some task on behalf of the attacker. There is also a potential risk that an attacker can gain access to a software agent's credentials and impersonate the agent when performing tasks.

# 6 Zero Trust Architecture and Existing Federal Guidance

There are several existing federal policies and guidance that intersect with the planning, deployment, and operation of a ZTA strategy. These policies do not prohibit an enterprise from moving to a more zero trust-oriented network strategy but can influence the development of a zero trust architecture for an agency. When complemented with existing cybersecurity policies and guidance; identity, credential, and access management (ICAM); continuous monitoring; and general cyber hygiene, ZTA may reinforce an organization's security posture and protect against common threats.

## 6.1 ZTA and NIST Risk Management Framework

A ZTA deployment involves developing access polices around acceptable risk to the designated mission or business process (see Section 7.3.3). It is possible to deny all network access to a resource and only allow access via a connected terminal, but this is disproportionately restrictive in most cases. In order for a federal agency to perform its mission, there is an acceptable level of risk. The risks associated with performing the given mission must be identified, evaluated, and mitigated. To assist in this, the NIST Risk Management Framework (RMF) was developed.

ZTA planning and implementation may change the authorization boundaries defined by the enterprise. This is due to the addition of new components (e.g., Policy Engine, Policy Administrator, and PEPs) and a reduction of the reliance on network perimeter defenses. The process described in the RMF will not change in a ZTA cybersecurity strategy.

## 6.2 ZTA and NIST Privacy Framework

Protecting the privacy of users and private information (e.g., personally identifiable information (PII)) is often a prime concern for organizations. Privacy and data protections are included in compliance programs such as FISMA and HIPAA. Recently, NIST issued a draft Privacy Framework[5] for public comment. This document provides a framework to describe privacy risks and mitigation strategies, as well as a process for an enterprise to identify, measure, and mitigate risks to user privacy and private information stored and processed by an organization.

Part of the core requirements for ZTA is that an enterprise should inspect and log all traffic on its network. This includes decrypting traffic as much as possible to enable inspection. Some of this traffic may contain private information or have associated privacy risks. Organizations will need to identify any possible risks associated with the interception, scanning, and logging of network traffic. This may include such things as informing users and obtaining consent (via a login page, banner, or similar) and educating enterprise users. The NIST Privacy Framework could help in developing a formal process to identify and mitigate any privacy-related risks to a ZTA network.

## 6.3 ZTA and Federal Identity, Credential, and Access Management Architecture (FICAM)

User provisioning is a key component of ZTA. The Policy Engine cannot determine if attempted connections are authorized to connect to a resource if the PE has insufficient information to identify associated users and resources. Strong user provision and authentication policies need to

[5] NIST Privacy Framework (DRAFT) https://www.nist.gov/privacy-framework/working-drafts

be in place before moving to a more zero trust-aligned deployment. Enterprises need to have a clear set of user attributes and policies that can be used by a PE to evaluate access requests.

Recently, the Office of Management and Budget (OMB) issued M-19-17 on improving identity management for the Federal Government. The goal of the policy is to develop "…a common vision for identity as an enabler of mission delivery, trust, and safety of the Nation" [M-19-17]. The memo calls on all federal agencies to form an ICAM office to govern efforts related to identity issuance and management. Many of these management policies should use the recommendations in NIST SP 800-63-3, *Digital Identity Guidelines* [SP800-63]. As ZTA is heavily dependent on precise identity management, any ZTA effort will need to integrate with an agency's ICAM policy.

## 6.4 ZTA and Trusted Internet Connection (TIC)

TIC is a federal cybersecurity initiative jointly managed by OMB, DHS, and the General Services Administration (GSA), and is intended to establish a network security baseline across the Federal Government. Historically, TIC was a perimeter-based cybersecurity strategy which required agencies to consolidate and monitor their external network connections. Inherent in TIC 1.0 and TIC 2.0 is the assumption that the inside of the perimeter is "trusted," whereas ZTA assumes that network location does not infer "trust" (i.e., there is no "trust" on an agency's internal network). TIC 2.0 provides a list of network-based security capabilities (e.g. content filtering, monitoring, authentication, and others) to be deployed at the TIC Access Point at the agency's perimeter; many of these capabilities are aligned with ZTA.

TIC 3.0 will be updated to accommodate cloud services and mobile devices [M-19-26]. In TIC 3.0, agencies can define trust zones as low-trust, moderate-trust, and high-trust based on the level of control, transparency, and verification an agency has over a particular computing environment, as well as the sensitivity of data associated with that environment. In addition, TIC 3.0 has updated the network-based security capabilities to be applied to multiple policy enforcement points (PEPs), which are located at the boundary of a trust zone and not at a single PEP at the agency perimeter. Many of these TIC 3.0 security capabilities directly support ZTA (e.g., encrypted traffic, default/deny, virtualization security, network and system inventory, and others). TIC 3.0 defines specific use cases that describe the implementation of trust zones and security capabilities across specific applications, services, and environments.

TIC 3.0 is focused on network-based security protections, whereas ZTA is a more inclusive architecture addressing application, user, and data protections. As TIC 3.0 evolves its use cases, it is likely that a ZTA TIC use case will be developed to define the network protections to be deployed at ZTA enforcement points.

## 6.5 ZTA and EINSTEIN (NCPS – National Cybersecurity Protection System)

NCPS (aka EINSTEIN) is an integrated system-of-systems that delivers intrusion detection, advanced analytics, information sharing, and intrusion prevention capabilities to defend the Federal Government from cyber threats. The goals of NCPS, which align with the overarching goals of Zero Trust, are to manage cyber risk, improve cyber protection, and empower partners to secure cyber space. EINSTEN sensors enable CISA's National Cybersecurity and Communications Integration Center (NCCIC) to defend federal networks and respond to significant incidents at federal agencies.

The placement of NCPS sensors is based on a perimeter network defense in the Federal Government, while Zero Trust Architectures move protections closer to the data and resources. If ZTA is adopted across the Federal Government, the NCPS implementation would need to evolve, or new capabilities would need to be deployed to fulfill NCPS objectives. Incident responders could potentially leverage information from the authentication, traffic inspection, and logging of agency traffic available to federal agencies that have implemented a Zero Trust Architecture. Information generated in a ZTA may better inform event impact quantification; machine learning tools could use ZTA data to improve detection; and additional logs from ZTA may be saved for after-the-fact analyses by incident responders.

## 6.6 ZTA and Continuing Diagnostics and Mitigations (CDM)

The DHS CDM program is an effort to improve federal agency IT security posture. Key to that posture is for agencies to have insight into the systems, configuration, and users within an agency. In order to protect a system, agencies need to set up processes to discover and understand the basic components and actors on their infrastructure:

- **What is connected?** What devices, applications, and services are used by the organization? This includes observing and improving the security posture of these artifacts as vulnerabilities and threats are discovered.
- **Who is using the network?** Which users are part of the organization or are external and allowed to access enterprise resources? This includes non-person entities (NPEs) that may be performing autonomous actions.
- **What is happening on the network?** Enterprises need insight into traffic patterns and messages between systems.
- **How is data protected?** The enterprise needs a set policy on how information is protected at rest and in transit.

Having a strong CDM program is key to the success of ZTA. For example, in order to move to ZTA, an enterprise must have a complete inventory of both physical and virtual assets. The DHS CDM program has initiated several efforts to build up the capabilities needed within federal agencies to move to a ZTA strategy. For example, the DHS Hardware Asset Management (HWAW) [HWAW] program is an effort to help agencies identify devices on their network infrastructure in order to deploy a secure configuration. This is similar to the first steps in developing a roadmap to ZTA. Agencies must have visibility into the systems active on the network in order to categorize, configure, and monitor its activity.

## 6.7 ZTA, Cloud Smart, and the Federal Data Strategy

The Cloud Smart[6] strategy, updated Data Center Optimization Initiative [M-19-19] policy, and the Federal Data Strategy[7] influence some requirements for agencies when planning a ZTA strategy for its enterprise. These policies require agencies to inventory and assess how they collect, store, and access data, both on-premises and in the cloud.

This inventory is critical in determining which business processes and resources would benefit

[6] Federal Cloud Computing Strategy https://cloud.cio.gov/strategy/
[7] Federal Data Strategy https://strategy.data.gov/

from implementing ZTA. Data resources and applications that are primarily cloud-based or primarily used by remote workers are good candidates for a ZTA approach (see Section 7.3.3) since the users and resources are located outside of the enterprise network perimeter.

One additional consideration with the Federal Data Strategy is how to make agency data assets accessible to other agencies or the public. This corresponds with the cross-enterprise collaboration ZTA use case (see Section 4.4). Agencies using a ZTA for these assets may need to take collaboration (or publication) requirements into account when developing the strategy.

# 7 Migrating to a Zero Trust Architecture

Implementing a ZTA strategy is a journey rather than a wholesale replacement of infrastructure or processes. Organizations should seek to incrementally implement zero trust principles, process changes, and technology solutions that protect its highest value data assets. Most enterprises will continue to operate in a hybrid zero trust-legacy mode for an indefinite period of time while continuing to invest in ongoing IT modernization initiatives.

How an enterprise migrates to a ZTA strategy depends on their current cybersecurity posture and operations. There is a baseline of competence that an enterprise should reach before it becomes possible to deploy a significant ZTA-focused network [7]. This baseline includes having the assets, users, and business processes identified and cataloged for the enterprise. The enterprise needs this information before it can develop a list of ZTA candidate business processes and the users/systems that are involved for this process.

## 7.1 Pure Zero Trust Architecture

In a green field approach, it would be possible to build a zero trust architecture network from the ground up. Assuming the enterprise knows the applications and work flows it wants to use for its operations, it can produce an architecture based on zero trust strategy tenets for those workflows. Once the workflows are identified, the enterprise can narrow down the components needed and begin to map how the individual components interact. From that point, it is an engineering exercise in building the network infrastructure and configuring the components.

In practice, this is rarely a viable option for federal agencies or any organization with an existing network. However, there may be times when an organization is asked to fulfill a new responsibility that would require building its own infrastructure. In these cases, it might be possible to introduce ZT concepts to some degree. For example, an agency may be given a new responsibility that entails building a new application and database. The agency could design the newly needed infrastructure around ZT principles, such as having users' trust evaluated before access is granted, having micro perimeters around new resources, etc. The degree of success depends on how dependent this new infrastructure is on existing resources (e.g., ID management systems).

## 7.2 Hybrid ZTA and Legacy Architecture

It is unlikely that any significant enterprise can migrate to a ZTA network in a single technology refresh cycle. There will be a (perhaps indefinite) period of time when ZTA workflows coexist in a traditional enterprise. The migration to a ZTA approach to the enterprise may take place one business process at a time. The enterprise needs to make sure that the common elements (e.g., ID management, device management, event logging, etc.) are flexible enough to operate in a ZTA and legacy hybrid security architecture. Enterprise architects may also want to restrict ZTA candidate solutions to those that can interface with existing components.

## 7.3 Steps to Introducing ZTA to a Legacy Architected Network

Migrating to ZTA requires an organization to have detailed knowledge of its assets (physical and virtual), users, and business processes. This knowledge is accessed by the PE when evaluating resource requests. Incomplete knowledge will most often lead to a business process failure where the PE denies requests due to insufficient information.

Before undertaking an effort to bring ZTA to an enterprise, there should be a survey of assets and users. This is the foundation state that must be reached before a ZTA deployment is possible. These surveys can be conducted in parallel, but both are tied to an examination of the business processes of the organization. These steps can be mapped to the steps in the Risk Management Framework (RMF) [SP800-37] as any move to ZTA can be seen as a process to reduce risk to an agency's business functions. The pathway to ZTA can be visualized in Figure 12.

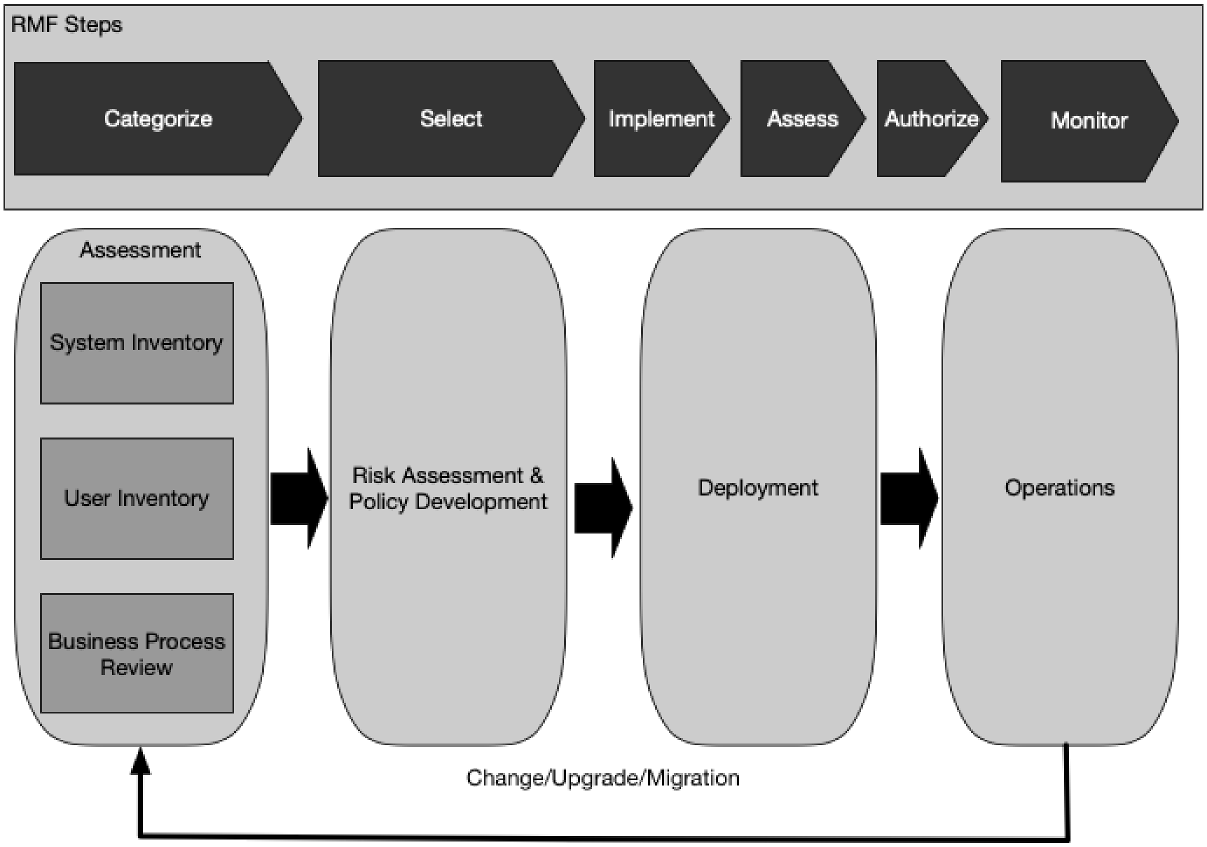

**Figure 12: ZTA Deployment Cycle**

After the initial inventory is created, there is a regular cycle of maintenance and updating. This updating may change business processes or not have any impact, but the evaluation of business processes should be conducted. For example, a change in digital certificate providers may not appear to have a significant impact but may involve certificate root store management, Certificate Transparency log monitoring, and other factors that are not apparent at first.

### 7.3.1 Identify Actors on the Enterprise

In order for a ZTA network to operate, the PE must have knowledge of enterprise subjects. "Subjects" encompasses both human and possible non-person entities (NPEs), such as service accounts that interact with resources.

Users with special privileges, such as developers or system administrators, need additional consideration when being assigned attributes or roles. In a traditional security architecture, these accounts may have blanket permission to access all enterprise resources. ZTA should allow for developers and administrators to have sufficient flexibility to satisfy their business requirements while logging and auditing behavior. ZTA deployments may require administrators to satisfy a more stringent score or criteria as outlined in NIST SP 800-63A, Section 5 [SP800-63A].

### 7.3.2 Identify Assets Owned by the Enterprise

As mentioned in Section 2.1, one of the key requirements of ZTA is the ability to identify and manage enterprise-owned devices. ZTA also requires the ability to identify and monitor non-enterprise owned devices that may be on enterprise-owned network infrastructure or that access enterprise resources. The ability to manage enterprise assets is key to the successful deployment of ZTA. This includes hardware components (e.g. laptops, phones, IoT devices, etc.) and digital artifacts (e.g., user accounts, applications, digital certificates, etc.).

This goes beyond simply cataloging and maintaining a database of enterprise assets. This also includes configuration management and monitoring. The ability to observe the current state of a system is part of the process of evaluating access requests (see Section 2.1). This means that the enterprise must have the ability to configure, survey, and update enterprise systems, including virtual systems, containers, etc. This also includes both its physical (as best estimated) and network location. This information should inform the PE when making resource access decisions.

Non-enterprise-owned assets should also be cataloged to the best extent possible. This may include whatever is visible by the enterprise (e.g., MAC address, network location) and augmented by administrator data entry. This information is not just used for access decisions (as collaborator and BOYD systems may need to contact PEPs) but for monitoring by the enterprise.

Many federal agencies have already begun the task of identifying enterprise assets. Agencies that have established CDM capabilities such as Hardware Asset Management (HWAM) [HWAM] and Software Asset Management (SWAM) [SWAM] have a rich set of data to draw from when enacting a ZTA strategy. Agencies may also have a list of ZTA candidate processes that involve High Value Assets (HVA) [M-19-03] that have been identified as key to the agency mission. This work would need to exist enterprise or agency-wide before any business process could be (re-)designed with a ZTA strategy. These programs must be designed to be expandable and adaptable to changes in the enterprise, not just when migrating to ZTA but to account for new systems, services, and business processes that become part of the enterprise.

### 7.3.3 Identify Key Processes and Evaluate Risks Associated with Executing Process

The third inventory that an agency should undertake is to identify and rank the business processes (i.e., missions) of the agency. Business processes should inform the circumstances under which resource access requests are granted and denied. An enterprise may wish to start with a low risk business process for the first transition to ZTA as disruptions will likely not negatively impact the entire organization. Once enough experience is gained, more critical business processes can become candidates.

Business processes that utilize cloud-based resources or are used by remote workers are often good candidates for ZTA. This is because the clients and resources are outside of the enterprise perimeter, one of the main advantages of ZTA over legacy enterprise network architecture. Rather than project the enterprise perimeter into the cloud or bring clients into the enterprise network via a Virtual Private Network (VPN), enterprise clients can request cloud services directly. The enterprise's PEPs ensure that enterprise policies are followed before resource access is granted to a client.

### 7.3.4 Formulating Policies for the ZTA Candidate

The process of identifying a candidate application or business workflow depends on several factors: the importance of the process to the organization, the group of users affected, and the current state of resources used for the workflow. The value of the asset or workflow based on risk to the asset or workflow can be evaluated using the NIST Risk Management Framework [SP800-37].

After the asset or workflow is identified, the next step is to identify the user set that would be affected. This may influence the candidate choice as a first migration to ZTA. An application used by an identified subset of enterprise users (e.g., a purchasing system) may be preferred over one vital to the entire user base of the enterprise (e.g., email).

The enterprise administrators then need to determine the set of criteria (if using a criteria-based TA) or trusted score weights (if using a score-based TA) for the resources used in the candidate business process (see Section 3.2.1). Administrators may need to make adjustments to these criteria or values during the tuning phase. These adjustments are necessary to make sure policies are effective but do not hinder necessary access to resources.

### 7.3.5 Identifying Candidate Solutions

Once a list of candidate business processes has been developed, enterprise architects can compose a list of candidate solutions. Some deployment models (see Section 3.1) are better suited to particular workflows and current enterprise ecosystems. Likewise, some vendor solutions are better suited to particular use cases than others. Some factors to take into consideration are:

- **Does the solution require that components be installed on the client system?** This may limit business processes where non-enterprise-owned systems are used or desired, such as BYOD or cross-agency collaborations.
- **Does the solution work where the business process resources exist entirely on enterprise premises?** Some solutions assume requested resources will reside in the cloud (so-called "north-south" traffic) and not within an enterprise perimeter ("east-west" traffic). The location of candidate business process resources will influence candidate solutions as well as the ZTA for the process.

One solution is to model an existing business process as a pilot program and not just as a replacement. This pilot program could be made general to apply to several business processes or specific to one use case. The pilot can be used as a "proving ground" for ZTA before transitioning users to the ZTA deployment and away from the traditional process infrastructure.

### 7.3.6 Initial Deployment and Monitoring

Once the candidate workflow and ZTA components are chosen, the initial deployment can start. Enterprise administrators must implement the developed policies using the selected components but may wish to make them more lenient at first. Few enterprise policy sets are complete in their first iterations: important user accounts (e.g., administrator accounts) may be denied access to resources they need or may not need all the access privileges they have assigned.

The new ZT business workflow could be operated in "reporting only mode" for some time to make sure the policies are effective and workable. "Reporting only" means that access should be granted for most requests, and logs and traces of connections should be compared to the initial developed policy. Basic policies such as denying requests that fail MFA or appear from known, blacklisted IP addresses should be enforced and logged, but after initial deployment, access polices should be more lenient to collect data on actual interactions of the ZT workflow. If it is not possible to operate in a more lenient nature, enterprise network operators should monitor logs closely and be prepared to modify access policies based on operational experience.

### 7.3.7 Expanding the ZTA

After enough confidence is gained in the workflow policy set, the enterprise enters the steady operational phase. The network and systems are still monitored, and traffic is logged (see Section 2.2.1), but responses and policy modifications are done at a lower tempo as they should not be severe. At this stage, the enterprise administrators can begin planning the next phase of ZT deployment. Like the previous rollout, a candidate workflow and solution set need to be identified and initial policies developed.

However, if a change occurs to the workflow, the operating ZT architecture needs to be reevaluated. Significate changes to the system—such as new devices, major updates to software (especially ZT logical components), or shifts in organizational structure—may result in changes to the workflow or policies. In effect, the entire process should be reconsidered with the assumption that some of the work has already been done. For example, new devices have been purchased, but no new user accounts have been created, so only the device inventory needs to be updated.

## References

[ACT-IAC] American Council for Technology and Industry Advisory Council (2019) *Zero Trust Cybersecurity Current Trends*. Available at https://www.actiac.org/zero-trust-cybersecurity-current-trends

[Anderson] Anderson B, McGrew D (2017) Machine Learning for Encrypted Malware Traffic Classification: Accounting for Noisy Labels and Non-Stationarity. *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM, Halifax, Nova Scotia, Canada), pp 1723-1732. https://doi.org/10.1145/3097983.3098163

[CSA-SDP] Cloud Security Alliance (2015) SDP Specification 1.0. April 2015. https://cloudsecurityalliance.org/artifacts/sdp-specification-v1-0/

[Gilman] Gilman E, Barth D (2017) *Zero Trust Networks: Building Secure Systems in Untrusted Networks* (O'Reilly Media, Inc., Sebastopol, CA), 1st Ed.

[HWAM] Department of Homeland Security (2015) *Hardware Asset Management (HWAM) Capability Description*. Available at https://www.us-cert.gov/sites/default/files/cdm_files/HWAM_CapabilityDescription.pdf

[JERICHO] The Jericho Forum (2007) *Jericho Forum Commandments*, version 1.2. Available at https://collaboration.opengroup.org/jericho/commandments_v1.2.pdf

[M-19-03] Office of Management and Budget (2018) Strengthening the Cybersecurity of Federal Agencies by Enhancing the High Value Asset Program. (The White House, Washington, DC), OMB Memorandum M-19-03, December 10, 2018. Available at https://www.whitehouse.gov/wp-content/uploads/2018/12/M-19-03.pdf

[M-19-17] Office of Management and Budget (2019) Enabling Mission Delivery through Improved Identity, Credential, and Access Management. (The White House, Washington, DC), OMB Memorandum M-19-17, May 21, 2019. Available at https://www.whitehouse.gov/wp-content/uploads/2019/05/M-19-17.pdf

[M-19-19] Office of Management and Budget (2019) Update on Data Center Optimization Initiative (DCOI). (The White House, Washington, DC), OMB Memorandum M-19-19, June 25, 2019. Available at https://datacenters.cio.gov/assets/files/m_19_19.pdf

[M-19-26] Office of Management and Budget (2019) Update to the Trusted Internet Connections (TIC) Initiative. (The White House, Washington, DC), OMB

Memorandum M-19-26, September 12, 2019. Available at https://www.whitehouse.gov/wp-content/uploads/2019/09/M-19-26.pdf

[NISTIR 7987] Ferraiolo DF, Gavrila S, Jansen W (2015) Policy Machine: Features, Architecture, and Specification. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Interagency or Internal Report (IR) 7987, Rev. 1. https://doi.org/10.6028/NIST.IR.7987r1

[SP800-37] Joint Task Force (2018) Risk Management Framework for Information Systems and Organizations: A System Life Cycle Approach for Security and Privacy. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Special Publication (SP) 800-37, Rev. 2. https://doi.org/10.6028/NIST.SP.800-37r2

[SP800-63] Grassi PA, Garcia ME, Fenton JL (2017) Digital Identity Guidelines. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Special Publication (SP) 800-63-3, Includes updates as of December 1, 2017. https://doi.org/10.6028/NIST.SP.800-63-3

[SP800-63A] Grassi PA, Fenton JL, Lefkovitz NB, Danker JM, Choong Y-Y, Greene KK, Theofanos MF (2017) Digital Identity Guidelines: Enrollment and Identity Proofing. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Special Publication (SP) 800-63A, Includes updates as of December 1, 2017. https://doi.org/10.6028/NIST.SP.800-63A

[SP800-160] Ross R, Pillitteri V, Graubart R, Bodeau D, and McQuaid R (2019) Developing Cyber Resilient Systems: A Systems Security Engineering Approach. (National Institute of Standards and Technology, Gaithersburg, MD), Final Public Draft NIST Special Publication (SP) 800-160, Vol. 2. Available at https://csrc.nist.gov/publications/detail/sp/800-160/vol-2/draft

[SP800-162] Hu VC, Ferraiolo DF, Kuhn R, Schnitzer A, Sandlin K, Miller R, Scarfone KA (2014) Guide to Attribute Based Access Control (ABAC) Definition and Considerations. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Special Publication (SP) 800-162, Includes updates as of February 25, 2019. https://doi.org/10.6028/NIST.SP.800-162

[SWAM] Department of Homeland Security (2015) *Software Asset Management (SWAM) Capability Description*. Available at https://www.us-cert.gov/sites/default/files/cdm_files/SWAM_CapabilityDescription.pdf

# Appendix A—Acronyms

| | |
|---|---|
| CDM | Continuous Diagnostics and Mitigation |
| DHS | Department of Homeland Security |
| NIST | National Institute of Standards and Technology |
| PA | Policy Administrator |
| PE | Policy Engine |
| PEP | Policy Enforcement Point |
| RMF | NIST Risk Management Framework |
| SIEM | Security Incident and Event Monitoring |
| ZTA | Zero Trust Architecture |
| ZTE | Zero Trust Ecosystem |

# Appendix B—Identified Gaps in the Current State-of-the-Art in ZTA

The current maturity of zero trust components and solutions was surveyed during the background research in producing this document. This survey concluded that the current state of the ZTA ecosystem is not mature enough for widespread adoption. While it is possible to use ZTA strategies to plan and deploy an enterprise network, there is no single solution that provides all the necessary components. Also, few ZTA components available today can be used for all of the various workflows present in an enterprise.

The following is a summary of identified gaps in the ZTA ecosystem and areas that need further investigation. Some of these areas have some foundation of work, but how ZTA tenets change these areas is not well-known as there is not enough experience with diverse ZTA-focused enterprise networks.

## B.1 Technology Survey

Multiple vendors were invited to present their products and views on zero trust. The goal of this survey was to identify missing pieces that prevent agencies from moving to a ZTA infrastructure now or maintaining an existing ZTA deployment. These gaps can be categorized into immediate deployment (immediate or short term), systemic gaps that affect maintenance or operations (short or mid-term), and missing knowledge (areas for future research). They are summarized in Table B-1:

**Table B-1: Summary of Identified Gaps**

| Category | Example Questions | Identified Gaps |
|---|---|---|
| **Immediate** | • How to write procurement requirements<br>• How a ZTA strategy works with TIC, FISMA, etc. | • Lack of a common framework and vocabulary for ZTA<br>• Perception that ZTA is in conflict with existing policy |
| **Systemic** | • How to prevent vendor lock-in<br>• How different ZTA environments interact | • Too much reliance on vendor APIs |
| **Research Areas** | • How threats will evolve in the face of ZTA<br>• How business processes change in the face of ZTA | • What a successful compromise looks like in an enterprise with a ZTA<br>• End user experience in an enterprise with a ZTA |

## B.2 Gaps that Prevent Immediate Move to ZTA

These are the issues that are slowing down the adoption of a ZTA strategy at present. These were classified as "immediate" issues, and no thought of future maintenance or migration were considered for this category. Forward-thinking enterprise may also consider the maintenance category to be of immediate concern in preventing the initial deployment of ZTA components, but they are considered a separate category for this analysis.

### B.2.1 Lack of Common Terms for ZTA Design, Planning, and Procurement

Zero trust as a strategy for the design and deployment of enterprise infrastructure is still a forming concept. Industry has not yet coalesced around a single set of terms or concepts to describe ZTA components and operations. This makes it very difficult for organizations (e.g., federal agencies) to develop coherent requirements and policies for designing ZTA infrastructure and procuring components.

The driver for Section 2.1 and Section 3.1 is an initial attempt to form a neutral base of terms and concepts to describe ZTA. The abstract ZTA components and deployment models were developed to serve as basic terms and ways to think about ZTA. The goal is to provide a common way to view, model, and discuss ZTA solutions when developing enterprise requirements and performing market surveys. The above sections may prove to be incomplete as more experience is gained with ZTA strategies in federal agencies, but they currently serve as a base for a common conceptual framework.

### B.2.2 Perception that ZTA is in Conflict with Existing Federal Cybersecurity Policies

There is a misconception that ZTA is a single framework with a set of solutions that are incompatible with the existing view of cybersecurity. Zero trust should actually be viewed as an evolution of current cybersecurity strategies as many of the concepts and ideas have been circulating for a long time. Federal agencies have been encouraged to take a more zero trust approach to cybersecurity through existing guidance (see Section 6). If an agency has a mature ID management system and robust CDM capabilities in place, it is on the road to a ZTA strategy (see Section 7.3). This gap is based on a misconception of ZTA and how it has evolved from previous cybersecurity paradigms.

## B.3 Systemic Gaps that Impact ZTA

These are the gaps that affect initial implementation and deployment of ZTA strategies and continued operation/maturity. These gaps could slow the adoption of ZTA strategies in agencies or result in a fragmentation of the ZTA component industry. Systemic gaps are areas where open standards (produced either by a Standards Development Organization (SDO) or industry consortium) can help.

### B.3.3 Standardization of Interfaces Between Components

During the technology survey, it became apparent that no one vendor provides a single solution that will provide zero trust. Furthermore, it might not be desirable to use a single vendor solution to achieve zero trust and risk "vendor lock-in." This leads to the issue of interoperability within components, not only at the time of purchase but over time.

The spectrum of components within the wider Zero Trust Ecosystem (ZTE) is vast, with many products focusing on a single niche within ZTE and relying on other products to provide either data or some service to another component (e.g., integration of multi-factor authentication (MFA) for resource access). Vendors too often rely on proprietary APIs provided by partner companies rather than standardized, vendor-independent APIs to achieve this integration. The problem with this approach is that these APIs are proprietary and single-vendor-controlled. The controlling vendor can change the API behavior, and integrators are required to update their products in response. This requires close partnerships between communities of vendors to assure early notification of modifications within APIs which may affect compatibility between products. This adds further burden on vendors and consumers: vendors need to expend resources to make changes to their products, and consumers need to apply updates to multiple products when one vendor makes a change to their proprietary API. Additionally, vendors are required to implement and maintain wrappers for each partner component to allow for maximum compatibility and interoperability. For example, many MFA product vendors are required to create a different wrapper for each cloud provider or identity management system in order to be usable in different kinds of client combinations.

On the customer side, this generates additional problems when developing requirements for purchasing products. There are no standards that purchasers can rely on to identify compatibility between products. Hence, it is very difficult to create a multi-year roadmap for moving into ZTA since it is impossible to identify a minimum set of compatibility requirements for components.

### B.3.4 Emerging Standards that Address Overreliance on Proprietary APIs

As there is not a single solution to deploying a ZTA strategy, there is no single set of tools or services for a zero trust architecture. Thus, it is impossible to have a single protocol or framework that enables an enterprise to move to a ZTA strategy. Currently, there are a wide variety of models and solutions seeking to become the leading authority of ZTA.

This indicates that there is an opportunity for a set of open, standardized protocols (or frameworks) to be developed to aid organizations in migrating to a ZTA strategy. Standards Development Organizations (SDOs) like the Internet Engineering Task Force (IETF) have specified protocols that may be useful in exchanging threat information (called XMPP-Grid [1]). The Cloud Security Alliance (CSA) has produced a framework for Software Defined Perimeter (SDP) [2] that may also be useful in ZTA. Efforts should be directed toward surveying the current state of ZTA-related frameworks or the protocols necessary for a useful ZTA strategy and toward identifying places where work is needed to produce or improve these specifications.

## B.4 Knowledge Gaps in ZTA and Future Areas of Research

The gaps listed here do not hinder an organization from adopting a ZTA strategy for their enterprise. These are gray areas in knowledge about operational ZTA environments. Most are due to a lack of time and experience with mature, zero trust deployments. These are areas of future work for researchers.

### B.4.5 Attacker Response to ZTA

A properly implemented ZTA strategy for an enterprise will improve its cybersecurity posture over legacy network perimeter-based security. The tenets of ZTA aim to reduce the exposure of resources to attackers and minimize (or prevent) lateral movement within an enterprise should a host system be compromised.

However, determined attackers will not sit idle but will, instead, change behavior in the face of ZTA. The open issue is how the attacks will change. One possibility is that attacks aimed at stealing credentials (e.g., phishing, social engineering) could become more prevalent as one of the main tenets of ZTA is frequent authentication before accessing resources. Another possibility is that in a hybrid ZTA/legacy enterprise, attackers will focus on the business processes that have not had ZTA tenets applied (i.e., follow traditional network perimeter-based security)—in effect, targeting the low-hanging fruit in an attempt to gain some foothold in the ZTA business process.

As ZTA matures, more deployments are seen, and experience gained, the effectiveness of ZTA over older methods of network perimeter-based security may become apparent. The metrics of "success" of ZTA over older cybersecurity strategies will also need to be developed.

### B.4.6 User Experience in a ZTA Environment

There has not been a rigorous examination of how end users act in an enterprise using a ZTA strategy. This is mainly due to the lack of large ZTA use cases available for studies. There have been studies on how users react to MFA and other security operations that are seen as part of a ZTA enterprise strategy. This work could form the basis of predicting end user experience and behavior when using ZTA workflows in an enterprise.

One set of studies that can predict how ZTA affects end user experience is the work done on the use of MFA in enterprises and "security fatigue." Security fatigue [3] is the phenomenon wherein end users are confronted with so many security policies and challenges that it begins to impact their productivity in a negative way. Other studies show that MFA may alter user behavior, but the overall change is mixed [4] [5]. Some users readily accept MFA if the process is streamlined and involves devices they are used to using or having with them (e.g., apps on a smartphone). However, some users resent having to use personally-owned devices for business processes or feel that they are being constantly monitored for possible violations of IT policies.

### B.4.7 Resilience of ZTA to Enterprise and Network Disruption

The survey of the ZTA vendor ecosystem displayed the wide range of infrastructure that an enterprise deploying a ZTA strategy would need to consider. As previously noted, there was no one single provider of a full zero trust solution. As a result, enterprises will purchase several different services and products. This can lead to a web of dependencies for components. If one vital component is disrupted or unreachable, there could be a cascade of failures that impact one or multiple business processes.

Most products and services surveyed relied on a cloud presence to provide robustness, but even cloud services have been known to become unreachable either through an attack or simple error. When this happens, key components used to make access decisions may be unreachable or may not be able to communicate with other components. For example, PE and PA components located in a cloud may be reachable during a distributed denial-of-service (DDoS) attack but may not be able to reach all PEPs located with resources. There will need to be research on discovering possible "choke points" of ZTA deployment models and the impact on network operations when a ZTA component is unreachable or has limited reachability.

The Continuity of Operations (COOP) plans for an enterprise will likely need revision when adopting a ZTA strategy. A ZTA strategy makes many COOP factors easier as remote workers may have the same access to resources as they had on premises. However, policies like MFA may also have a negative impact if users are not properly trained and lack experience. Users may forget (or not have access to) tokens and enterprise devices during an emergency, and that will impact the speed and effectiveness of enterprise business processes.

## B.5 ZTA Test Environment

TBD – describe NCCoE test lab and tests to be performed

## B.6 References

[1] Cam-Winget N (ed.), Appala S, Pope S, Saint-Andre P (2019) Using Extensible Messaging and Presence Protocol (XMPP) for Security Information Exchange. (Internet Engineering Task Force (IETF)), IETF Request for Comments (RFC) 8600. https://doi.org/10.17487/RFC8600

[2] Software Defined Perimeter Working Group "SDP Specification 1.0" Cloud Security Alliance. April 2014. https://cloudsecurityalliance.org/artifacts/software-defined-perimeter/

[3] Stanton B, Theofanos MF, Spickard Prettyman S, Furman S (2016) Security Fatigue. *IT Professional* 18(5):26-32. https://doi.org/10.1109/MITP.2016.84

[4] Strouble D, Shechtman GM, Alsop AS (2009) Productivity and Usability Effects of Using a Two-Factor Security System. *SAIS 2009 Proceedings* (AIS, Charleston, SC), p 37. Available at http://aisel.aisnet.org/sais2009/37

[5] Weidman J, Grossklags J (2017) I Like It but I Hate It: Employee Perceptions Towards an Institutional Transition to BYOD Second-Factor Authentication. *Proceedings of the 33rd Annual Computer Security Applications Conference (ACSAC 2017)* (ACM, Orlando, FL), pp 212-224. https://doi.org/10.1145/3134600.3134629