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

C O M P U T E R   S E C U R I T Y

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

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
Cybersecurity & Infrastructure Security Agency
Department of Homeland Security

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

August 2020

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, 59 pages (August 2020)
CODEN: NSPUE2

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

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.

Comments on this publication may be submitted to:

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 (ZT) is the term for an evolving set of cybersecurity paradigms that move defenses from static, network-based perimeters to focus on users, assets, and resources. A zero trust architecture (ZTA) uses zero trust principles to plan industrial and enterprise infrastructure and workflows. Zero trust assumes there is no implicit trust granted to assets or user accounts based solely on their physical or network location (i.e., local area networks versus the internet) or based on asset ownership (enterprise or personally owned). Authentication and authorization (both subject and device) are discrete functions performed before a session to an enterprise resource is established. Zero trust is a response to enterprise network trends that include remote users, bring your own device (BYOD), and cloud-based assets that are not located within an enterprise-owned network boundary. Zero trust focuses on protecting resources (assets, services, workflows, network accounts, etc.), 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 zero trust architecture (ZTA) and gives general deployment models and use cases where zero trust could improve an enterprise's overall information technology security posture.

Keywords

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

Acknowledgments

This document is the product of a collaboration between multiple federal agencies and is overseen by the Federal CIO Council. The architecture subgroup is responsible for development of this document, but there are specific individuals who deserve recognition. These include Greg Holden, project manager of the Federal CIO Council ZTA project; Alper Kerman, project manager for the NIST/National Cybersecurity Center of Excellence ZTA effort; and Douglas Montgomery.

Audience

This document is intended to describe zero trust for enterprise security architects. It is meant to aid understanding of zero trust for civilian unclassified systems and provide a road map to migrate and deploy zero trust security concepts to an enterprise environment. Agency cybersecurity managers, network administrators, and managers may also gain insight into zero trust and ZTA from this document. It is not intended to be a single deployment plan for ZTA as an enterprise will have unique business use cases and data assets that require safeguards. Starting with a solid understanding of the organization's business and data will result in a strong approach to zero trust.

Trademark Information

Patent Disclosure Notice

NOTICE: The Information Technology Laboratory (ITL) has requested that holders of patent claims whose use may be required for compliance with the guidance or requirements of this publication disclose such patent claims to ITL. However, holders of patents are not obligated to respond to ITL calls for patents and ITL has not undertaken a patent search in order to identify which, if any, patents may apply to this publication.

Following the ITL call for the identification of patent claims whose use may be required for compliance with the guidance or requirements of this publication, notice of one or more such claims has been received.

By publication, no position is taken by ITL with respect to the validity or scope of any patent claim or of any rights in connection therewith. The known patent holder(s) has (have), however, provided to NIST a letter of assurance stating either (1) a general disclaimer to the effect that it does (they do) not hold and does (do) not currently intend holding any essential patent claim(s), or (2) that it (they) will negotiate royalty-free or royalty-bearing licenses with other parties on a demonstrably nondiscriminatory basis with reasonable terms and conditions.

Details may be obtained from **zerotrust-arch@nist.gov.**

No representation is made or implied that this is the only license that may be required to avoid patent infringement in the use of this publication.

Table of Contents

List of Appendices

List of Figures

List of Tables

1 Introduction

A typical enterprise's 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 legacy methods of perimeter-based network security as there is no single, easily identified perimeter for the enterprise. Perimeter-based network security has also been shown to be insufficient since once attackers breach the perimeter, further lateral movement is unhindered.

This complex enterprise has led to the development of a new model for cybersecurity known as "zero trust" (ZT). A ZT approach is primarily focused on data and service protection but can and should be expanded to include all enterprise assets (devices, infrastructure components, applications, virtual and cloud components) and subjects (end users, applications and other non-human entities that request information from resources). Throughout this document, "subject" will be used unless the section relates directly to a human end user in which "user" will be specifically used instead of the more generic "subject." Zero trust security models assume that an attacker is present in the environment and that an enterprise-owned environment is no different—or no more trustworthy—than any nonenterprise-owned environment. In this new paradigm, an enterprise must assume no implicit trust and continually analyze and evaluate the risks to its assets and business functions and then enact protections to mitigate these risks. In zero trust, these protections usually involve minimizing access to resources (such as data and compute resources and applications/services) to only those subjects and assets identified as needing access as well as continually authenticating and authorizing the identity and security posture of each access request.

A zero trust architecture (ZTA) is an enterprise cybersecurity architecture that is based on zero trust principles and designed to prevent data breaches and limit internal lateral movement. This publication discusses ZTA, its logical components, possible deployment scenarios, and threats. It also presents a general road map for organizations wishing to migrate to a zero trust design approach and discusses relevant federal policies that may impact or influence a zero trust architecture.

ZT is not a single architecture but a set of guiding principles for workflow, system design and operations that can be used to improve the security posture of any classification or sensitivity level [FIPS199]. Transitioning to ZTA is a journey concerning how an organization evaluates risk in its mission and cannot simply be accomplished with 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 their data assets and business functions by use case. Most enterprise infrastructures will operate in a hybrid zero trust/perimeter-based mode while continuing to invest in IT modernization initiatives and improve organization business processes.

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

1.1 History of Zero Trust Efforts Related to Federal Agencies

The concept of zero trust has been present in cybersecurity since before the term "zero trust" was coined. The Defense Information Systems Agency (DISA) and the Department of Defense published their work on a more secure enterprise strategy dubbed "black core" [BCORE]. Black core involved moving from a perimeter-based security model to one that focused on the security of individual transactions. The work of the Jericho Forum in 2004 publicized the idea of de-perimeterization—limiting implicit trust based on network location and the limitations of relying on single, static defenses over a large network segment [JERICHO]. The concepts of de-perimeterization evolved and improved into the larger concept of zero trust, which was later coined by John Kindervag[1] while at Forrester.[2] Zero trust then became the term used to describe various cybersecurity solutions that moved security away from implied trust based on network location and instead focused on evaluating trust on a per-transaction basis. Both private industry and higher education have also undergone this evolution from perimeter-based security to a security strategy based on zero trust principles.

Federal agencies have been urged to move to security based on zero trust principles for more than a decade, building capabilities and policies such as the Federal Information Security Modernization Act (FISMA) followed by the Risk Management Framework (RMF); Federal Identity, Credential, and Access Management (FICAM); Trusted Internet Connections (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 continually analyze and evaluate access requests in a dynamic and granular fashion to a "need to access" basis to mitigate data exposure due to compromised accounts, attackers monitoring a network, and other threats.

1.2 Structure of This Document

The rest of the document is organized as follows:

- **Section 2** defines ZT and ZTA and lists some assumptions when designing a ZTA for an enterprise. This section also includes a list of the tenets of ZT design.
- **Section 3** documents the logical components, or building blocks, of ZT. It is possible that unique implementations compose ZTA components differently yet serve the same logical functionality.

---

[1] https://go.forrester.com/blogs/next-generation-access-and-zero-trust/

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

- **Section 4** lists some possible use cases where a ZTA may make enterprise environments more secure and less prone to successful exploitation. These include enterprises with remote employees, cloud services, and guest networks.
- **Section 5** discusses threats to an enterprise using a ZTA. Many of these threats are similar to any 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 enterprise infrastructure that are guided by ZT tenets.

2 Zero Trust Basics

Zero trust is a cybersecurity paradigm focused on resource protection and the premise that trust is never granted implicitly but must be continually evaluated. Zero trust architecture is an end-to-end approach to enterprise resource and data security that encompasses identity (person and non-person entities), credentials, access management, operations, endpoints, hosting environments, and the interconnecting infrastructure. The initial focus should be on restricting resources to those with a need to access and grant only the minimum privileges (e.g., read, write, delete) needed to perform the mission. Traditionally, agencies (and enterprise networks in general) have focused on perimeter defense and authenticated subjects are given authorized access to a broad collection of resources once on the internal network. As a result, unauthorized lateral movement within the environment 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 and cannot protect subjects outside of the enterprise perimeter (e.g., remote workers, cloud-based services, edge devices, etc.).

An operative definition of zero trust and zero trust architecture is as follows:

> *Zero trust* (ZT) provides a collection of concepts and ideas designed to minimize uncertainty in enforcing accurate, least privilege per-request access decisions in information systems and services in the face of a network viewed as compromised. *Zero trust architecture* (ZTA) is an enterprise's cybersecurity plan that utilizes zero trust concepts and encompasses component relationships, workflow planning, and access policies. Therefore, a zero trust enterprise is the network infrastructure (physical and virtual) and operational policies that are in place for an enterprise as a product of a zero trust architecture plan.

An enterprise decides to adopt zero trust as its core strategy and generate a zero trust architecture as a plan developed with zero trust principles (see Section 2.1 below) in mind. This plan is then deployed to produce a zero trust environment for use in the enterprise.

This definition focuses on the crux of the issue, which is the goal to *prevent unauthorized access to data and services* coupled with making the *access control enforcement as granular a*s *possible*. That is, authorized and approved subjects (combination of user, application (or service), and device) 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 ZT and ZTA are about resource access (e.g., printers, compute resources, Internet of Things [IoT] actuators) and not just data access.

To lessen uncertainties (as they cannot be eliminated), the focus is on authentication, authorization, and shrinking implicit trust zones while maintaining availability and minimizing temporal delays in authentication mechanisms. Access rules are made as granular as possible to enforce least privileges needed to perform the action in the request.

In the abstract model of access shown in Figure 1, a subject needs access to an enterprise resource. Access is granted through a policy decision point (PDP) and corresponding policy enforcement point (PEP).[3]

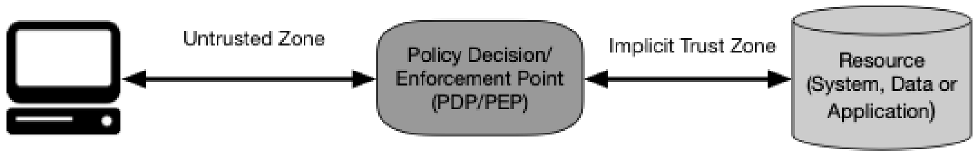

Figure 1: Zero Trust Access

The system must ensure that the subject is authentic 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 to two basic areas: authentication and authorization. What is the level of confidence about the subject's identity for this unique request? Is access to the resource allowable given the level of confidence in the subject's identity? Does the device used for the request have the proper security posture? Are there other factors that should be considered and that change the confidence level (e.g., time, location of subject, subject's security posture)? Overall, enterprises need to develop and maintain dynamic risk-based policies for resource access and set up a system to ensure that these policies are enforced correctly and consistently for individual resource access requests. This means that an enterprise should not rely on implied trustworthiness wherein if the subject has met a base authentication level (e.g., logging into an asset), all subsequent resource requests are assumed to be equally valid.

The "implicit 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 checkpoint (PDP/PEP) to access the boarding gates. The passengers, airport employees, aircraft crew, etc., mill about in the terminal area, and all the individuals are considered trusted. In this model, the implicit trust zone is the boarding area.

The PDP/PEP applies a set of controls so that all traffic beyond the PEP has a common level of trust. The PDP/PEP cannot apply additional policies beyond its location in the flow of traffic. To allow the PDP/PEP to be as specific as possible, the implicit trust zone must be as small as possible.

Zero trust provides a set of principles and concepts around moving the PDP/PEPs closer to the resource. The idea is to explicitly authenticate and authorize all subjects, assets and workflows that make up the enterprise.

[3] Part of the concepts defined in OASIS XACML 2.0 https://docs.oasis-open.org/xacml/2.0/access_control-xacml-2.0-core-spec-os.pdf

2.1 Tenets of Zero Trust

Many definitions and discussions of ZT stress the concept of removing wide-area perimeter defenses (e.g., enterprise firewalls) as a factor. However, most of these definitions continue to define themselves in relation to perimeters in some way (such as micro-segmentation or micro-perimeters; see Section 3.1) as part of the functional capabilities of a ZTA. The following is an attempt to define ZT and ZTA in terms of basic tenets that should be involved rather than what is excluded. These tenets are the ideal goal, though it must be acknowledged that not all tenets may be fully implemented in their purest form for a given strategy.

A zero trust architecture is designed and deployed with adherence to the following zero trust basic tenets:

1. **All data sources and computing services are considered resources.** A network may be composed of multiple classes of devices. A network may also have small footprint devices that send data to aggregators/storage, software as a service (SaaS), systems sending instructions to actuators, and other functions. Also, an enterprise may decide to classify personally owned devices as resources if they can access enterprise-owned resources.
2. **All communication is secured regardless of network location.** Network location alone does not imply trust. Access requests from assets 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 nonenterprise-owned network. In other words, trust should not be automatically granted based on the device being on enterprise network infrastructure. All communication should be done in the most secure manner available, protect confidentiality and integrity, and provide source authentication.
3. **Access to individual enterprise resources is granted on a per-session basis.** Trust in the requester is evaluated before the access is granted. Access should also be granted with the least privileges needed to complete the task. This could mean only "sometime recently" for this particular transaction and may not occur directly before initiating a session or performing a transaction with a resource. However, authentication and authorization to one resource will not automatically grant access to a different resource.
4. **Access to resources is determined by dynamic policy—including the observable state of client identity, application/service, and the requesting asset—and may include other behavioral and environmental attributes.** An organization protects resources by defining what resources it has, who its members are (or ability to authenticate users from a federated community), and what access to resources those members need. For zero trust, client identity can include the user account (or service identity) and any associated attributes assigned by the enterprise to that account or artifacts to authenticate automated tasks. Requesting asset state can include device characteristics such as software versions installed, network location, time/date of request, previously observed behavior, and installed credentials. Behavioral attributes include, but not limited to, automated subject analytics, device analytics, and measured deviations from observed usage patterns. Policy is the set of access rules based on attributes that an organization assigns to a subject, data asset, or application. Environmental attributes may include such factors as requestor

network location, time, reported active attacks, etc. These rules and attributes are based on the needs of the business process and acceptable level of risk. Resource access and action permission policies can vary based on the sensitivity of the resource/data. Least privilege principles are applied to restrict both visibility and accessibility.

5. **The enterprise monitors and measures the integrity and security posture of all owned and associated assets.** No asset is inherently trusted. The enterprise evaluates the security posture of the asset when evaluating a resource request. An enterprise implementing a ZTA should establish a continuous diagnostics and mitigation (CDM) or similar system to monitor the state of devices and applications and should apply patches/fixes as needed. Assets that are discovered to be subverted, have known vulnerabilities, and/or are not managed by the enterprise may be treated differently (including denial of all connections to enterprise resources) than devices owned by or associated with the enterprise that are deemed to be in their most secure state. This may also apply to associated devices (e.g., personally owned devices) that may be allowed to access some resources but not others. This, too, requires a robust monitoring and reporting system in place to provide actionable data about the current state of enterprise resources.
6. **All resource authentication and authorization are dynamic and strictly enforced before access is allowed.** This is a constant cycle of obtaining access, scanning and assessing threats, adapting, and continually reevaluating trust in ongoing communication. An enterprise implementing a ZTA would be expected to have Identity, Credential, and Access Management (ICAM) and asset management systems in place. This includes the use of multifactor authentication (MFA) for access to some or all enterprise resources. Continual monitoring with possible reauthentication and reauthorization occurs throughout user transactions, as defined and enforced by policy (e.g., time-based, new resource requested, resource modification, anomalous subject activity detected) that strives to achieve a balance of security, availability, usability, and cost-efficiency.
7. **The enterprise collects as much information as possible about the current state of assets, network infrastructure and communications and uses it to improve its security posture.** An enterprise should collect data about asset security posture, network traffic and access requests, process that data, and use any insight gained to improve policy creation and enforcement. This data can also be used to provide context for access requests from subjects (see Section 3.3.1).

The above tenets attempt to be technology agnostic. For example, “user identification (ID)” could include several factors such as username/password, certificates, and onetime password. These tenets apply to work done within an organization or in collaboration with one or more partner organizations and not to anonymous public or consumer-facing business processes. An organization cannot impose internal policies on external actors (e.g., customers or general internet users) but may be able to implement some ZT-based policies on nonenterprise users who have a special relationship with the organization (e.g. registered customers, employee dependents, etc.).

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 operating on nonenterprise-owned network infrastructure (e.g., public Wi-Fi or public cloud providers). These assumptions are used to direct the formation of a ZTA. The network in an enterprise implementing a ZTA should be developed with the ZTA tenets outlined above and with the following assumptions.

1. **The entire enterprise private network is not considered an implicit trust zone.** Assets should always act as if an attacker is present on the enterprise network, and communication should be done in the most secure manner available (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 nonenterprise-owned assets that need network access to perform their role. This includes bring-your-own-device (BYOD) policies that allow enterprise subjects to use nonenterprise-owned devices to access enterprise resources.
3. **No resource is inherently trusted.** Every asset must have its security posture evaluated via a PEP before a request is granted to an enterprise-owned resource (similar to tenet 6 above for assets as well as subjects). This evaluation should be continual for as long as the session lasts. Enterprise-owned devices may have artifacts that enable authentication and provide a confidence level higher than the same request coming from nonenterprise-owned devices. Subject credentials alone are insufficient for device authentication to an enterprise resource.
4. **Not all enterprise resources are on enterprise-owned infrastructure.** Resources include remote enterprise subjects as well as cloud services. Enterprise-owned or -managed assets may need to utilize the local (i.e., nonenterprise) network for basic connectivity and network services (e.g., DNS resolution).
5. **Remote enterprise subjects and assets cannot fully trust their local network connection.** Remote subjects should assume that the local (i.e., nonenterprise-owned) network is hostile. Assets should assume that all traffic is being monitored and potentially modified. All connection requests should be authenticated and authorized, and all communications should be done in the most secure manner possible (i.e., provide confidentiality, integrity protection, and source authentication). See the tenets of ZTA above.
6. **Assets and workflows moving between enterprise and nonenterprise infrastructure should have a consistent security policy and posture.** Assets and workloads should retain their security posture when moving to or from enterprise-owned infrastructure. This includes devices that move from enterprise networks to nonenterprise networks (i.e. remote users). This also includes workloads migrating from on-premises data centers to nonenterprise cloud instances.

3 Logical Components of Zero Trust Architecture

There are numerous logical components that make up a ZTA deployment in an enterprise. These components may be operated as an on-premises service or through a cloud-based service. The conceptual framework model in Figure 2 shows the basic relationship between 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). The ZTA logical components use a separate control plane to communicate, while application data is communicated on a data plane (see Section 3.4).

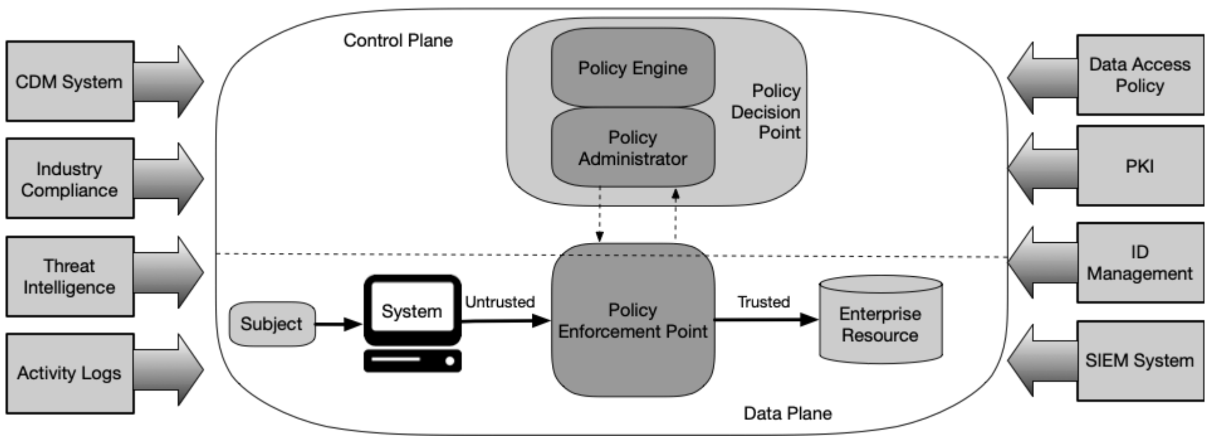

Figure 2: Core Zero Trust Logical Components

The component descriptions:

- **Policy engine (PE):** This component is responsible for the ultimate decision to grant access to a resource for a given subject. The PE uses enterprise policy as well as input from external sources (e.g., CDM systems, threat intelligence services described below) as input to a trust algorithm (see Section 3.3 for more details) to grant, deny, or revoke access to the resource. The PE is paired with the policy administrator component. The policy engine makes and logs the decision (as approved, or denied), and the policy administrator executes the decision.
- **Policy administrator (PA):** This component is responsible for establishing and/or shutting down the communication path between a subject and a resource (via commands to relevant PEPs). It would generate any session-specific authentication and authentication token or credential used by a client to access an enterprise resource. It is closely tied to the PE and relies on its decision to ultimately allow or deny a session. If the session is authorized and the request authenticated, the PA configures the PEP to allow the session to start. If the session is denied (or a previous approval is countermanded), the PA signals to the PEP to shut down the connection. Some implementations may treat the PE and PA as a single service; here, it is divided into its

two logical components. The PA communicates with the PEP when creating the communication path. This communication is done via the control plane.

- **Policy enforcement point (PEP):** This system is responsible for enabling, monitoring, and eventually terminating connections between a subject and an enterprise resource. The PEP communicates with the PA to forward requests and/or receive policy updates from the PA. This is a single logical component in ZTA but may be broken into two different components: the client (e.g., agent on a laptop) and resource side (e.g., gateway component in front of resource that controls access) or a single portal component that acts as a gatekeeper for communication paths. Beyond the PEP is the trust zone (see Section 2) hosting the enterprise resource.

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

- **Continuous diagnostics and mitigation (CDM) system:** This gathers information about the enterprise asset's current state and applies updates to configuration and software components. An enterprise CDM system provides the policy engine with the information about the asset making an access request, such as whether it is running the appropriate patched operating system (OS), the integrity of enterprise-approved software components or presence of non-approved components and whether the asset has any known vulnerabilities. CDM systems are also responsible for identifying and potentially enforcing a subset of polices on nonenterprise devices active on enterprise infrastructure.

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

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

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

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

- **Enterprise public key infrastructure (PKI):** This system is responsible for generating and logging certificates issued by the enterprise to resources, subjects, services and applications. This also includes the global certificate authority ecosystem and the Federal PKI,[4] which may or may not be integrated with the enterprise PKI. This could also be a PKI that is not built upon X.509 certificates.
- **ID management system:** This is responsible for creating, storing, and managing enterprise user accounts and identity records (e.g., lightweight directory access protocol (LDAP) server). This system contains the necessary subject information (e.g., name, email address, certificates) and other enterprise characteristics such as role, access attributes, and assigned assets. This system often utilizes other systems (such as a PKI) for artifacts associated with user accounts. This system may be part of a larger federated community and may include nonenterprise employees or links to nonenterprise assets for collaboration.
- **Security information and event management (SIEM) system:** This collects security-centric information for later analysis. This data is then used to refine policies and warn of possible attacks against enterprise assets.

3.1 Variations of Zero Trust Architecture Approaches

There are several ways that an enterprise can enact a ZTA for workflows. These approaches vary in the components used and in the main source of policy rules for an organization. Each approach implements all the tenets of ZT (see Section 2.1) but may use one or two (or one component) as the main driver of policies. A full ZT solution will include elements of all three approaches. The approaches include enhanced identity governance–driven, logical micro-segmentation, and network-based segmentation.

Certain approaches lend themselves to some use cases more than others. An organization looking to develop a ZTA for its enterprise may find that its chosen use case and existing policies point to one approach over others. That does not mean the other approaches would not work but rather that other approaches may be more difficult to implement and may require more fundamental changes to how the enterprise currently conducts business flows.

3.1.1 ZTA Using Enhanced Identity Governance

The enhanced identity governance approach to developing a ZTA uses the identity of actors as the key component of policy creation. If it were not for subjects requesting access to enterprise resources, there would be no need to create access polices. For this approach, enterprise resource access policies are based on identity and assigned attributes. The primary requirement for resource access is based on the access privileges granted to the given subject. Other factors such as device used, asset status, and environmental factors may alter the final confidence level calculation (and ultimate access authorization) or tailor the result in some way, such as granting only partial access to a given data source based on network location. Individual resources or PEP

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

components protecting the resource must have a way to forward requests to a policy engine service or authenticate the subject and approve the request before granting access.

Enhanced identity governance-based approaches for enterprises are often employed using an open network model or an enterprise network with visitor access or frequent nonenterprise devices on the network (such as with the use case in Section 4.3 below). Network access is initially granted to all assets but access to enterprise resources are restricted to identities with the appropriate access privileges. There is a downside in granting basic network connectivity as malicious actors could still attempt network reconnaissance and/or use the network to launch denial of service attacks either internally or against a third party. Enterprises still need to monitor and respond to such behavior before it impacts workflows.

The identity-driven approach works well with the resource portal model (see Section 3.2.3) since device identity and status provide secondary support data to access decisions. Other models work as well, depending on policies in place. Identity-driven approaches also work well for enterprises that use cloud-based applications/services that may not allow for enterprise-owned or -operated ZT security components to be used (such as many SaaS offerings). The enterprise can use the identity of requestors to form and enforce policy on these platforms.

3.1.2 ZTA Using Micro-Segmentation

An enterprise may choose to implement a ZTA based on placing individual or groups of resources on a unique network segment protected by a gateway security component. In this approach, the enterprise places infrastructure devices such as intelligent switches (or routers) or next generation firewalls (NGFWs) or special purpose gateway devices to act as PEPs protecting each resource or small group of related resources. Alternatively (or additionally), the enterprise may choose to implement host-based micro-segmentation using software agents (see Section 3.2.1) or firewalls on the endpoint asset(s), These gateway devices dynamically grant access to individual requests from a client, asset or service. Depending on the model, the gateway may be the sole PEP component or part of a multipart PEP consisting of the gateway and client-side agent (see Section 3.2.1).

This approach applies to a variety of use cases and deployment models as the protecting device acts as the PEP, with management of said devices acting as the PE/PA component. This approach requires an identity governance program (IGP) to fully function but relies on the gateway components to act as the PEP that shields resources from unauthorized access and/or discovery.

The key necessity to this approach is that the PEP components are managed and should be able to react and reconfigure as needed to respond to threats or change in the workflow. It is possible to implement some features of a micro-segmented enterprise by using less advanced gateway devices and even stateless firewalls, but the administration cost and difficulty to quickly adapt to changes make this a very poor choice.

3.1.3 ZTA Using Network Infrastructure and Software Defined Perimeters

The last approach uses the network infrastructure to implement a ZTA. The ZTA implementation could be achieved by using an overlay network (i.e., layer 7 but also could be set up lower of the

OSI network stack). These approaches are sometimes referred to as software defined perimeter (SDP) approaches and frequently include concepts from Software Defined Networks (SDN) [SDNBOOK] and intent-based networking (IBN) [IBNVN]. In this approach, the PA acts as the network controller that sets up and reconfigures the network based on the decisions made by the PE. The clients continue to request access via PEPs, which are managed by the PA component.

When the approach is implemented at the application network layer (i.e., layer 7), the most common deployment model is the agent/gateway (see Section 3.2.1). In this implementation, the agent and resource gateway (acting as the single PEP and configured by the PA) establish a secure channel used for communication between the client and resource. There may be other variations of this model, as well for cloud virtual networks, non-IP based networks, etc.

3.2 Deployed Variations of the Abstract Architecture

All of the above components are logical components. They do not necessarily need to be unique systems. A single asset 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-managed 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 some ZT product offerings currently available on the market, the PE and PA components are combined in a single service.

There are several variations on 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.

3.2.1 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 asset 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 communicates only with the gateway, essentially serving as a proxy for the resource. The agent is a software component that directs some (or all) traffic to the appropriate PEP in order for requests to be evaluated. The gateway is responsible for communicating with the policy administrator and allowing only approved communication paths configured by the policy administrator (see Figure 3).

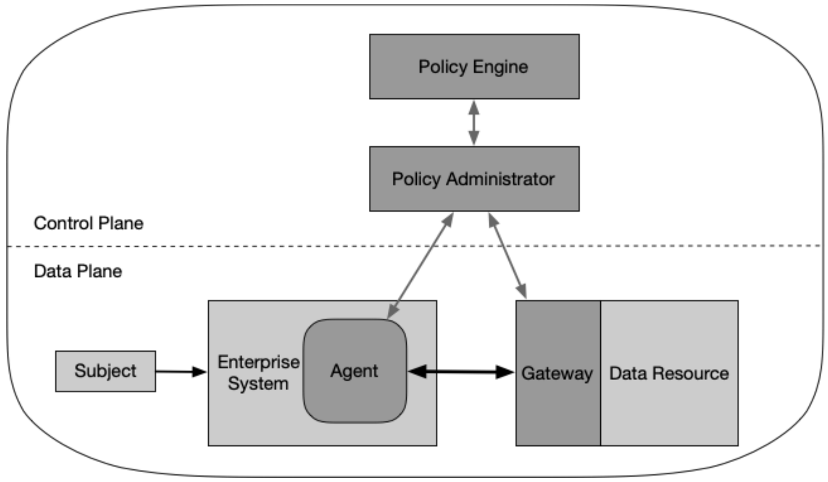

Figure 3: Device Agent/Gateway Model

In a typical scenario, a subject with an enterprise-issued laptop wishes to connect to an enterprise resource (e.g., human resources application/database). The access request is taken by the local agent, and the request is forwarded to the policy administrator. The policy administrator and policy engine could be an enterprise local asset 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 information such as an internet protocol (IP) address, port information, session key, or similar security artifacts. The device agent and gateway then connect, and encrypted application/service 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 time-out, failure to reauthenticate).

This model is best utilized for enterprises that have a robust device management program in place as well as 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 appropriate for enterprises that do not want a BYOD policy in place. Access is possible only via the device agent, which can be placed on enterprise-owned assets.

3.2.2 Enclave-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 assets 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 application programming interface [API] that can be used to communicate with a gateway). This deployment model may also be useful for enterprises that use cloud-based micro-services for a single business processes (e.g., user notification, database lookup, salary disbursement). In this model, the entire private cloud is located behind a gateway.

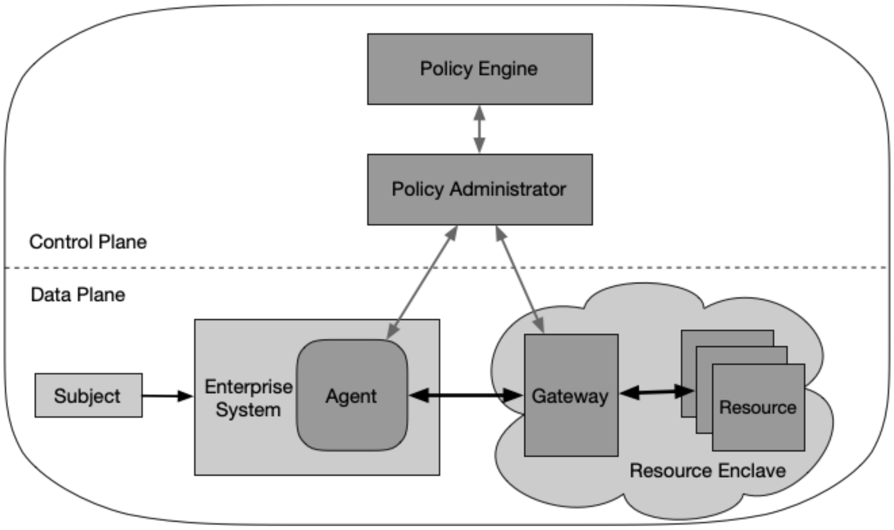

Figure 4: Enclave Gateway Model

It is possible for this model to be a hybrid with the device agent/gateway model. In this model, enterprise assets have a device agent that is used to connect to enclave 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 a robust asset and configuration management program in place to install/configure the device agents. The downside is that the gateway protects a collection of resources and may not be able to protect each resource individually. This may also allow for subjects to see resources which they do not have privileges to access.

3.2.3 Resource Portal-Based Deployment

In this deployment model, the PEP is a single component that acts as a gateway for subject requests. The gateway portal can be for an individual resource or a secure enclave 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.

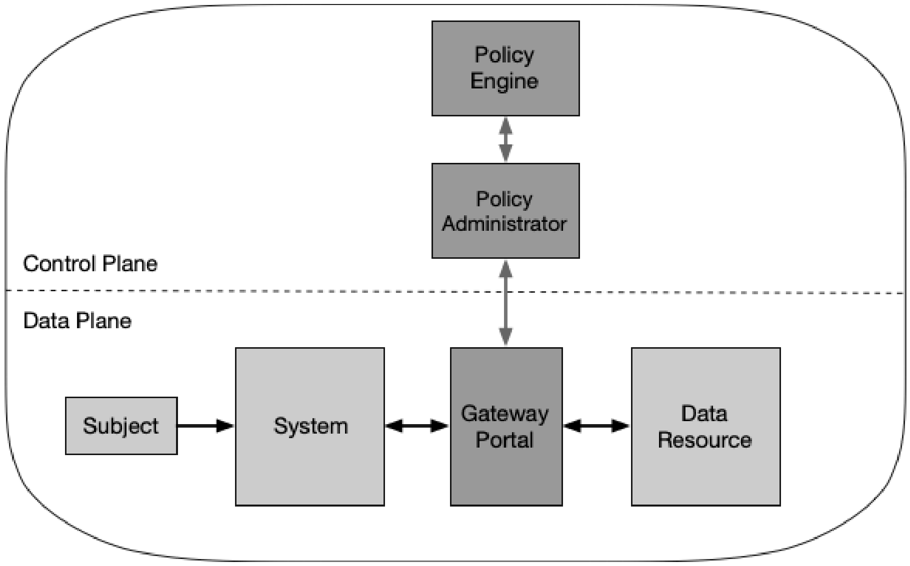

Figure 5: Resource Portal Model

The primary benefit of this model over the others is that a software component does not need to be installed on all client devices. This model is also more flexible for BYOD policies and inter-organizational 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. This model can only scan and analyze assets and devices once they connect to the PEP portal and may not be able to continuously monitor them for malware, unpatched vulnerabilities, and appropriate configuration.

The main difference with this model is there is no local agent that handles requests, and so the enterprise may not have full visibility or arbitrary control over assets as it can only see/scan them when they connect to a portal. The enterprise may be able to employ measures such as browser isolation to mitigate or compensate. These assets may be invisible to the enterprise between these sessions. This model also allows attackers to discover and attempt to access the portal or attempt a denial-of-service (DoS) attack against the portal. The portal systems should be well-provisioned to provide availability against a DoS attack or network disruption.

3.2.4 Device Application Sandboxing

Another variation of the agent/gateway deployment model is having vetted applications or processes run compartmentalized on assets. These compartments could be virtual machines, containers, or some other implementation, but the goal is the same: to protect the application or instances of applications from a possibly compromised host or other applications running on the asset.

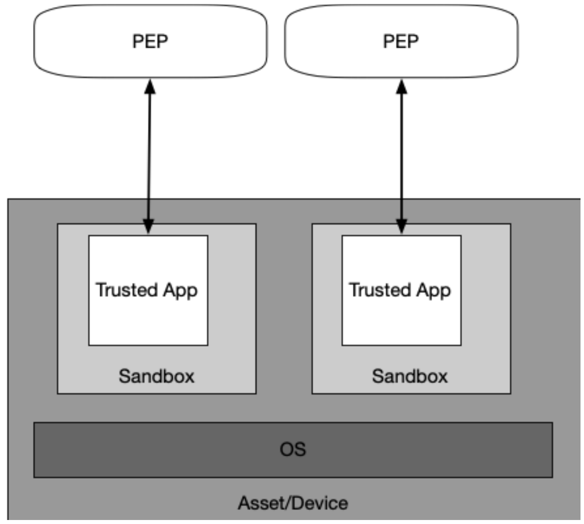

Figure 6: Application Sandboxes

In Figure 6, the subject device runs approved, vetted applications in a sandbox. The applications can communicate with the PEP to request access to resources, but the PEP will refuse requests from other applications on the asset. 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 from the rest of the asset. If the asset cannot be scanned for vulnerabilities, these individual, sandboxed applications may be protected from a potential malware infection on the host asset. One of the disadvantages of this model is that enterprises must maintain these sandboxed applications for all assets and may not have full visibility into client assets. The enterprise also needs to make sure each sandboxed application is secure, which may require more effort than simply monitoring devices.

3.3 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 as its primary thought process. The trust algorithm (TA) 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 (see Section 3): the policy database with observable information about subjects, subject attributes and roles, historical subject behavior patterns, threat intelligence sources, and other metadata sources. The process can be grouped into broad categories and visualized in Figure 7.

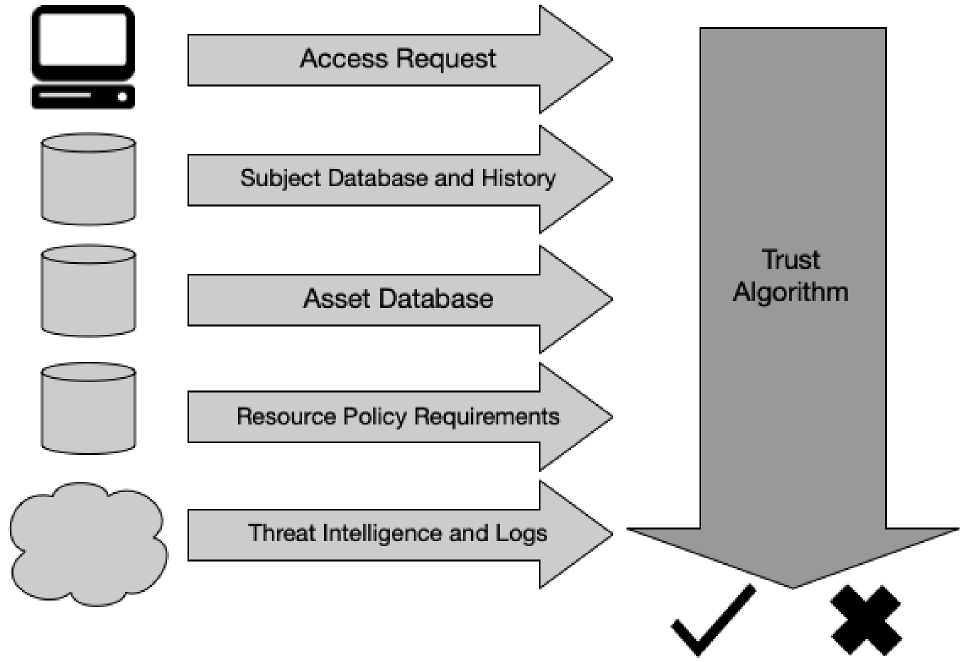

Figure 7: Trust Algorithm Input

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

- **Access request:** This is the actual request from the subject. The resource requested is the primary information used, but information about the requester is also used. This can include OS version, software used (e.g., does the requesting application appear on a list of approved applications?), and patch level. Depending on these factors and the asset security posture, access to assets might be restricted or denied.
- **Subject database:** This is the "who" that is requesting access to a resource [SP800-63]. This is the set of subjects (human and processes) of the enterprise or collaborators and a collection of subject attributes/privileges assigned. These subjects and attributes form the basis of policies for resource access [SP800-162] [NISTIR 7987]. User identities can include a mix of logical identity (e.g., account ID) and results of authentication checks performed by PEPs. Attributes of identity that can be factored into deriving the confidence level include time and geolocation. A collection of privileges given to multiple subjects could be thought of as a role, but privileges should be assigned to a subject on an individual basis and not simply because they may fit into a particular role in the organization. This collection should be encoded and stored in an ID management system and policy database. This may also include data about past observed subject behavior in some (TA) variants (see Section 3.3.1).
- **Asset database (and observable status):** This is the database that contains the known status of each enterprise-owned (and possibly known nonenterprise/BYOD) asset (physical and virtual, to some extent). This is compared to the observable status of the asset making the request and can include OS version, software present, and its integrity,

location (network location and geolocation), and patch level. Depending on the asset state compared with this database, access to assets might be restricted or denied.

- **Resource requirements:** This set of policies complements the user ID and attributes database [SP800-63] and defines the minimal requirements for access to the resource. Requirements may include authenticator assurance levels, such as MFA network location (e.g., deny access from overseas IP addresses), data sensitivity, and requests for asset 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 could also include specific information about communication seen from the device that may be suspect (such as queries for possible malware command and control nodes). These feeds can be external services or internal scans and discoveries and can include attack signatures and mitigations. This is the only component that will most likely be under the control of a service rather than the enterprise.

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 execution. The PA's job is to configure the necessary PEPs to enable authorized communication. Depending on how the ZTA is deployed, this may involve sending authentication results and connection configuration information to gateways and agents or resource portals. PAs may also place a hold or pause on a communication session to reauthenticate and reauthorize the connection in accordance with policy requirements. The PA is also responsible for issuing the command to terminate the connection based on policy (e.g., after a time-out, when the workflow has been completed, due to a security alert).

3.3.1 Trust Algorithm Variations

There are different ways to implement a TA. Different implementers may wish to weigh the above factors differently according to the factors' perceived importance. There are two other major characteristics that can be used to differentiate TAs. The first is how the factors are evaluated, whether as binary decisions or weighted parts of a whole "score" or confidence level. The second is how requests are evaluated in relation to other requests by the same subject, application/service, or device.

- **Criteria- versus score-based:** A criteria-based TA assumes a set of qualified attributes that must be met before access is granted to a resource or an action (e.g., read/write) is allowed. These criteria are configured by the enterprise and should be independently configured for every resource. Access is granted or an action applied to a resource only if all the criteria are met. A score-based TA computes a confidence level 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, or the action is performed.

Otherwise, the request is denied, or access privileges are reduced (e.g., read access is granted but not write access for a file).

- **Singular versus contextual:** A singular TA treats each request individually and does not take the subject history into consideration when making its evaluation. This can allow faster evaluations, but there is a risk that an attack can go undetected if it stays within a subject's allowed role. A contextual TA takes the subject or network agent's recent history into consideration when evaluating access requests. This means the PE must maintain some state information on all subjects 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 subject. This also means that the PE must be informed of user behavior by the PA (and PEPs) that subjects interact with when communicating. Analysis of subject behavior can be used to provide a model of acceptable use, and deviations from this behavior could trigger additional authentication checks or resource request denials.

The two factors are not always dependent on each other. It is possible to have a TA that assigns a confidence level to every subject and/or device and still considers every access request independently (i.e., singular). However, contextual, score-based TAs would provide the ability to offer more dynamic and granular access control, since the score provides a current confidence level for the requesting account and adapts to changing factors more quickly than static policies modified by human administrators.

Ideally, a ZTA trust algorithm should be contextual, but this may not always be possible with the infrastructure components available to the enterprise. A contextual TA can mitigate threats where an attacker stays close to a "normal" set of access requests for a compromised subject account or insider attack. It is important to balance security, usability, and cost-effectiveness when defining and implementing trust algorithms. Continually prompting a subject for reauthentication 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 to 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. A contextual TA may also send an alert if someone is making access requests after normal business hours as this could be an attacker exfiltrating records by using a compromised HR account. These are examples where a contextual TA can detect an attack whereas a singular TA may fail to detect the new behavior. In another example, an accountant who typically accesses the financial system during normal business hours 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 subject to satisfy a more stringent confidence level or other criteria as outlined in NIST Special Publication 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 implementation 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. How long this tuning phase lasts depends on the enterprise-defined metrics for progress and tolerance for incorrect access denials/approvals for the resources used in the workflow.

3.4 Network/Environment Components

In a ZT environment, there should be a separation (logical or possibly physical) of the communication flows used to control and configure the network and application/service 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/service communication flows [Gilman].

The control plane is used by various infrastructure components (both enterprise-owned and from service providers) to maintain and configure assets; judge, grant, or deny access to resources; and perform any necessary operations to set up communication paths between resources. The data plane is used for actual communication between software components. This communication channel may not be possible before the path has been established via the control plane. For example, the control plane could be used by the PA and PEP to set up the communication path between the subject and the enterprise resource. The application/service workload would then use the data plane path that was established.

3.4.1 Network Requirements to Support ZTA

1. **Enterprise assets have basic network connectivity.** The local area network (LAN), enterprise controlled or not, provides basic routing and infrastructure (e.g., DNS). The remote enterprise asset may not necessarily use all infrastructure services.

2. **The enterprise must be able to distinguish between what assets are owned or managed by the enterprise and the devices' current security posture.** This is determined by enterprise-issued credentials and not using information that cannot be authenticated information (e.g., network MAC addresses that can be spoofed).

3. **The enterprise can observe all network traffic.** The enterprise records packets seen on the data plane, even if it is not be able to perform application layer inspection (i.e., OSI layer 7) on all packets. The enterprise filters out metadata about the connection (e.g., destination, time, device identity) to dynamically update policies and inform the PE as it evaluates access requests.

4. **Enterprise resources should not be reachable without accessing a PEP.** Enterprise resources do not accept arbitrary incoming connections from the internet. Resources accept custom-configured connections only after a client has been authenticated and authorized. These communication paths are set up by the PEP. Resources may not even be discoverable without accessing a PEP. This prevents attackers from identifying targets via scanning and/or launching DoS attacks against resources located behind PEPs. Note that not all resources should be hidden this way; some network infrastructure components (e.g., DNS servers) must be accessible.

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 not directly accessible by enterprise assets and resources. The data plane is used for

application/service data traffic. The policy engine, policy administrator, and PEPs use the control plane to communicate and manage communication paths between assets. The PEPs must be able to send and receive messages from both the data and control planes.

6. **Enterprise assets can reach the PEP component.** Enterprise subjects must be able to access the PEP component to gain access to resources. This could take the form of a web portal, network device, or software agent on the enterprise asset that enables the connection.

7. **The PEP is the only component that accesses the policy administrator as part of a business flow.** Each PEP operating on the enterprise network has a connection to the policy administrator to establish communication paths from clients to resources. All enterprise business process traffic passes through one or more PEPs.

8. **Remote enterprise assets should be able to access enterprise resources without needing to traverse enterprise network infrastructure first.** For example, a remote subject should not be required to use a link back to the enterprise network (i.e., virtual private network [VPN]) to access services utilized by the enterprise and hosted by a public cloud provider (e.g., email).

9. **The infrastructure used to support the ZTA access decision process should be made scalable to account for changes in process load.** The PE(s), PA(s), and PEPs used in a ZTA become the key components in any business process. Delay or inability to reach a PEP (or inability of the PEPs to reach the PA/PE) negatively impacts the ability to perform the workflow. An enterprise implementing a ZTA needs to provision the components for the expected workload or be able to rapidly scale the infrastructure to handle increased usage when needed.

10. **Enterprise assets may not be able to reach certain PEPs due to policy or observable factors.** For example, there may be a policy stating that mobile assets may not be able to reach certain resources if the requesting asset is located outside of the enterprise's home country. These factors could be based on location (geolocation or network location), device type, or other criteria.

4 Deployment Scenarios/Use Cases

Any enterprise environment can be designed with zero trust tenets in mind. Most organizations already have some elements of zero trust in their enterprise infrastructure or are on their way through implementation of information security and resiliency policies and best practices. Several deployment scenarios and use cases lend themselves 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 can benefit from a zero trust architecture.

In the use cases below, ZTA is not explicitly indicated since the enterprise likely has both perimeter-based and possibly ZTA infrastructures. As discussed in Section 7.2, there will likely be a period when ZTA components and perimeter-based network infrastructure are concurrently in operation in an enterprise.

4.1 Enterprise with Satellite Facilities

The most common scenario involves 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 to perform their tasks. The enterprise may have a Multiprotocol Label Switch (MPLS) link to the enterprise HQ network but may not have adequate bandwidth for all traffic or may not wish for traffic destined for cloud-based applications/services to traverse through the enterprise HQ network. Likewise, employees may be teleworking or in a remote location and 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 or restrict actions to more sensitive resources (e.g., HR database).

In this use case, the PE/PA(s) is often hosted as a cloud service (which usually provides superior availability and would not require remote workers to rely on enterprise infrastructure to access cloud resources) with end assets having an installed agent (see Section 3.2.1) or accessing a resource portal (see Section 3.2.3). It may not be most responsive to have the PE/PA(s) hosted on the enterprise local network as remote offices and workers must send all traffic back to the enterprise network to reach applications/services hosted by cloud services.

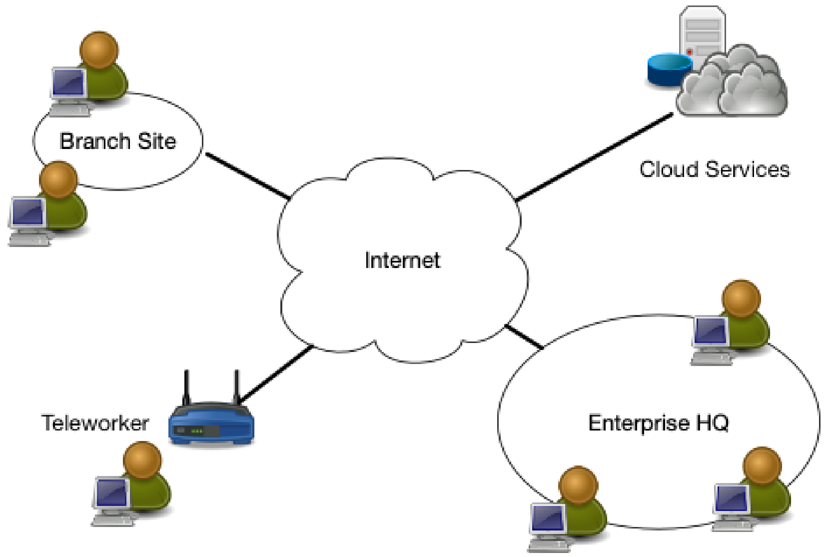

Figure 8: Enterprise with Remote Employees

4.2 Multi-cloud/Cloud-to-Cloud Enterprise

One increasingly common use case for deploying a ZTA 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/services and data. Sometimes, the application/service is hosted on a cloud service that is separate from 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.

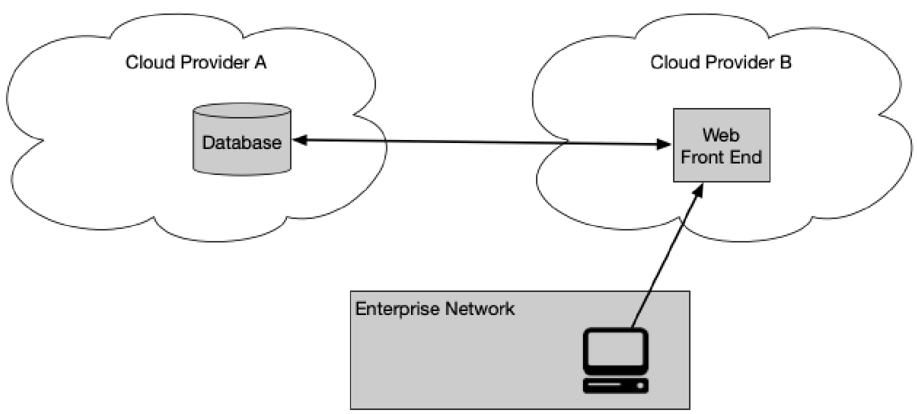

Figure 9: Multi-cloud Use Case

This use case is the server-server implementation of the CSA's software defined perimeter (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, ZT principles take the view that there should be no difference between enterprise-owned and -operated network infrastructure and infrastructure owned and operated by any other service provider. The zero trust approach to multi-cloud use is to place PEPs at the access points of each application/service and data source. The PE and PA may be services 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 the enterprise. One challenge is that different cloud providers have unique ways of implementing similar functionality. Enterprise architects will need to be aware of the how to implement their enterprise ZTA with each cloud provider they utilize.

4.3 Enterprise with Contracted Services and/or Nonemployee 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/services, databases, and assets. These include services contracted out to providers who may occasionally be on-site to provide maintenance (e.g., smart heating and lighting systems that are owned and managed by external providers). These visitors and service providers will need network connectivity to perform their tasks. A zero trust enterprise could facilitate this by allowing these devices and any visiting service technician access to the internet while obscuring enterprise resources.

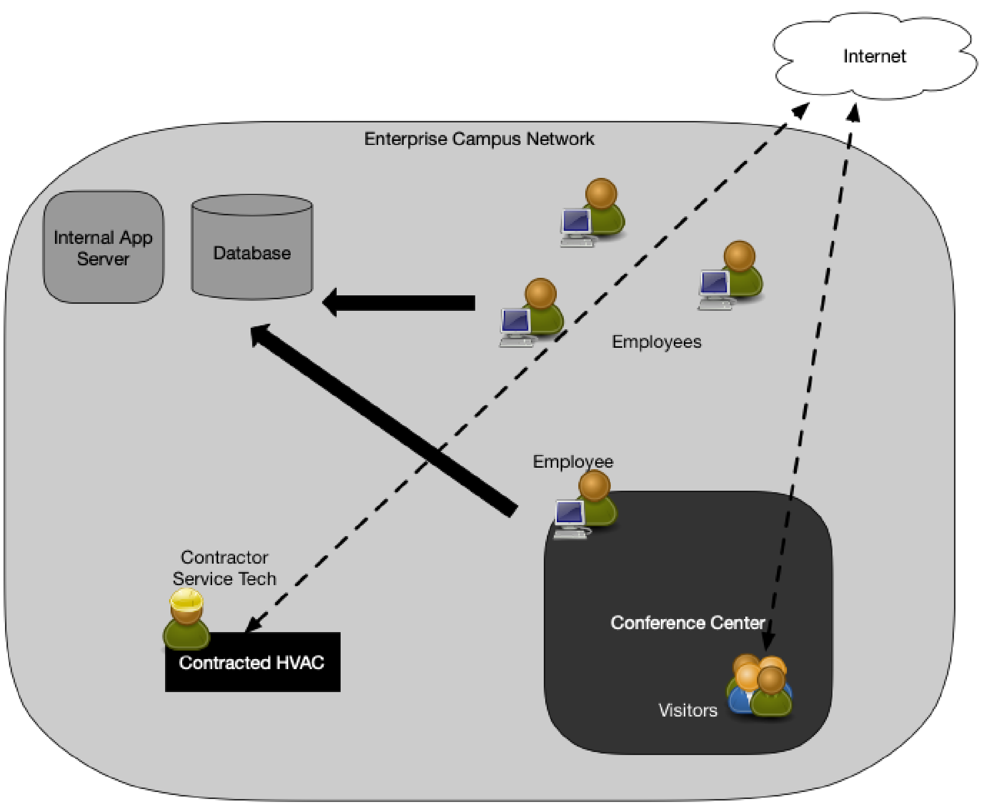

Figure 10: Enterprise with Nonemployee Access

In this example, the organization also has a conference center where visitors interact with employees. Again, with a ZTA approach of SDPs, employee devices and subjects 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 may not even be able to discover enterprise services via network scans (i.e., prevent active network reconnaissance/east-west movement).

In this use case, the PE(s) and PA(s) could be hosted as a cloud service or on the LAN (assuming little or no use of cloud-hosted services). The enterprise assets could have an installed agent (see Section 3.2.1) or access resources via a portal (see Section 3.2.3). The PA(s) ensures that all nonenterprise assets (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, but this can quickly become difficult to manage. Having both organizations enrolled in a federated ID management system would allow quicker establishment of these relationships provided that both organizations' PEPs can authenticate subjects in a federated ID community.

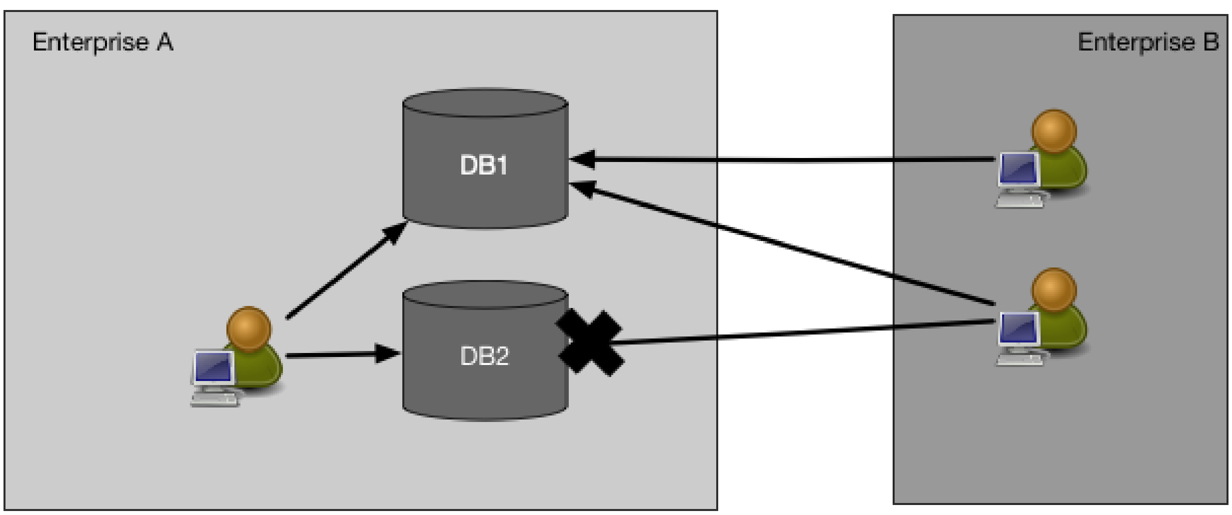

Figure 11: Cross-Enterprise Collaboration

This scenario can be similar to Use Case 1 (Section 4.1) as employees of both enterprises may not be located on their organizations' network infrastructures, and the resource they need to access may be within one enterprise environment or hosted in the cloud. This means that there do not need to be complex firewall rules or enterprise-wide access control lists (ACLs) allowing certain IP addresses belonging to Enterprise B to access resources in Enterprise A based on Enterprise A's access policies. How this access is accomplished depends on the technology in use. Similar to Use Case 1, a PE and PA hosted as a cloud service may provide availability to all parties without having to establish a VPN or similar. The employees of Enterprise B may be asked to install a software agent on their asset or access the necessary data resources through a web gateway (see Section 3.2.3).

4.5 Enterprise with Public- or Customer-Facing Services

A common feature in many enterprises is a public-facing service that may or may not include user registration (i.e., users must create or have been issued a set of login credentials). Such services could be for the general public, a set of customers with an existing business relationship, or a special set of nonenterprise users such as employee dependents. In all cases, it is likely that requesting assets are not enterprise-owned, and the enterprise is constrained as to what internal cybersecurity polices can be enforced.

For a general, public-facing resource that does not require login credentials to access (e.g., public web page), the tenets of ZTA do not directly apply. The enterprise cannot strictly control the state of requesting assets, and anonymous public resources (e.g., a public web page) do not require credentials in order to be accessed.

Enterprises may establish policies for registered public users such as customers (i.e., those with a business relationship) and special users (e.g., employee dependents). If the users are required to produce or are issued credentials, the enterprise may institute policies regarding password length, life cycle, and other details and may provide MFA as an option or requirement. However, enterprises are limited in the policies they can implement for this class of user. Information about incoming requests may be useful in determining the state of the public service and detecting possible attacks masquerading as legitimate users. For example, a registered user portal is known to be accessed by registered customers using one of a set of common web browsers. A sudden increase in access requests from unknown browser types or known outdated versions could indicate an automated attack of some kind, and the enterprise could take steps to limit requests from these identified clients. The enterprise should also be aware of any statutes or regulations regarding what information can be collected and recorded about the requesting users and assets.

5 Threats Associated with Zero Trust Architecture

No enterprise can eliminate cybersecurity risk. When complemented with existing cybersecurity policies and guidance, identity and access management, continuous monitoring, and general cyber hygiene, a properly implemented and maintained ZTA can reduce overall risk and protect against common threats. However, some threats have unique features when implementing a ZTA.

5.1 Subversion of ZTA Decision Process

In ZTA, the policy engine and policy administrator are the key components of the entire enterprise. No communication 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 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 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(s) or PE/PA (i.e., DoS attack or route hijack), it can adversely impact enterprise operations. Enterprises can mitigate this threat by having the policy enforcement reside in a properly secured cloud environment or be replicated in several locations following guidance on cyber resiliency [SP 800-160v2].

This mitigates the risk but does not eliminate it. Botnets such as Mirai produce massive DoS attacks against key internet service providers and disrupt service to millions of internet users.[5] 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 subjects is affected. This is also possible in legacy remote-access VPNs and is not unique to ZTA.

A hosting provider may also accidentally take a cloud-based PE or PA offline. Cloud services have experienced disruptions in the past, both infrastructure as a service (IaaS)[6] and SaaS.[7] An operational error could prevent an entire enterprise from functioning if the policy engine or policy administrator component becomes inaccessible from the network.

---

5 https://blog.cloudflare.com/inside-mirai-the-infamous-iot-botnet-a-retrospective-analysis/
6 https://aws.amazon.com/message/41926/
7 https://www.nzherald.co.nz/business/news/article.cfm?c_id=3&objectid=12286870

There is also the risk that enterprise resources may not be reachable from the PA, so even if access is granted to a subject, the PA cannot configure the communication path from the network. This could happen due to a DDoS attack or simply due to unexpected heavy usage. This is similar to any other network disruption in that some or all enterprise subjects cannot access a particular resource due to that resource not being available for some reason.

5.3 Stolen Credentials/Insider Threat

Properly implemented ZT, information security and resiliency policies, and best practices reduce the risk of an attacker gaining broad access via stolen credentials or insider attack. The ZT principle of no implicit trust based on network location means attackers need to compromise an existing account or device to gain a foothold in an enterprise. A properly developed and implemented ZTA should prevent a compromised account or asset from accessing resources outside its normal purview or access patterns. This means that accounts with access policies around resources that an attacker is interested in would be the primary targets for attackers.

Attackers may use phishing, social engineering, or a combination of attacks to obtain credentials of valuable accounts. “Valuable” may mean different things based on the attacker’s motivation. For instance, enterprise administrator accounts may be valuable, but attackers interested in financial gain may consider accounts that have access to financial or payment resources of equal value. Implementation of MFA for access requests may reduce the risk of information loss from a compromised account. However, 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 asset of a valid human resources employee may still be able to access an employee database.

ZTA reduces risk and prevents any compromised accounts or assets from moving laterally throughout the network. If the compromised credentials are not authorized to access a particular resource, they will continue to be denied access to that resource. In addition, a contextual trust algorithm (see Section 3.3.1) is more likely to detect and respond quickly to this attack than when occurring in a legacy, perimeter-based 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 3.4.1, 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 (possibly the majority) of the traffic on the enterprise network may be opaque to layer 3 network analysis tools. This traffic may originate from nonenterprise-owned assets (e.g., contracted services that use the enterprise infrastructure to access the internet) or applications/services that are resistant to passive monitoring. The enterprise that cannot perform deep packet inspection or examine the encrypted traffic and must use other methods to assess 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 (e.g., source and destination addresses, etc.) about the encrypted traffic and use that to detect an active attacker or possible malware communicating on the network. 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.

5.5 Storage of System and Network Information

A related threat to enterprise monitoring and analysis of network traffic is the analysis component itself. If monitor scans, network traffic, and metadata are being stored for building contextual policies, forensics, or later 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 this information, they may be able to gain insight into the enterprise architecture and identify assets for further reconnaissance and attack.

Another source of reconnaissance information for an attacker in a ZT enterprise 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).

As for 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 be accessible only from designated or dedicated administrator accounts.

5.6 Reliance on Proprietary Data Formats or Solutions

ZTA relies on several different data sources to make access decisions, including information about the requesting subject, asset used, enterprise and external intelligence, and threat analysis. Often, the assets 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 assets) 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 because 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 considering factors such as vendor security controls, enterprise switching costs, and supply chain risk management in addition to more typical factors such as performance, stability, etc.

5.7 Use of Non-person Entities (NPE) in ZTA Administration

Artificial intelligence 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), sometimes in lieu of a human administrator. How these components authenticate themselves in an enterprise implementing a

ZTA is an open issue. It is assumed that most automated technology systems will use some means to authenticate when using an API to resource components.

The biggest risk when using automated technology for configuration and policy enforcement is the possibility of false positives (innocuous actions mistaken for attacks) and false negatives (attacks mistaken for normal activity) impacting the security posture of the enterprise. This can be reduced with regular retuning analysis to correct mistaken decisions and improve the decision process.

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

6 Zero Trust Architecture and Possible Interactions with Existing Federal Guidance

Several existing federal policies and guidance intersect with the planning, deployment, and operation of a ZTA. These policies do not prohibit an enterprise from moving to a more zero trust-oriented architecture but can influence development of a zero trust strategy for an agency. When complemented with existing cybersecurity policies and guidance, 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 allow access only via a connected terminal, but this is disproportionately restrictive in the majority of cases and could inhibit work from being accomplished. 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 and evaluated, and either accepted or mitigated. To assist in this, the NIST Risk Management Framework (RMF) was developed [SP800-37].

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 reliance on network perimeter defenses. The overall process described in the RMF will not change in a ZTA.

6.2 Zero Trust and NIST Privacy Framework

Protecting the privacy of users and private information (e.g., personally identifiable information) is a prime concern for organizations. Privacy and data protections are included in compliance programs such as FISMA and the Heath Insurance Portability and Accountability Act (HIPAA). In response, NIST produced a Privacy Framework for use by organizations [NISTPRIV]. 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. This includes personal information used by the enterprise to support ZTA operations and any biometric attributes used in access request evaluations.

Part of the core requirements for ZT is that an enterprise should inspect and log traffic (or at least log and inspect metadata when dealing with traffic that cannot be decrypted by monitoring systems) in its environment. Some of this traffic may contain private information or have associated privacy risks. Organizations will need to identify any possible risks associated with intercepting, scanning, and logging network traffic [NISTIR 8062]. This may include actions such as informing users, obtaining consent (via a login page, banner, or similar), and educating enterprise users. The NIST Privacy Framework [NISTPRIV] could help in developing a formal process to identify and mitigate any privacy-related risks to an enterprise developing a zero trust architecture.

6.3 ZTA and Federal Identity, Credential, and Access Management Architecture

Subject 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 subjects and resources. Strong subject provision and authentication policies need to be in place before moving to a more zero trust–aligned deployment. Enterprises need a clear set of subject attributes and policies that can be used by a PE to evaluate access requests.

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 the agency's ICAM policy.

6.4 ZTA and Trusted Internet Connections 3.0

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 ZT principles.

TIC 3.0 has been updated to accommodate cloud services and mobile devices [M-19-26]. In TIC 3.0, it is recognized that the definition of "trust" may vary across specific computing contexts and that agencies have different risk tolerances for defining trust zones. In addition, TIC 3.0 has an updated TIC Security Capability Handbook, which defines two types of security capabilities: (1) Universal Security Capabilities that apply at the enterprise level, and (2) PEP Security Capabilities that are network-level capabilities to be applied to multiple policy enforcement points (PEPs), as defined in TIC use cases. The PEP Security Capabilities may be applied at any appropriate PEP located along a given data flow instead of at a single PEP at the agency perimeter. Many of these TIC 3.0 security capabilities directly support ZTA (e.g., encrypted traffic, strong authentication, microsegmentation, 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 (operationally known as 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. EINSTEIN 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 for DHS situational awareness is based on a perimeter network defense in the Federal Government, while ZTA moves protections closer to the assets, data and all other resources. The NCPS program is evolving to ensure that situational awareness is preserved through utilization of security information about cloud-based traffic, helping to set the foundation for expanded situational awareness telemetry from ZTA systems. NCPS intrusion prevention functions would also require evolution to be able to inform policy enforcement at both the current NCPS locations as well as ZTA systems. As ZTA is adopted across the Federal Government, the NCPS implementation would need to continually 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 DHS Continuous Diagnostics and Mitigations (CDM) Program

The DHS CDM program is an effort to improve federal agency information technology (IT). Vital to that posture is an agency's insight into the assets, configuration, and subjects within itself. To protect a system, agencies need to set up processes to discover and understand the basic components and actors in 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?** What users are part of the organization or are external and allowed to access enterprise resources? These include NPEs that may be performing autonomous actions.
- **What is happening on the network?** An enterprise needs 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, in transit, and in use.

Having a strong CDM program implementation is key to the success of ZTA. For example, to move to ZTA, an enterprise must have a system to discover and record physical and virtual assets to create a usable inventory. The DHS CDM program has initiated several efforts to build the capabilities needed within federal agencies to move to a ZTA. For example, the DHS

Hardware Asset Management (HWAM) [HWAM] program is an effort to help agencies identify devices on their network infrastructure to deploy a secure configuration. This is similar to the first steps in developing a road map to ZTA. Agencies must have visibility into the assets active on the network (or those accessing resources remotely) to categorize, configure, and monitor the network's activity.

6.7 ZTA, Cloud Smart, and the Federal Data Strategy

The Cloud Smart[8] strategy, updated Data Center Optimization Initiative [M-19-19] policy, and Federal Data Strategy[9] all influence some requirements for agencies planning a ZTA. 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 to determining what business processes and resources would benefit from implementing a ZTA. Data resources and applications and services that are primarily cloud-based or primarily used by remote workers are good candidates for a ZTA approach (see Section 7.3.3) because the subjects and resources are located outside of the enterprise network perimeter and are likely to see the most benefit in use, scalability, and security.

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.

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

7 Migrating to a Zero Trust Architecture

Implementing a ZTA is a journey rather than a wholesale replacement of infrastructure or processes. An organization 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/perimeter-based mode for an indefinite period while continuing to invest in ongoing IT modernization initiatives. Having an IT modernization plan that includes moving to an architecture based on ZT principles may help an enterprise form roadmaps for small scale workflow migrations.

How an enterprise migrates to a strategy depends on its current cybersecurity posture and operations. An enterprise should reach a baseline of competence before it becomes possible to deploy a significant ZT-focused environment [ACT-IAC]. This baseline includes having assets, subjects, business processes, traffic flows and dependency mappings identified and cataloged for the enterprise. The enterprise needs this information before it can develop a list of candidate business processes and the subjects/assets involved in this process.

7.1 Pure Zero Trust Architecture

In a greenfield approach, it would be possible to build a zero trust architecture from the ground up. Assuming the enterprise knows the applications/services and workflows that it wants to use for its operations, it can produce an architecture based on zero trust 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 and organizational exercise in building the infrastructure and configuring the components. This may include additional organizational changes depending on how the enterprise is currently set up and operating.

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, service, or database. The agency could design the newly needed infrastructure around ZT principles and secure system engineering [SP8900-160v1], such as evaluating subjects' trust before granting access and establishing micro-perimeters around new resources. 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 Perimeter-Based Architecture

It is unlikely that any significant enterprise can migrate to zero trust in a single technology refresh cycle. There may be an indefinite period when ZTA workflows coexist with non-ZTA workflows in an enterprise. 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) are flexible enough to operate in a ZTA and perimeter-based hybrid security architecture. Enterprise architects may also want to restrict ZTA candidate solutions to those that can interface with existing components.

Migrating an existing workflow to a ZTA will likely require (at least) a partial redesign. Enterprises may take this opportunity to adopt secure system engineering [SP800-160v1] practices if they have not already done so for workflows.

7.3 Steps to Introducing ZTA to a Perimeter-Based Architected Network

Migrating to ZTA requires an organization to have detailed knowledge of its assets (physical and virtual), subjects (including user privileges), 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. This is especially an issue if there are unknown "shadow IT" deployments operating within an organization.

Before undertaking an effort to bring ZTA to an enterprise, there should be a survey of assets, subjects, data flows, and workflows. This awareness forms the foundational state that must be reached before a ZTA deployment is possible. An enterprise cannot determine what new processes or systems need to be in place if there is no knowledge of the current state of operations. These surveys can be conducted in parallel, but both are tied to examination of the business processes of the organization. These steps can be mapped to the steps in the RMF [SP800-37] as any adoption of a ZTA is a process to reduce risk to an agency's business functions. The pathway to implementing a ZTA can be visualized in Figure 12.

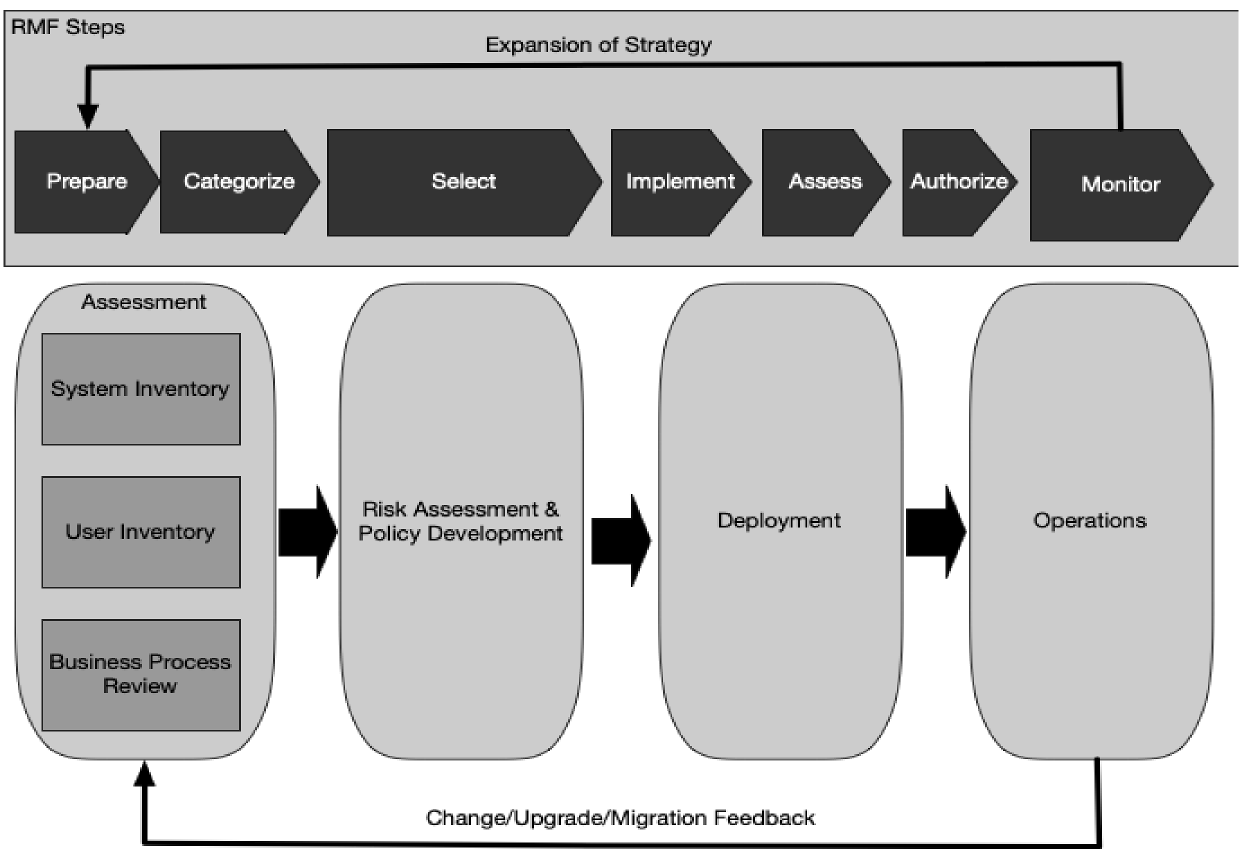

Figure 12: ZTA Deployment Cycle

After the initial inventory is created, there is a regular cycle of maintenance and updating. This updating may either change business processes or not have any impact, but an 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

For a zero trust enterprise to operate, the PE must have knowledge of enterprise subjects. Subjects could encompass both human and possible NPEs, such as service accounts that interact with resources.

Users with special privileges, such as developers or system administrators, require additional scrutiny when being assigned attributes or roles. In many legacy security architectures, these accounts may have blanket permission to access all enterprise resources. ZTA should allow developers and administrators to have sufficient flexibility to satisfy their business requirements while using logs and audit actions to identify access behavior patterns. ZTA deployments may require administrators to satisfy a more stringent confidence level 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 devices. ZTA also requires the ability to identify and monitor nonenterprise-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) and digital artifacts (e.g., user accounts, applications, digital certificates). It may not be possible to conduct a complete census on all enterprise-owned assets, so an enterprise should consider building the capability to quickly identify, categorize, and assess newly discovered assets that are on enterprise-owned infrastructure.

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 an asset is part of the process of evaluating access requests (see Section 2.1). This means that the enterprise must be able to configure, survey, and update enterprise assets, such as virtual assets and containers. This also includes both its physical (as best estimated) and network location. This information should inform the PE when making resource access decisions.

Nonenterprise-owned assets and enterprise-owned “shadow IT” should also be cataloged as well as 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 only used for access decisions (as collaborator and BYOD assets may need to contact PEPs) but also for monitoring and forensics logging by the enterprise. Shadow IT presents a special problem in that these resources are enterprise-owned but not managed like other resources. Certain ZTA approaches (mainly network-based) may even cause shadow IT components to become unusable as they may not be known and included in network access policies.

Many federal agencies have already begun identifying enterprise assets. Agencies that have established CDM program capabilities, such as HWAM [HWAM] and Software Asset Management (SWAM) [SWAM], have a rich set of data to draw from when enacting a ZTA. 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. These programs must be designed to be expandable and adaptable to changes in the enterprise, not only when migrating to ZTA but also when accounting for new assets, 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, data flows, and their relation in the 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 and would likely see improvements to availability and security. Rather than project the enterprise perimeter into the cloud or bring clients into the enterprise network via a 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. Planners should also consider potential tradeoffs in performance, user experience, and possible increased workflow fragility that may occur when implementing ZTA for a given business process.

7.3.4 Formulating Policies for the ZTA Candidate

The process of identifying a candidate service or business workflow depends on several factors: the importance of the process to the organization, the group of subjects 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, identify all upstream resources (e.g., ID management systems, databases, micro-services), downstream resources (e.g., logging, security monitoring), and entities (e.g., subjects, service accounts) that are used or affected by the workflow. This may influence the candidate choice as a first migration to ZTA. An application/service used by an identified subset of enterprise subjects (e.g., a purchasing system) may be preferred over one that is vital to the entire subject 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 confidence level weights (if using a score-based TA) for the resources used in the candidate business process (see Section 3.3.1). Administrators may need to adjust these criteria or values during the tuning phase. These adjustments are necessary to ensure that policies are effective but do not hinder 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 some use cases than others. These are some factors to consider:

- **Does the solution require that components be installed on the client asset?** This may limit business processes where nonenterprise-owned assets 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 that 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.
- **Does the solution provide a means to log interactions for analysis?** A key component of ZT is the collection and use of data related to the process flow that feeds back into the PE when making access decisions.
- **Does the solution provide broad support for different applications, services, and protocols?** Some solutions may support a broad range of protocols (web, secure shell [SSH], etc.) and transports (IPv4 and IPv6), while others may only work with a narrow focus such as web or email.
- **Does the solution require changes to subject behavior?** Some solutions may require additional steps to perform a given workflow. This may change how enterprise subjects perform the workflow.

One solution is to model an existing business process as a pilot program rather than just a replacement. This pilot program could be made general to apply to several business processes or be made specific to one use case. The pilot can be used as a "proving ground" for ZTA before transitioning subjects to the ZTA deployment and away from the legacy 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 by using the selected components but may wish to operate in an observation and monitoring mode 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 been 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. This also allows the enterprise to gain an understanding of baseline asset and resource access requests, behavior, and communication patterns. Reporting-only means that access should be granted for most requests, and logs and traces of connections should be compared with the initial developed policy. Basic policies such as denying requests that fail MFA or appear from known, attacker controlled or subverted IP addresses should be enforced and logged, but after initial deployment, access polices should be more lenient to collect data from actual interactions of the ZT workflow. Once the baseline activity patterns for the workflow has been established, anomalous behavior can be more easily identify. 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

When enough confidence is gained and the workflow policy set is refined, the enterprise enters the steady operational phase. The network and assets are still monitored, and traffic is logged (see Section 2.1), but responses and policy modifications are done at a lower tempo as they should not be severe. The subjects and stakeholders of the resources and processes involved should also provide feedback to improve operations. 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), and 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

[BCORE] Department of Defense CIO (2007). Department of Defense Global Information Grid Architecture Vision Version 1.0 June 2007. Available at http://www.acqnotes.com/Attachments/DoD%20GIG%20Architectural%20Vision,%20June%2007.pdf

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

[FIPS199] National Institute of Standards and Technology (2004) Standards for Security Categorization of Federal Information and Information Systems. (U.S. Department of Commerce, Washington, DC), Federal Information Processing Standards Publication (FIPS) 199. https://doi.org/10.6028/NIST.FIPS.199

[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

[IBNVN] Cohen R, Barabash K, Rochwerger B, Schour L, Crisan D, Birke R, Minkenberg C, Gusat M, Recio R, Jain V (2013) An Intent-based Approach for Network Virtualization. *2013 IFIP/IEEE International Symposium on Integrated Network Management (IM 2013)*. (IEEE, Ghent, Belgium), pp 42-50. Available at https://ieeexplore.ieee.org/document/6572968

[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

[NISTIR 8062] Brooks SW, Garcia ME, Lefkovitz NB, Lightman S, Nadeau EM (2017) An Introduction to Privacy Engineering and Risk Management in Federal Systems. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Interagency or Internal Report (IR) 8062. https://doi.org/10.6028/NIST.IR.8062

[NISTPRIV] National Institute of Standards and Technology (2020) Privacy Framework: A Tool for Improving Privacy Through Enterprise Risk Management, Version 1.0. (National Institute of Standards and Technology, Gaithersburg, MD). https://doi.org/10.6028/NIST.CSWP.01162020

[SDNBOOK] Nadeau T, Gray K (2013) *SDN: Software Defined Networks: An Authoritative Review of Network Programmability Technologies*. (O'Reilly) 1st Ed.

[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 March 2, 2020. 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 March 2, 2020. https://doi.org/10.6028/NIST.SP.800-63A

[SP800-160v1] Ross R, McEvilley M, Oren JC (2016) Systems Security Engineering: Considerations for a Multidisciplinary Approach in the Engineering of Trustworthy Secure Systems. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Special Publication (SP) 800-160, Vol. 1, Includes updates as of March 21, 2018. https://doi.org/10.6028/NIST.SP.800-160v1

[SP800-160v2] Ross R, Pillitteri V, Graubart R, Bodeau D, McQuaid R (2019) Developing Cyber Resilient Systems: A Systems Security Engineering Approach. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Special Publication (SP) 800-160, Vol. 2. https://doi.org/10.6028/NIST.SP.800-160v2

[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 August 2, 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

API	Application Programming Interface
BYOD	Bring Your Own Device
CDM	Continuous Diagnostics and Mitigation
DHS	Department of Homeland Security
DoS	Denial of Service
G2B	Government to Business (private industry)
G2G	Government to Government
NIST	National Institute of Standards and Technology
NPE	Non-Person Entity
PA	Policy Administrator
PDP	Policy Decision Point
PE	Policy Engine
PEP	Policy Enforcement Point
PKI	Public Key Infrastructure
RMF	NIST Risk Management Framework
SDN	Software Defined Network
SDP	Software Defined Perimeter
SIEM	Security Information and Event Monitoring
TIC	Trusted Internet Connections
VPN	Virtual Private Network
ZT	Zero Trust
ZTA	Zero Trust Architecture

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 research conducted in the development of 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 environment, 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 environments.

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 zero trust based enterprise infrastructure now or maintaining an existing ZTA implementation. These gaps can be categorized into immediate deployment (immediate or short term), systemic gaps that affect maintenance or operations (short or midterm), and missing knowledge (areas for future research). They are summarized in Table B-1.

Table B-1: Summary of Identified Deployment Gaps

Category	Example Questions	Identified Gaps
Immediate deployment	• How should procurement requirements be written? • How does a ZTA plan work with TIC, FISMA, and other requirements?	• Lack of a common framework and vocabulary for ZTA • Perception that ZTA conflicts with existing policy
Systemic	• How can vendor lock-in be prevented? • How do different ZTA environments interact?	• Too much reliance on vendor APIs
Areas needing more research	• How will threats evolve in the face of ZTA? • How will business processes change in the face of ZTA?	• What a successful compromise looks like in an enterprise with a ZTA • Documented end user experience in an enterprise with a ZTA

B.2 Gaps that Prevent an Immediate Move to ZTA

These are the issues that are slowing adoption of a ZTA at present. These were classified as immediate issues, and no thought of future maintenance or migration was considered for this category. A forward-thinking enterprise may also consider the maintenance category to be of immediate concern in preventing the initial deployment of ZTA components, but these issues 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 difficult for organizations (e.g., federal agencies) to develop coherent requirements and policies for designing zero trust enterprise infrastructure and procuring components.

The driver for Sections 2.1 and 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 in federal agencies, but they currently serve as a base for a common conceptual framework.

B.2.2 Perception that ZTA Conflicts 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 instead 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 (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 and continued operation/maturity. These gaps could slow the adoption of ZTA in agencies or result in 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 offers 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 thereby risk vendor lock-in. This leads to interoperability within components not only at the time of purchase but also over time.

The spectrum of components within the wider enterprise is vast, with many products focusing on a single niche within zero trust and relying on other products to provide either data or some service to another component (e.g., integration of 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 ensure early notification of modifications within APIs, which may affect compatibility between products. This adds an additional burden on vendors and consumers: vendors need to expend resources to change their products, and consumers need to apply updates to multiple products when one vendor makes a change to its proprietary API. Additionally, vendors are required to implement and maintain wrappers for each partner component to allow 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 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 multiyear road map for moving into ZTA because 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 no single solution to developing a ZTA, there is no single set of tools or services for a zero trust enterprise. Thus, it is impossible to have a single protocol or framework that enables an enterprise to move to a ZTA. Currently, there is 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. 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 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 for its enterprise. These are gray areas in knowledge about operational ZTA environments, and most arise from 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 for an enterprise will improve the enterprise's cybersecurity posture over traditional 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 asset 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 will be expanded to target MFA (e.g., phishing, social engineering). Another possibility is that in a hybrid ZTA/perimeter-based 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 is gained, the effectiveness of ZTA in shrinking the attack surface of resources 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 that is using a ZTA. This is mainly due to the lack of large ZTA use cases available for analysis. There have, however, been studies on how users react to MFA and other security operations that are part of a ZTA enterprise, and 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., applications 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 would need to consider. As previously noted, there is no single provider of a full zero trust solution at this time. As a result, enterprises will purchase several different services and products, which 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 through either 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. Research is needed on discovering the 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. A ZTA makes many COOP factors easier as remote workers may have the same access to resources that they had on-premises. However, policies like MFA may also have a negative impact if users are not properly trained or 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 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.

[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

sustainability

MDPI

Review

Security of Zero Trust Networks in Cloud Computing: A Comparative Review

Sirshak Sarkar [1], Gaurav Choudhary [2], Shishir Kumar Shandilya [1], Azath Hussain [1] and Hwankuk Kim [3,*]

[1] School of Computing Science and Engineering, VIT Bhopal University, Bhopal 466114, Madhya Pradesh, India
[2] DTU Compute, Department of Applied Mathematics and Computer Science, Technical University of Denmark, 2800 Kongens Lyngby, Denmark
[3] Department of Information Security Engineering, Sangmyung University, Cheonan 31066, Korea
* Correspondence: rinyfeel@smu.ac.kr

Abstract: Recently, networks have shifted from traditional in-house servers to third-party-managed cloud platforms due to its cost-effectiveness and increased accessibility toward its management. However, the network remains reactive, with less accountability and oversight of its overall security. Several emerging technologies have restructured our approach to the security of cloud networks; one such approach is the zero-trust network architecture (ZTNA), where no entity is implicitly trusted in the network, regardless of its origin or scope of access. The network rewards trusted behaviour and proactively predicts threats based on its users' behaviour. The zero-trust network architecture is still at a nascent stage, and there are many frameworks and models to follow. The primary focus of this survey is to compare the novel requirement-specific features used by state-of-the-art research models for zero-trust cloud networks. In this manner, the features are categorized across nine parameters into three main types: zero-trust-based cloud network models, frameworks and proofs-of-concept. ZTNA, when wholly realized, enables network administrators to tackle critical issues such as how to inhibit internal and external cyber threats, enhance the visibility of the network, automate the calculation of trust for network entities and orchestrate security for users. The paper further focuses on domain-specific issues plaguing modern cloud computing networks, which leverage choosing and implementing features necessary for future networks and incorporate intelligent security orchestration, automation and response. The paper also discusses challenges associated with cloud platforms and requirements for migrating to zero-trust architecture. Finally, possible future research directions are discussed, wherein new technologies can be incorporated into the ZTA to build robust trust-based enterprise networks deployed in the cloud.

Keywords: zero trust; cloud security; zero-trust cloud networks; cloud computing; zero-trust models

Citation: Sarkar, S.; Choudhary, G.; Shandilya, S.K.; Hussain, A.; Kim, H. Security of Zero Trust Networks in Cloud Computing: A Comparative Review. *Sustainability* **2022**, *14*, 11213. https://doi.org/10.3390/su141811213

Academic Editor: Tin-Chih Toly Chen

Received: 8 August 2022
Accepted: 30 August 2022
Published: 7 September 2022

1. Introduction

In the modern era of distributed computing, there have been many advances in the adoption and implementation of networked security systems for cloud servers. Since 2010, the global cloud services industry has had a year-on-year increase which sums up-to a USD 370 billion valuation in 2020, posting a 380 percent growth in a decade. As a consequence of such breakneck adoption in 2022, over 60 percent of all corporate data is stored in the cloud. This increased by almost 30 percent in 2015 and has seen continual growth as firms rapidly migrate their resources and business applications into cloud environments with the hope of improving security, reliability and ease of business [1,2].

This dramatic expansion has caused issues with corporate and government networks hosted on cloud platforms, each deployed with a highly proprietary set of security mechanisms such as service-level agreements (SLA), identity management and access controls, intrusion detection systems (IDS) and application service management. It is quite evident that cloud service customers (CSC) deploy these services on their networks with their own specifications, based on their past operational experiences and convenience. Some of the leading cloud network environments are Amazon Web Services, Microsoft Azure, IBM Cloud, VMware and Google Cloud [3,4].

However, these complex security systems have not stopped cloud platforms and networks from being exploited by ransomware groups, botnets and advanced persistent threats (APT), the main culprit being poor security practices and configuration and internal vulnerabilities [5,6]. Cloud networks can also be exploited by third-party applications which introduce unforeseen bugs and even zero-day vulnerabilities, giving attackers access to sensitive customer data. Additionally, unless organisations verify sources, third-party applications can come from anyone inside the network, including an APT. According to a study conducted by Palo Alto Network's Unit 42, approximately 96% of application containers in cloud infrastructure have known exploits and vulnerabilities [7,8].

The current threat landscape naturally leads us to believe that a trust-based authorisation mechanism is needed in a cloud network environment which monitors and assists different nodes of that network [9,10]. This network authority must also have the privilege of authorising users' access to services and distributing responsibilities based on the authenticity of their identity. The technology we have just described is called a zero-trust network model. The zero-trust network model is such that no entity inside such a network is explicitly trusted. For each individual action by that entity where it must make use of some mission critical data or service, the network management authority must first give clearance to that action. This can be set up with existing technologies such as IDS, real-time resource management, segmentation of resources and behaviour tracking, which provides visibility, granular control and access of endpoint devices to network security teams [11–13].

Unfortunately, there are many issues which organisations must handle before an environment is fully functional, such as issues with legacy hardware, lack of applications to manage endpoint devices and training employees to use complex virtualization software, etc. Moreover, the widespread use of cloud computing platforms has caused the boundaries of the network to collapse. When an institution deploys a hybrid cloud platform and stores business-critical data on-site and on the platform, it increases their threat surface area [14,15]. Thankfully, there are continual efforts by many government institutions and private companies, including cloud service providers (CSP), to streamline varying rules and guidelines into adaptable frameworks and models.

For example, The National Institute of Standards and Technology (NIST) released a special publication on Zero-Trust Architecture in 2020 [16]. On 26 January 2022, The Office of Management and Budget (OMB) released a federal strategy to move the U.S. Government toward a "zero trust" approach to cybersecurity; this was in line with Executive Order (EO) 14028, "Improving the Nation's Cybersecurity", which pushes U.S. agencies to adopt zero-trust cybersecurity principles and adjust their network architectures accordingly [17]. The memorandum by the OMB requires agencies to achieve specific zero-trust security goals by the end of Fiscal Year (FY) 2024. Many private firms have also developed and offer state-of-the-art zero-trust network security solutions such as VMware's NSX Advanced Threat Detection and Carbon Black Cloud, Google's BeyondCorp, Palo Alto Networks' Next Generation Firewall, Citrix's Workspace, Microsoft's Azure and 365 Security.

However, the adoption of such frameworks and technologies have been overlooked by smaller and medium-sized firms and institutions who do not have sufficient resources, time or the inclination to implement such a framework to their cloud ecosystem. It is thus important to engage in a survey of the current implementations of zero-trust-based cloud network models. The survey will compare the different methods and approaches to validate identity and authorise the use of critical services in trust-based cloud networks. Thus, the objective of our paper is to help firms, institutions and governments achieve Zero-Trust Maturity for their cloud networks.

1.1. Contributions of This Paper

1. This paper discusses the current trends in the adoption of cloud infrastructure to host networks, how cost-effectiveness has prompted organisations to switch to cloud infrastructure and the issues faced when using such services;
2. The paper introduces the origin of trust evaluation as a precursor of zero trust. It links the natural progression of different projects and precursor technologies for distributed networks into the zero-trust architecture model. It also lists the core motivations behind the model, why it was needed and details of the workings of mature state-of-the-art research categorically;
3. The paper discusses the many challenges to the security of cloud networks, such as the interoperability of different cloud architectures, flaws in perimeter-based network design, vulnerabilities in virtualisation and poor practices prevalent in the cloud industry regarding the access to and oversight of data;
4. The paper explains the concept of zero-trust architecture, enumerates the properties a network must have to be based on ZTA and compares it with traditional network security architecture, as well as the required attributes of the algorithm for calculating trust.
5. The paper also maps out the specific changes that a pre-existing cloud network can undertake to migrate to ZTA and briefly talks about some problems that might arise operating this complex network architecture;
6. The paper then provides a type-focused comparison of the related state-of-the-art-works, emphasising their features and categories such as ZTA-based network models, frameworks and proof-of-concept technologies;
7. Conclusively, the paper outlines ongoing research and future directions for including emerging as well as proven technologies to enhance the ZTA as a whole with some use cases

1.2. Outline of the Survey

The paper is structured as follows: Section 1 introduces the current scenario of cloud platform adoption and issues that cloud service providers face in relation to the challenges of traditional networks and states the recent advancements undertaken by academia and industry to remediate them. Section 2 elucidates the development of the ZTA, followed by a detailed discussion of related works on ZTA models, frameworks, surveys and road maps according to their types. Section 3 includes the taxonomy of challenges faced by cloud networks. Section 4 highlights how the migration of traditional networks can be executed to mitigate detrimental effects and the long-term consequences of migrating to a ZTA. Section 5 compares related works based on various features, and Section 6 concludes the paper. Section 7 consists of recommendations and future directions.

2. Comparison with Existing Survey Articles

There are many studies which survey the architectural and technical features of ZTA. Table 1 includes a summary of existing surveys about Zero Trust. Buck et al. [18] provided a survey which analysed papers written on ZTA using a search model which distinguished academic literature from grey literature, the latter being from non-academic, commercial or private sources. Alevizos et al. [19] covered the fusion of blockchain's immutability to use intrusion prevention and detection at network endpoints. He et al. [20] provide a study on the advantages and disadvantages of access control models and authentication protocols and compare popular evaluation methods for trust. This work about access control methods and authenticating protocols in a network is also the focus of Syed et al. [21]. They discuss the challenges to such an architecture and expand its scope towards software-defined perimeters and micro-segmentation. Pittman et al. [22] survey a novel idea, applying zero-trust tenets and principles to data objects instead of pathways that allow users to access data. They conclusively state that the calculation of trust in a dynamic system such as a network is both a classification problem and a regression problem.

Most surveys primarily focused on the development of the architecture and management of ZTA or specific derived topics such as micro-segmentation, software-defined perimeter and intrusion prevention systems. This paper provides a comparison of the properties of the network or distinct features which are commonly used. As zero-trust architecture is not a monolithic one: it employs many proven and emerging technologies; comparing these is essential to sorting out the best-fit features. Of the papers surveyed, many authors also state that ZTNs have not been able to replace existing approaches to network security.

Table 1. A comparison of existing surveys and reviews (discussed: ✓; never mentioned: x; partially mentioned: -). P1: Categorisation of types of works surveyed, P2: Comparison of Models based on novel features, P3: Comparison of works based on independent parameters, P4: Details about challenges to cloud networks, P5: Discussion on Zero-Trust Lifecycle (maturity model), P6: Specifies possible domains of future research.

Author(s)	Primary Contributions	P1	P2	P3	P4	P5	P6
Buck et al. [18]	Consolidation of works based on Zero Trust, analysis of knowledge gaps in industry and academia	✓	x	x	-	x	✓
Alevizos et al. [19]	Analysis of ZTA-based models, blockchain-based intrusion detection and prevention to augment end-point security	x	✓	x	-	x	✓
He et al. [20]	Analysis of core technologies mainly relied on in Zero Trust and comparison of pros and cons	✓	✓	✓	x	x	-
Syed et al. [21]	Discussion about access control and authentication in different scenarios and impact of ZT implementation	✓	-	✓	-	x	✓
Pittman et al. [22]	Application of Zero Trust tenets and principles to data objects instead of data access pathways	✓	x	x	-	x	x
This Paper	Categorization and comparison of novel features used in Zero Trust Models for Cloud Networks	✓	✓	✓	✓	✓	✓

3. Related Work

The idea of a continuous 'trust evaluation'-based computer network has been around for a long time, having been proposed and even designed by the U.S. Department of Defence (DoD) in association with the Defence Information Systems Agency (DISA). It was named Black Core. Black Core was a communication network architecture in which user data traversing a global IP (Internet Protocol) network was encrypted end-to-end at the IP layer [23]. It was an experiment to re-focus network security across distributed servers from the model of perimeter security to request-based security.

But the very first instance of the term 'Zero Trust' being used was in 2010 by John Kindervag, from Forrester Research. It described the objective of the zero-trust model as 'to look at everything from a data-centric perspective, we can design networks from the inside out and make them more efficient, more elegant, simpler, and more cost-effective' [24]. Shortly after, further papers were published describing proposed methods of continuous interactions between users and the network authorisation mechanism [25].

The cloud industry and academia have started to formally develop many zero-trust cloud network models [16,18,26]. It must be understood that the Zero-Trust Model is a collection of different technologies and methods implemented in a system, by which the trust of an entity is determined. Thus, there are three main types of implementations, which we shall consolidate and compare according to their type. They are, namely, frameworks for the trust model, which can be used to develop and design use cases, practical zero-trust models and theoretical proof-of-concepts, which showcase individual cases of specific technologies that fall under the umbrella term of Zero Trust [27,28].

3.1. Models

Casimer et al. produced a model which used tokens included inside the first TCP (Transmission Control Protocol) packet to verify and validate user identity during packet authentication. This was used to show that their network model could prevent DDoS attacks, spoofing of identity and fingerprinting of the network by adversaries in different environments such as enterprise-class servers, cloud computing data centres and a campus-based network connecting different physical locations [23,29]. This approach of first packet authentication using tokens was further developed for geographically distributed higher education cloud networks [30].

Dayna et al. published a proposed zero-trust cloud data centre network. Their model used identity management along with automated threat response and packet-based authentication for establishing trust. The model then dynamically managed eight distinct network trust levels it had generated [31]. An incredibly unique approach to a lateral problem in the domain of privacy for location-based service users was put forward by Anwar et al. Location information such as history and overall regional proximity to certain 'business-specific' areas can give third-party information processors a huge insight. Their model distributes user location data into different servers according to a partitioning model based on multi-level policy. Third-party applications are granted access only to designated servers where the privacy of the user profile is also ensured, as these applications are not trustable. Zero-Trust identity management is used in a cloud environment to manage the access to systems containing sensitive data, using different trust levels [32].

Further use of Zero Trust can be seen in different domains. One such domain was a distributed volunteer cloud network, which, in practice, might be comparable to a blockchain volunteer network where nodes are awarded with higher trust. Abdullah et al. [33] proposed a client-server model to verify the trust of nodes participating as volunteer nodes in a zero-trust cloud network. The nodes are initially not trusted by the system. They proposed an adaptive behaviour-based system which assigns tasks to the most trusted nodes and manages their lifecycle. Nodes with low trust scores are added to a blocklist and given less essential or no tasks at all. This trust score is generated by their analysis of the entire lifecycle of the node, which includes its behaviour, efficiency and availability, among other factors [33].

3.2. Frameworks

Frameworks are also essential to building zero-trust networks in the cloud. They help cloud architects and network administrators design, setup and manage essential procedures for trust-based identity management. Many times, such frameworks outline stringent policy-based enforcement, such as in paper by Romans et al., where a policy enforcement framework called FURZE (Fuzzy Risk Framework for Zero Trust Networks) was created to address challenges in Zero-Trust Networks. The researchers outline specific language choices for the design of a risk-based access control framework, created using fuzzy logic and with an emphasis on continuity updates to the system. They also put forth some generic firewall policy language and rules which may help in creating specific firewall rules [34].

Other authors have leaned towards creating typologies and philosophies such as by Mehraj et al. The authors compiled a list of typologies and strategies related to zero-trust networks. One such strategy is the complete automation of trust calculation using a 'Trust Engine', which is an integrated system which controls the data, users and applications. The proposed system would then dynamically calculate the consolidated trust of a user, device or application by applying a trust score in a particular segment of the network.

The trust engine would work on a Zero-Trust triangle, much like the CIA Triad, the trio being application, user and device [35].

An enterprise-oriented Software-Defined Perimeter (SDP) Framework was proposed by Abdallah Moubayed et al., which adopted a zero-trust architecture by authenticating and verifying a host for every session using a client-gateway architecture. It could address lateral threats and internal pivoting attacks often found in such environments. Their performance analysis showed the potential as an alternative to VPNs for internal enterprise networks [36]. Ahmed and Petrova proposed a federated IAM framework using zero trust as a basis to stop CSP's from accessing virtual assets of their customers [37]. For zero-trust deployments in multi-cloud environments, a performance analysis was performed by Simone et al., which pointed to no negative effects [27].

3.3. Proof-of-Concepts

The much talked about eZTrust by Zirak Zaheer et al. provides a network-independent perimeter solution for micro-services. It focuses on contextual and granular control of workload identities by tracing them using the Berkeley Packet Filter. It also enables data centres to create and enforce access control policies based on the previously mentioned workload identity. The system can verify and trace authentic workload identities and tags to packets received [27]. A research project by Weever and Andreou highlighted the importance of securing data in-transit from one containerized application to another. It is during the transit that attacks can occur and data can be stolen. By securing the data, the authors were able to regulate the flow of traffic and find out the attack's origin [38].

4. Security Challenges to Cloud Computing Networks

This section discusses the internal security challenges that networks hosted on cloud platforms face. One such example is the inherent flaw of the model of the implicit trust of entities inside the network. Although this is not an exhaustive list of issues, the primary focus of this section is to highlight the core design flaws of modern cloud networks with traditional network security controls. Cyber threats such as malware, ransomware, data breaches and phishing attacks are influenced by external factors that have leeway to affect systems due to vulnerabilities arising from the issues discussed below. For example, vulnerabilities in virtualization constitute a significant aspect of the spread of ransomware from a host OS to its hypervisor and eventually to other host operating systems. Poor security controls and lack of oversight are pivotal problems that allow attackers to cause damage to the IT systems of CSCs and CSPs.

4.1. Architecture Interoperability Issues

There is no one type of conformal cloud computing architecture. Several different cloud computing models, types and services have emerged to serve evolving requirements of organisations and institutions. Currently, there are three types of cloud computing infrastructure:

(a) **Public Cloud Infrastructure** Cloud services where the hosting infrastructure and computing capabilities are available publicly. The CSP develops, operates and manages the service which is made available for public use, either free-of-cost or for a nominal fee. Google Cloud offers its services for free to Google users, such as Drive for hosting data, Classroom for educational management, Meet for video–audio calls, Gmail for email services and YouTube as a separate video platform ecosystem which is hosted on Google Cloud, etc. These types of cloud platforms are for general purpose use, with little or no option for customisable/modular features.

(b) **Private Cloud Infrastructure** Cloud services where the hosting infrastructure and computing capabilities are available to a limited user base. In this architecture, the resources are used by one firm or organisation. The servers which host the platform are physically located at an on-site data centre or they are hosted by a third-party CSP. The infrastructure is dedicated solely to one firm or a group of firms, and as such, the services and applications hosted on the cloud platform are not available to entities outside the organisation. Microsoft's Azure has many such services which are available to customer organisations on a subscription basis. These services are also highly customised and modular in accordance to the customer's request. Leading examples are Microsoft Azure, Oracle Cloud and VMware.

(c) **Hybrid Cloud Infrastructure** Cloud services are hosted both on-site at the customer location and on a private or public cloud platform. This approach is way more scalable and modular; however, it is not more secure. The hybrid cloud environment usually includes a plethora of application suites and network software with micro-segmentation, where networks are segmented according to user groups, departments or physical locations such as offices or outlets. Hybrid Cloud deployments are scalable as resources, and parts of the network can be opened for routine use in a phased manner. Innovative technologies can also be integrated much faster than private-only cloud networks as they can be deployed on a certain segment of the network for testing and evaluation.

According to a survey conducted in 2019, of 786 cloud professionals at large and small enterprises across a multitude of industries, such as software, financial services, telecommunications, education, government and healthcare, 22% of CSC's were utilizing public-only cloud services, 3% were using private and 68% were using hybrid models of cloud infrastructure. This is clearly an issue since each firm has a specific standard operating procedure when a network anomaly is detected. Cloud professionals and network specialists must be trained each time a new component is brought online or joins the firm, increasing costs and time consumed due to training.

4.2. Network Design

The fundamental model of cloud security still makes use of a reactive model, that of perimeter security where the connections and flow of data originating internally are trusted more than that originating externally, and mission-critical components of the network are cordoned off using firewalls. This model often includes a DMZ or demilitarized zone where untrusted requests for services present inside the 'protected' are accepted or denied as per the security policy. However, some non-essential network components, such as an email server or a publicly available FTP server, as commonly found in academic institutions, are exposed for convenience. This introduces insecurity into the network. This is most often how network penetration occurs, by gaining entry into a non-privileged part of the network and pivoting towards administrated areas. The Perimeter Based Security Model of Cloud Network is shown in Figure 1.

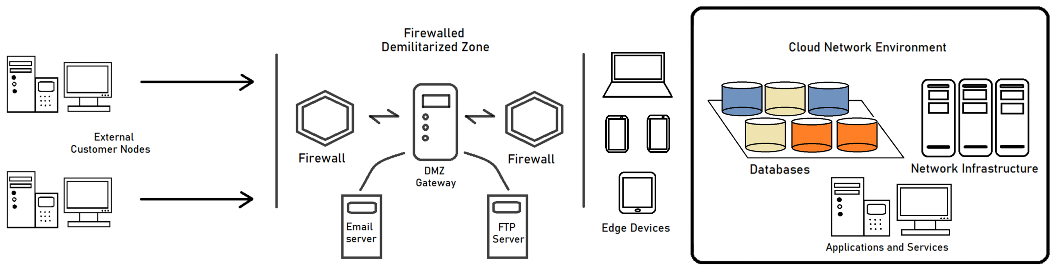

Figure 1. Perimeter Based Security Model of Cloud Network.

This is a reactive and static approach to networks where both physical and virtual services of the network are either secured or left insecure based on business requirements; however, this breaks down in modern cloud computing and ad-hoc endpoint device environments, where dynamic changes to the composition of the network and available resources make the DMZ irrelevant. Conventional networks verify the identity of users and the applications they use based on unique parameters such as MAC addresses, IP addresses, group policy and access privileges. However, these unique parameters can be spoofed and changed at will by attackers with moderate network knowledge, as noted by an NIST report in 2013.

4.3. Vulnerabilities in Virtualisation

The use of virtualisation software to emulate the hardware stack has allowed CSPs to create and offer complex cloud platforms, complete with operating systems, applications and code running natively inside virtual machines. This use of multiple operating systems and applications with the same type or types of hardware that CSPs use has enabled feature creep and thus introduced vulnerabilities associated with such software and hardware.

Virtual machines (VM) use a software system called a hypervisor. which is often called the Virtual Machine Monitor (VMM). The sole purpose of the VMM is to manage and assign resources such as virtualised OSs (Operating Systems) to users. The hypervisor is responsible for creating, stopping, and modifying the VMs. It is evident that such a software system which is responsible for the management of many VMs, will most definitely have vulnerabilities. According to Guodong Zhu et al., there are three main types of virtualisation errors which give rise to vulnerabilities, hardware logic errors, device state management errors and resource availability errors. The most common being the first hardware logic errors [39]. The authors state that the implementation of VMs works differently than their physical counterparts. Such slight erroneous differences cause technical issues which are worked around for operational continuity. It is at this time that the vulnerabilities are introduced due to oversights in the logic. Moreover, the hypervisor may have distinct levels of control for a server farm. Cloud customer often use a variety of services, such as the cloud infrastructure (IaaS), cloud developer platform (PaaS) or just the application software hosted by the CSP (SaaS).

A second issue being during migration or backup of VMs, device state errors can occur. Every time a VM is started, it does so by remembering its previous backup state in a specific configuration. This is essential for repeated use of the same VM by different users, who have different tasks. Bad handling of the VM either by CSCs or the hypervisor can lead to corruption of guest the OS (Operating System) and possibly the entire VM stack. Such errors can be minute and trivial to the guest OS during operation; however, on boot-up, a few corrupted CPU register values or logical errors of the OS kernel can cause catastrophic damage to the VM hardware stack, which further corrupts other guest operating systems.

The final issue being resource availability errors, it becomes relevant as the scale and complexity of the cloud providers available VM resource pool increases. Errors may be introduced into the VMM implementation when managing a large data centre with multiple types of hardware stacks, this combined with ad-hoc use of resources by VMs, and their guest OS can lead to lack of available resources.

As stated by Guodong Zhu et al., a virtualised OS cannot differentiate between virtual and physical hardware. The authors also state that physical isolation of such VMMs is essential to alleviate this issue. Other essential work has also been performed by Gábor Pék et al. where the different hardware virtualisation vulnerabilities are classified according to their source, attack strategies and adversary models, as well as structure of attack vectors [40].

4.4. Poor Security Practices and Insider Threats

There exists a huge gap in the quality and quantity of a cloud security workforce, especially that of Security Operation Centers (SOC) in-charge of cloud networks. The overall challenges that SOCs are struggling with are real-time visibility of the infrastructure they protect, compliance by CSCs and design of security policies that remain enforceable across architectures, on-site and off-site [41]. Another alarming trend for cloud networks is insider threats, an insider threat is a possible threat to the network by an individual with heightened privileges and access. Most often insider threats are accidental and not purposely done. This issue arises due to lack of transparency by CSPs. To a potential customer firm, the pipeline of procedures that a CSP employee can use to access and exploit customer data is not visible. CSPs do not give adequate information to their users, about the scope and scale of information available to its own employees. This is security through obscurity, which is flawed. An insider or a CSP employee can focus on a specific part of the cloud platform and may have any reasons to exploit the system ranging personal financial incentive to sociopathic behaviour. There are also no concrete methods to precisely predict such an attack, the attacker can plan in advance and even make use of the cloud platform to schedule the exploitation [42]. The main reason cloud networks attract such attention from adversaries internal and external is due to the varied nature of the data stored in these systems. A meagre 16 percent of firms using cloud platform to host their business networks say that current security tools and technologies are sufficient [41]. It is thus imperative that further research be conducted based on a human-centric approach, to effectively combat insider threats and poor security practices in cloud networks.

5. Concepts of Zero-Trust Architecture

The traditional security model is the perimeter (boundary) security model with the concept "Trust but verify". This concept has been operated that trusts internal users who have passed the security functions in the system or network but is wary of external attacks. However, the Zero-Trust model does not have a 'trust zone' and is based on verifying without trust even if it is an internal user. While the perimeter security model focuses on blocking, the zero-trust model focuses on thorough and continuous verification rather than blocking. The comparison between traditional security model and zero-trust model is presented in Table 2.

Table 2. The comparison between traditional security model and zero-trust model.

Features	Traditional Security Model	Zero-Trust Model
Approach	Trust but verify	Trust nothing and verify everything
Trust Boundary	External (Non-trust), Internal (Trust)	Micro Segmentation
Access Control	IP (Port, Protocol) based access control	Data-centric access control
Communication Encryption	External (Encryption)/Internal (No Encryption)	Full traffic encryption
Authentication	Once verification at initial access	Before access and continuous verification
Security Policy	Pre-defined rules and common policies	Fine-grained rules and adaptive policies (Needs Security Assessment)
Security Managements	Individual Monitoring and visibility	Visibility, automation orchestration of behaviour, devices, services and security

The Zero-Trust Architecture (ZTA) can quite simply be described as a set of coordinated system design principles based on the core concept that threats to a computer network or integrated networks can originate both internally and externally. This proposed network system design concept requires near-continuous verification and analysis of network nodes, services, applications and groups of users. This is simply because the network does not implicitly trust any entity. Only after the process of verification and subject to a decreasing time interval can elements of the network be entrusted to access, utilize and even modify network resources such as databases, other network nodes, servers present on the network and network policies. Once the pre-established time interval expires, the node must go through the verification process again.

It is important to note that even authorised nodes are given least-privilege access, which means access to only necessary components, 'no more no less' than what is required. To employ such a system, the network must have the following properties:

- Automation of system security;
- Dynamic access control policies based on risk management;
- Internal and external network traffic monitoring;
- Behaviour analysis for network nodes;
- Complete infrastructure visibility;
- Focus on protection of mission critical assets ensuring business continuity.

5.1. Leverage Zero-Trust Design Concepts

Each following subsection of this section is supposed to represent the stages in achieving Zero-Trust Maturity. The following points are meant to represent the stages of Zero-Trust Maturity:

1. **Define Architecture Priorities—**Derive your priorities from organization-specific mission requirements. The design team must identify the critical assets, services and data (ASD).
2. **Design the network architecture from the inside out—**Primarily, the focus should be on protecting the ASD; after this has been achieved, list and harden methods of accessing ASD.
3. **Devise access control policies—**Creation of a consistent security policy and standard operating procedures (SOPs) must be performed across all environments: endpoint, internal, edge and perimeter.
4. **Establishing visibility and Automation—**This is the last stage of maturity. Every activity/event occurring across all environments must be collected and analyzed by an automated threat detection system. Archiving major threats and events of interest should be conducted. Leveraging data analytics for detecting rogue users should be the outcome of this stage.

5.2. Transitioning to ZTA

When the process of transition begins, many challenges will arise. The first hurdle can be a lack of support from department administrators, executive members of management, normal employees and sometimes even top-level leadership. The second challenge may come in the implementation phase, oversights in access controls and proper configuration of security responses may allow some parts of the network to be left accessible by non-essential entities. So, it is essential that network engineers and company security teams are motivated to follow-up on these issues and plug in the gaps. Meanwhile, the management must also have the will to support the initiative financially and operationally.

Finally, the most prominent issue would come after the system is operational, work fatigue may set in and as time progresses the security posture can degrade due to overwork and the embodying the mindset of constant network compromise [43].

5.3. Achieving Zero-Trust Maturity

Before we begin to modify our network, we must adopt the mindset crucial to effectively make use of the security controls. It is essential to define expected outcomes when the system is fully operational. The organisation must identify its mission critical assets deployed in the network, cloud and on-site. The designers should have complete visibility over the components which would be added to the final network, this would enable them to assess and create an overall strategy for the protection of the entire system right from the design phase.

There must also be security measures which align themselves with the normal functioning of the network, they should not be more intrusive than they need to be, so it is essential to define the acceptable boundaries of surveillance.The National Security Agency

(NSA) provides four main principles for adopting the Zero-Trust mindset [43]. They are as follows:

(a) Coordinated management and monitoring of the system as well as its defensive capabilities;
(b) Assume all requests for critical resources and network traffic to be malicious;
(c) Assume that the network infrastructure and devices are already compromised;
(d) Assume all approvals for critical resources as risky and be prepared to perform damage control and recovery operations.

Implementing the above for, say, an enterprise-grade network cannot be done quickly, and transitioning everything and everyone all at once can cause faults to be introduced. So, we must have a maturity model, where the network is modified in a transitional manner, phase by phase. Most firms have existing infrastructure which requires the acquisition of additional software and hardware for this transition. The Figure 2 provides a visual discernment of what we have discussed so far.

Second, the Cybersecurity and Infrastructure Security Agency (CISA) in the U.S. released a zero-trust maturity model (ZTMM). The ZTMM model was classified into five pillars: identity, device, network/environment, application/workload and data from an asset perspective to be protected. Each pillar includes common elements regarding visibility and analysis, automation and orchestration, and governance. When the organization applied the zero-trust model, zero-trust maturity was divided into traditional, advanced and optimal levels. The traditional stage means a level without zero trust implementation. The traditional stage means a level without the implementation of a Zero-Trust model. The advanced stage means some implementation of a Zero-Trust model and the optimal stage means the fully automated implementation.

Third, Microsoft published a maturity model to implement the zero-trust model. MS's maturity model is similar to CISA's model, divided into six security elements: ID, device, application, infrastructure, network and data. In addition, the level of maturity was divided into traditional, advanced and optimal stages. The comparisons of ZTMM models (NIST, CISA, Microsoft) is shown in Table 3.

Table 3. The comparisons of ZTMM models (NIST, CISA, Microsoft).

Category		**NSA Model**	**CISA Model**	**MS Model**
Maturity Levels		**5 Stages**	**3 Stages**	**3 Stages**
Security Elements	Identities	✓	✓	✓
	Device	✓	✓	✓
	Network	✓	✓	✓
	Application		✓	✓
	Workload		✓	✓
	Data		✓	✓
	Infrastructure (VM, Cloud etc.)	✓		✓
	Visibility and Analytics	✓	✓	
	Automation/Orchestration	✓	✓	
	Governance (Polices)	✓	✓	

We need to understand that as the Zero-Trust Model is brought online, issues will most certainly occur, but coordination of the network improves over time as it is used. The enhanced infrastructure visibility and automation of security controls will give network administrators the ability to better thwart threats and mitigate risks before major damage can occur, much more than a traditional perimeter security system.

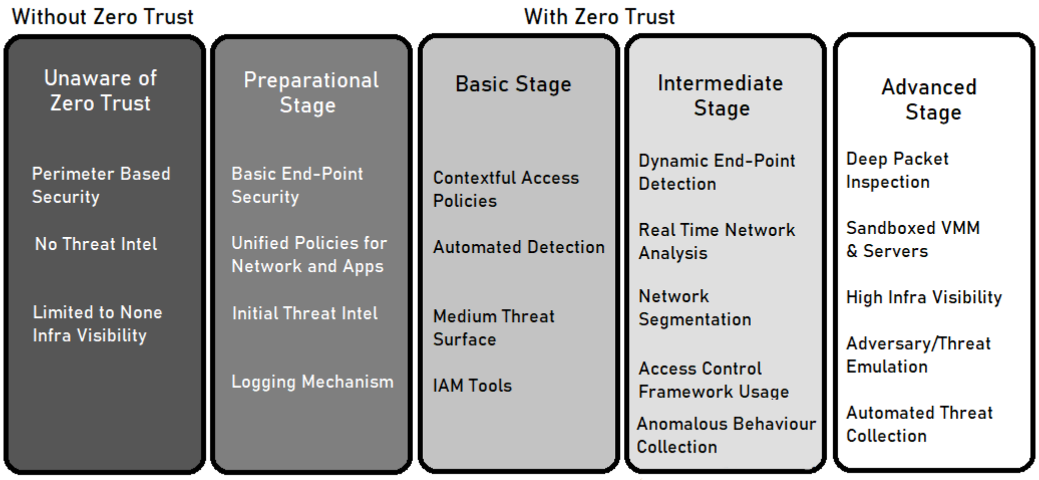

Figure 2. Stages of Zero-Trust Maturity.

5.4. Implementing the ZTA

A thorough implementation of a Zero-Trust Network must make use of several network sensors to achieve elevated levels of awareness inside and on the perimeter of the network. This system should not negatively impact performance; security is there only to prevent unauthorised events. Additional measures may have to be taken when implementing the model for a network, for example, east–west security controls (segmentation of network), grouping of users and nodes based on access policies, end-point detection response (EDR) and sandboxing of certain network components such as the VMM.

The ZTA is incomplete without an algorithm which calculates trust. As seen in Section 2, there are many variations in the method of calculating trust. However, it is the algorithm which will decide whether to deny or allow each individual request for access and use of network resources. Such an algorithm can be thought of as a function with inputs being relevant attributes such as access request type, previous behaviour of the requesting entity, resource usage history, previous history of penalties, current IAM policies, trust of the group to which the entity belongs to and current threat scenario, etc. Some of these are described in further detail below:

(a) **Current Threat Scenario**: This may include the type, severity and time intervals of recently detected attacks or network anomalies. This is a collection of threat intelligence to create a threat matrix for the network. The trust algorithm can use certain attributes selectively from the pool of threat attributes as needed.

(b) **Resource Usage Behaviour**: Previous usage of resources such as databases requests and network protocols used to access administrated areas, files and directories accessed on web servers and applications generally used by the user can be used as attributes for trust calculation. This would provide a histological aspect to the trust calculated: how a specific network resource was used and how it related to the work assigned to that entity.

(c) **Access Request**: The current request by the entity for entry to a protected component must have some additional metadata, such as time of request, number of previous requests made for the same component, level of authority required for access, current details of service or application being used to make the request (software version, OS version) and logical network identifiers such as MAC addresses and IP address.

(d) **Resource Status and Requirements**: In a dynamic network, the individual security of certain resources can be elevated or decreased as needed. The network may increase scrutiny of requests being made for access to that particular resource. So, in a case where the network might have recently had unauthorised access or anomalous behaviour, the ZTA authority may require the entity requesting access to meet certain security standards such as updated software version, updated firmware version, use of only secure protocols and accepting responsibility for their actions when given access.

The algorithm may assign more attributes specific to the current network condition and give different weights to each attribute. This would allow granular, on-the-fly changes to the network security policy by security administrators as needed, as proposed by NIST.

5.5. Expanding the ZTA

Once a flow of actions has been integrated into the network response mechanism, a pre-set response policy can help automate responses to suspicious behaviour in the network. Now would be the time to establish a database of past recorded anomalies, which the network authority can use to further calculate trust values for requests and flow of data. If this is not possible, network responders should try to use operational experience to modify access policies as needed. However, automation would be more suitable for this task since the key concept is alleviating human errors and lapses in oversight.

It is now that the ZTA would be operational and extremely mission-critical networks can be further integrated into the established network, creating a distributed zero-trust network. It must be noted that with a major overhaul of the network such as the introduction of completely new virtualisation hardware and software, critical web server migration, shifting endpoint network terminals to a different OS, software updates to network component firmware, etc., there must be a re-evaluation of the ZTA and how the flow of response occurs. It is not necessary to change the ZTNA but rather re-evaluate as needed. Needless overhauls can also introduce security flaws. Another area of focus might be in Internet-of-Things (IoT) networks where Zero Trust can be used to validate transactions and nodes can have varying levels of trust, as discussed by Samaniego and Deters in [44]. Zero Trust can also be combined and used in conjunction with other novel technologies such as blockchain and the IoT, as proposed by Dhar and Bose [45].

6. Comparative Analysis of Zero-Trust Cloud Network Technologies

As zero trust is not a monolithic architecture, it employs many proven and emerging technologies. Comparing these is essential to sorting out the best- or worst-fit features. As adversaries change, regressive designs can be retired from the domain in favour of ones which work. The operational necessity is given precedence over the economic efficiency of this model. Furthermore, the various authors provide crucial insight into how most papers have contributed mainly to the architectural design and approach of ZTNs. These parameters were used for comparisons, reflecting the common key necessities found across many different implementations of cloud networks. The basis for using these parameters is discussed below, with references to works supporting their requirement. Although this is not an exhaustive list of parameters, the above set can be considered a higher priority based on the pre-established challenges.

1 **Variable Trust Levels**—In an unreliable environment such as a computer network, cooperation of beneficiary nodes as a distinct group has been shown to be a key component of the network's safety [46]. Since the security of resources is of paramount importance for the proper operation of cloud networks, determining the trustworthiness of a request must be based on available historical information. Judging every request with an unchanging standard causes gaps to arise in the authentication. Thus, the separation of trustworthiness of nodes according to different levels is optimal. A sound analogy can be drawn from a study on trust management in ad-hoc wireless networks, where rogue nodes' selfish behaviour regularly disrupts network operations, causing drops in throughput [47].

2 **Access Control Policies**—Implementing access controls for users in a fragmented network with different policies and standard procedures is not a new design choice [48]. A homogeneous policy globally across many data centres for a company can be catastrophic. User access controls in industrial control systems in standard practice, especially those connected to cloud-dependent hardware [49].

3 **Includes DMZ**—Creating a buffer zone with limited visibility for networks (DMZ) is the surface area of the main network. Unauthorized users may be able to exploit vulnerabilities in this zone, but access to the main network protected by a hard firewall would be difficult and detectable. Defense-in-depth, although a traditional aspect of network security, can be the front-line rather than the only counter available to a network [50].

4 **Logging Mechanism**—Maintaining network logs can be performed relatively quickly, but efficiency is the key in the cloud. Storing logs in a standard format which can be parsed by automated software such as intrusion detection and prevention can maximize the potential of preemptively safeguarding the network from attacks [51].

5 **Supports Segmented Networks**—Cloud networks are often deployed block by block, with services becoming operational in a transitory phase. This is practical as demand fuels the growth of a company's capabilities, and cloud platforms amply provide flexibility in the form of subscription plans for additional storage or computational needs. Segmenting sections is essential to keeping the confidentiality and integrity of data used by each cloud service or department in a company. Network devices need to segregate and direct traffic based on the affiliation of the service in the overall organizational structure, as demonstrated by [52]. Segmentation of the network also has some security benefits, as discussed by Du et al. [53], and improves the automation of the network's defenses against the enumeration of its resources [54,55].

6 **Supports Multi-Cloud Environment**—The cloud industry's gravitation towards offering support for many different cloud platforms can be attributed to its many advantages. There may be unique features and offerings by disparate CSPs that, when coupled, enhance the business capabilities of CSC, such as maintenance of information, preservation of data confidentiality, management of delicate infrastructure and improvements in performance [56–60].

7 **Supports Geographical Distribution of System**—Due to natural disasters, businesses can be severely affected if their offerings can no longer be available. Thus, cloud networks require reliability and fall-back options. CSCs are also unevenly distributed across the globe, and parallel servers must be provided to reduce issues such as latency, traffic load balancing, non-repudiation of data, and primarily balancing the allocation of resources offered to users [61,62].

8 **Supports Open-Source Tools**—Open-Source has long been a cornerstone of public engagement in software development. The use of open-source tools and software which are free and publicly available gives government institutions and corporations the ability to modify the software as per their need and quickly deploy it for use. Open-Source software is generally more secure and has had many improvements, which comes with the pooling of human resources and skillsets common in Open-Source communities. A case study by Rodriguez-Martinez et al. [63] on the use of Open-Source software for weather systems hosted on the cloud showed that the cost of ownership was relatively low, the reliability of the system high and the huge potential of scalability for their system. The only downsides were the learning curve was challenging at first and the lack of specific management tools such as a Graphical User Interface (GUI)-based system for operating their cloud stack. A similar study by Huang et al. [64] on open-source cloud computing solutions for geo-sciences showed that the performance of such systems was better in most and comparative in others to commercially available counterparts. Finally, other potential benefits were described by I. Voras et al. [65].

9 **Support for Containers and Micro-services**—The final parameter support for micro-services and containerized applications results from consumer demand. The use of containers provides higher levels of scalability, reliability, and isolation of sensitive resources. Packaging entire programs or software suites is efficient and streamlines maintenance. This is potentially valuable for governance-related scenarios such as healthcare or use in Internet of Things (IoT)-enabled networks for vehicles [66–69].

These parameters were used after a study of the aforementioned papers, and they reflect the key common necessities found across many different implementations of cloud networks. Although this is not an exhaustive list of parameters, the above set can be considered higher priority based on the pre-established challenges.Variable levels of trust for different areas of the network in a cloud environment is a key necessity. Separation of bias during analysis of malicious events in one domain of the network would provide a reasonably specific picture of the daily events occurring in a certain part of the network and eliminate false-positives. It would also enable administrators to assess users not just individually but as a whole group and adopt specific countermeasures. This is where Access Control Policies come into use, although not essential, having a configurable default policy setting globally across many data-centers for a company can be considered essential. Logging mechanisms are important as they provide transparency providing detailed reasoning behind the calculation of trust for a particular entity, based on known events. The comparison of existing research models based on parameters associated with Zero-Trust Cloud Networks is shown in Table 4.

Table 4. A detailed comparison of existing research models based on parameters associated with Zero-Trust Cloud Networks. {P1: Variable Trust Levels, P2: Access Control Policies, P3: Includes DMZ, P4: Logging Mechanism, P5: Supports Segmented Networks, P6: Supports Multi-Cloud Environment, P7: Supports Geographical Distribution of System, P8: Supports Open-Source Tools, P9: Support for Containers and Micro-services}.

Authors	Novel Features	P1	P2	P3	P4	P5	P6	P7	P8	P9
Casimer et al. [23]	Transport Access Control and First Packet Authentication		✓		✓	✓		✓	✓	
Dayna et al. [31]	Autonomic control plane threat response using Boyd's OODA framework	✓	✓		✓	✓			✓	
Casimer et al. [30]	Transport Access Control and First Packet Authentication with geolocation attributes		✓		✓	✓		✓	✓	✓
Anwar et al. [32]	User data privacy protection for location-based services using data partitioning		✓		✓	✓				
Abdullah et al. [33]	Identity verification using client-server model and adaptive behaviour evaluation model of trusted nodes	✓	✓	✓	✓	✓				

As part of a maturity cycle, having a pre-established DMZ can be useful. It would filter out many commonly used attack patterns originating externally and it would carry no additional commitment of resources. In 2020, almost 93% of all organisations using the cloud were adopting a multi-cloud strategy [70]; this is why multi-cloud support is a future-proof solution for networked security systems, especially ZTA-based solutions which make use of a wide scope of emerging and proven technologies. Considering that the big three cloud services are based across different geographies and have large distances and are under many differing jurisdictions, it would be pertinent to have some semblance of support for distributed cloud networks. Taking into account latency in situations where multiple cases of repudiation show up or unwanted feedback loops are formed in cloud environments communicating across great physical distances, it would be wise to have a well-tested remediation system in place for any ZTCN. The comparison of existing research frameworks based on parameters associated with Zero-Trust Cloud Networks is shown in Table 5.

Table 5. A comparison of existing research frameworks based on parameters associated with Zero-Trust Cloud Networks. {P1: Inclusion of Variable Trust Levels, P2: Develops Access Control Policy Language, P3: Includes DMZ, P4: Includes Logging Mechanisms, P5: Includes Segmentation of Networks, P6: Includes Multi-Cloud Strategy, P7: Accounts for Geographical Distribution of System, P8: Includes Performance Analysis, P9: Includes Containers and Micro-services}.

Authors	Novel Feature	P1	P2	P3	P4	P5	P6	P7	P8	P9
Romans et al. [34]	Fuzzy Risk evaluation-based access control enforcement framework		✓	✓		✓				
Abdallah et al. [36]	SDP-based framework using a client-gateway architecture			✓	✓	✓			✓	
Mehraj et al. [35]	Typologies of Trust in Zero-Trust context. Use of Zero-Trust Triangle for calculation of trust for an entity		✓	✓	✓	✓				
Ahmed et al. [37]	Zero-trust framework for federated Identity Access Management in Cloud Computing using decentralised audit logs		✓		✓	✓	✓			
Simone et al. [27]	Performance Analysis of the cloud data plane under load and impact on the control plane			✓		✓	✓		✓	✓

Segmentation of a network is the only parameter other than variable level of trust which is common across all solutions compared in this paper, except Zirak Zaheer's perimeter security solution. This clearly shows that segmenting areas is highly beneficial to the security posture of any system in the long run. The comparison of existing proof-of-concept technologies associated with Zero-Trust Cloud Networks based on the various parameters is shown in Table 6.

Table 6. A comparison of existing proof-of-concept technologies associated with Zero-Trust Cloud Networks based on the following parameters. {P1: Variable Trust Levels, P2: Access Control Policies, P3: Supports DMZ, P4: Logging Mechanism, P5: Supports Segmented Networks, P6: Supports Multi-Cloud, P7: Supports Geographically Distributed Cloud System, P8: Supports FOSS, P9: Supports Containers and Micro-services}.

Authors	Novel Feature	P1	P2	P3	P4	P5	P6	P7	P8	P9
Zirak et al. [27]	Network-independent perimeter solution which traces authentic identities using per-packet tagging and verification.		✓	✓	✓			✓	✓	✓
Weever et al. [38]	Operational controls which mitigate data leaks during service-to-service transit of data in public cloud.					✓	✓	✓	✓	✓

In the many implementations of Zero-Trust Cloud Networks we encountered, almost all models included support for segmentation of the cloud network. This indicates an increase in complexity of the systems being hosted on cloud networks, leading to the need for segmentation. Again, almost all implementations included a logging mechanism. Even if not visible to the administrators, the logging mechanism stores necessary proof for future trust calculations and enforcement of access control policies.

Open-source tool integration is also on the rise as administrators need tools which give them fine-grained control of the cloud network. Another interesting observation was that of geographical distribution of the systems or support for networks hosted across a huge physical distance. This may indicate a wider use of network resources across different regions. The addition of micro-segmentation and Zero Trust does not have an enormous impact on performance of the network as stated by Muji M. et al. [71].

There currently exist proof-of-concepts for key components of a fully independent zero-trust network, purpose-built for cloud micro-services. eZTrust is the most mature in a practical and deployment use-case. An operational control and monitoring tool for in transit data by Weever and Andreou adds to potential features for future ZTCNs [36].

In the future, more support is required for applications hosted as micro-services and containers on the cloud, as well as multi-cloud environments, where hardware and software from different vendors are in use.

7. Potential Future Work

Zero Trust as the basis for cloud networks should be used to re-prioritize existing technologies available to the end-user and design a streamlined system to minimize authentication delays while ensuring business continuity. ZTA currently uses security information and event management (SIEM), data analytics, trust calculation using event logging, modification of file system permissions using active directory and multi-factor authentication (MFA). These are individual technologies used in conjunction. There remains a vast scope of improvement and addition to the compendium of ZTCN technologies; some of the areas of particular focus are listed below:

- **Internet-of-Things and Blockchains**: Zero-Trust can be used to validate transactions, and nodes can have varying levels of Trust. As proposed by Dhar and Bose, Zero Trust can also be combined and used with other novel technologies such as blockchain and the IoT.
- **5G/6G Networks**: Given the advances in data transfer rates, traditional security controls will become overwhelmed by the sheer quantity and variety of data they have to process and verify. This calls for revolutionary changes in the design of network protocols and how the routing of data works in 5G-enabled networks [72]. Artificial Intelligence algorithms can deter malicious requests and thwart network performance degradation in real-time or almost real-time [73]. This would be essential in mission-critical sectors such as healthcare, air defense and autonomous vehicle networks [74].
- **Military Networks**: The need for Zero Trust arose from concerns about the reliability of military communication networks spanning many different operational environments. It has now come full circle and matured into a technology used to keep adversaries out of military networks, many of which use cloud services ranging from non-essential to highly critical. In conventional armed forces, offensive and defensive cyber capabilities are often developed in isolation in different branches and agencies, called 'silos', and there is no real-time sharing of capabilities. This enables adversarial nation-states and APTs to exploit this gap in the command structure and breach government data servers [75]. Thus, with this philosophy, the United States Army is migrating to a Zero-Trust architecture as mandated by EO 14028. The zero trust road-map developed by the U.S government involves the NSA, DoD, CISA, NIST and OMB all pooling their resources to convert the entire government ecosystem into a ZT-based one. A simultaneous transformation effort by the various agencies will result in a system with a common operational picture devised by the DoD's Zero Trust Reference Architecture and meet the national security objective.
- **Containerized software and micro-services**: Using micro-services as a basis for comparison shows it requires further research and improvements. Micro-services are essentially programs running in an austere environment; the problems associated with scheduling and managing processes generated by such micro-services can metastasize into more significant problems [69]. More performance improvements are necessary to analyze the micro-services' impact on cloud networks thoroughly.
- **Sustainable Cloud Systems**: Any IT system's security relies on its availability just as much as its confidentiality and integrity. Thus, future cloud networks must not only balance security and reliability, but effective use of electricity and overall sustainability

of their network design must also be a focus. Sustainable cloud networks would have a higher tolerance for unwarranted disruptions [76].

8. Conclusions

This paper presented a comprehensive discussion of current zero-trust cloud network technology implementations along with their strengths and limitations. Zero Trust provides a highly granular and case-specific solution to network security issues in the cloud. It is a highly agile approach; however, further research and commercial use are required to present comprehensive conclusions about its effectiveness in real-world deployments. The scope of our paper is limited to publicly available research projects. This paper compares security-based features of recently published zero-trust-based cloud network models, frameworks and proofs-of-concept employed for network security. Comparing these models and frameworks used in zero-trust networks will enable future researchers to focus on security issues and oversights plaguing modern cloud computing networks. It allows them to create robust zero-trust cloud networks and implement intelligent Security Orchestration, Automation and Response. Commercial software products based on Zero-Trust Architecture for the cloud exist and require further extrapolation of their effectiveness. VMware's Carbon Black is a great example, which provides east–west security using multi-hop network traffic analysis.

The scope of this paper is such that future researchers would be able to follow the general timeline and milestones in developing the Zero-Trust Architecture their cloud platform needs. This would be beneficial to map the actual capabilities and operational needs of their network. It would inhibit feature creep in their design while allowing their network to become more agile, automated and transparent in its decision making. Currently, many network test-beds and proofs-of-concept are available for specific purposes; unfortunately, none provide a comprehensive platform or a 'one-size-fits-all' type of system. To alleviate this, a cloud network architecture which uses a modular, add-drop style of trust-based technologies can be developed in the future. This can provide the right balance of business flexibility and adaptive security.

Author Contributions: Conceptualization, S.S., G.C., S.K.S., A.H. and H.K.; methodology, S.S., G.C. and A.H.; validation, G.C., S.K.S., A.H. and H.K.; investigation, G.C. and H.K.; resources, G.C., S.K.S., A.H. and H.K.; data curation, S.S., G.C., S.K.S., A.H. and H.K.; writing—original draft preparation, S.S., G.C. and H.K.; writing—review and editing, G.C. and H.K.; visualization, G.C., S.K.S., A.H. and H.K.; supervision, G.C., S.K.S., A.H. and H.K.; project administration, G.C. and H.K.; funding acquisition, H.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Institute of Information and communications Technology Planning and Evaluation (IITP) grant funded by the Korea Government (MSIT) (No. 2021- 0-00358, AI Big data based Cyber Security Orchestration and Automated Response Technology Development).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Share of Corporate Data Stored in the Cloud in Organizations Worldwide from 2015 to 2022. Available online: https://www.statista.com/statistics/1062879/worldwide-cloud-storage-of-corporate-data/ (accessed on 30 July 2022).
2. El-Shrkawey, M.; Alalfi, M.; Al-Mahdi, H. An Enhanced Intrusion Detection System Based on Multi-Layer Feature Reduction for Probe and DoS Attacks. *J. Internet Serv. Inf. Secur.* **2021**, *11*, 61–78.
3. Rahmadika, S.; Firdaus, M.; Lee, Y.H.; Rhee, K.H. An Investigation of Pseudonymization Techniques in Decentralized Transactions. *J. Internet Serv. Inf. Secur.* **2021**, *11*, 1–18.
4. Pagano, F.; Verderame, L.; Merlo, A. Understanding Fuchsia Security. *J. Wirel. Mob. Netw. Ubiquitous Comput. Dependable Appl.* **2021**, *12*, 47–64.

5. Gupta, T.; Choudhary, G.; Sharma, V. A survey on the security of pervasive online social networks (POSNs). *arXiv* **2018**, arXiv:1806.07526.
6. Velumani, R.; Sudalaimuthu, H.; Choudhary, G.; Bama, S.; Jose, M.V.; Dragoni, N. Secured Secret Sharing of QR Codes Based on Nonnegative Matrix Factorization and Regularized Super Resolution Convolutional Neural Network. *Sensors* **2022**, *22*, 2959. [CrossRef] [PubMed]
7. Unit 42 Cloud Threat Report 2H 2021. Available online: https://www.paloaltonetworks.com/prisma/unit42-cloud-threat-research-2h21 (accessed on 30 July 2022).
8. Teerakanok, S.; Uehara, T.; Inomata, A. Migrating to zero trust architecture: Reviews and challenges. *Secur. Commun. Netw.* **2021**, *2021*, 9947347. [CrossRef]
9. Greitzer, F.L.; Purl, J.; Sticha, P.J.; Yu, M.C.; Lee, J. Use of Expert Judgments to Inform Bayesian Models of Insider Threat Risk. *J. Wirel. Mob. Netw. Ubiquitous Comput. Dependable Appl.* **2021**, *12*, 3–47.
10. Rahmadika, S.; Astillo, P.V.; Choudhary, G.; Duguma, D.G.; Sharma, V.; You, I. Blockchain-based Privacy Preservation Scheme for Misbehavior Detection in Lightweight IoMT Devices. *IEEE J. Biomed. Health Inform.* **2022**.[CrossRef]
11. Alagappan, A.; Venkatachary, S.K.; Andrews, L.J.B. Augmenting Zero Trust Network Architecture to enhance security in virtual power plants. *Energy Rep.* **2022**, *8*, 1309–1320. [CrossRef]
12. Tyler, D.; Viana, T. Trust No One? A Framework for Assisting Healthcare Organisations in Transitioning to a Zero-Trust Network Architecture. *Appl. Sci.* **2021**, *11*, 7499. [CrossRef]
13. Astillo, P.V.; Choudhary, G.; Duguma, D.G.; Kim, J.; You, I. TrMAps: Trust management in specification-based misbehavior detection system for IMD-enabled artificial pancreas system. *IEEE J. Biomed. Health Inform.* **2021**, *25*, 3763–3775. [CrossRef] [PubMed]
14. RightScale 2019 State of the Cloud Report, (March 2022). Available online: https://www.flexera.com/about-us/press-center/rightscale-2019-state-of-the-cloud-report-from-flexera-identifies-cloud-adoption-trends#:~:text=In (accessed on 30 July 2022).
15. Garbis, J.; Chapman, J.W. *Zero Trust Security: An Enterprise Guide*; Springer: Berlin/Heidelberg, Germany, 2021.
16. Stafford, V. Zero trust architecture. *NIST Spec. Publ.* **2020**, *800*, 207.
17. Moving the U.S. Government Toward Zero Trust Cybersecurity Principles. Available online: https://www.whitehouse.gov/omb/briefing-room/2022/01/26/office-of-management-and-budget-releases-federal-strategy-to-move-the-u-s-government-towards-a-zero-trust-architecture (accessed on 30 July 2022).
18. Buck, C.; Olenberger, C.; Schweizer, A.; Völter, F.; Eymann, T. Never trust, always verify: A multivocal literature review on current knowledge and research gaps of zero-trust. *Comput. Secur.* **2021**, *110*, 102436. [CrossRef]
19. Alevizos, L.; Ta, V.T.; Hashem Eiza, M. Augmenting zero trust architecture to endpoints using blockchain: A state-of-the-art review. *Secur. Priv.* **2022**, *5*, e191. [CrossRef]
20. He, Y.; Huang, D.; Chen, L.; Ni, Y.; Ma, X. A Survey on Zero Trust Architecture: Challenges and Future Trends. *Wirel. Commun. Mob. Comput.* **2022**, *2022*, 6476274. [CrossRef]
21. Syed, N.F.; Shah, S.W.; Shaghaghi, A.; Anwar, A.; Baig, Z.; Doss, R. Zero Trust Architecture (ZTA): A Comprehensive Survey. *IEEE Access* **2022**, 10, 57143–57179. [CrossRef]
22. Pittman, J.M.; Alaee, S.; Crosby, C.; Honey, T.; Schaefer, G.M. Towards a Model for Zero Trust Data. *AJSE* **2022**, *3*, 18–24.
23. DeCusatis, C.; Liengtiraphan, P.; Sager, A.; Pinelli, M. Implementing zero trust cloud networks with transport access control and first packet authentication. In Proceedings of the 2016 IEEE International Conference on Smart Cloud (SmartCloud), New York, NY, USA, 18–20 November 2016; pp. 5–10.
24. Kindervag, J. *No More Chewy Centers: The Zero Trust Model of Information Security*; Forrester Research Inc.: Cambridge, MA, USA, 2016.
25. Kindervag, J. *Build Security into Your Network's Dna: The Zero Trust Network Architecture*; Forrester Research Inc.: Cambridge, MA, USA, 2010.
26. Zhang, P.; Tian, C.; Shang, T.; Liu, L.; Li, L.; Wang, W.; Zhao, Y. Dynamic access control technology based on zero-trust light verification network model. In Proceedings of the 2021 International Conference on Communications, Information System and Computer Engineering (CISCE), Beijing, China, 14–16 May 2021; pp. 712–715.
27. Rodigari, S.; O'Shea, D.; McCarthy, P.; McCarry, M.; McSweeney, S. Performance Analysis of Zero-Trust multi-cloud. In Proceedings of the 2021 IEEE 14th International Conference on Cloud Computing (CLOUD), Chicago, IL, USA, 5–10 September 2021; pp. 730–732.
28. Shore, M.; Zeadally, S.; Keshariya, A. Zero Trust: The What, How, Why, and When. *Computer* **2021**, *54*, 26–35. [CrossRef]
29. D'Silva, D.; Ambawade, D.D. Building a zero trust architecture using Kubernetes. In Proceedings of the 2021 6th international conference for convergence in technology (i2ct), Maharashtra, India, 2–4 April 2021; pp. 1–8.
30. DeCusatis, C.; Liengtiraphan, P.; Sager, A. Advanced intrusion prevention for geographically dispersed higher education cloud networks. In *Online Engineering & Internet of Things*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 132–143.
31. Eidle, D.; Ni, S.Y.; DeCusatis, C.; Sager, A. Autonomic security for zero trust networks. In Proceedings of the 2017 IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON), New York, NY, USA, 19–21 October 2017; pp. 288–293.

32. Jasim, A.C.; Hassoon, I.A.; Tapus, N. Cloud: Privacy For Locations Based-services' through Access Control with dynamic multi-level policy. In Proceedings of the 2019 6th International Conference on Control, Decision and Information Technologies (CoDIT), Paris, France, 23–26 April 2019; pp. 1911–1916.
33. Albuali, A.; Mengistu, T.; Che, D. ZTIMM: A zero-trust-based identity management model for volunteer cloud computing. In Proceedings of the International Conference on Cloud Computing, Honolulu, HI, USA, 18–20 September 2020; Springer: Cham, Switzerland, 2020; pp. 287–294.
34. Vanickis, R.; Jacob, P.; Dehghanzadeh, S.; Lee, B. Access control policy enforcement for zero-trust-networking. In Proceedings of the 2018 29th Irish Signals and Systems Conference (ISSC), Belfast, UK, 21–22 June 2018; pp. 1–6.
35. Mehraj, S.; Banday, M.T. Establishing a zero trust strategy in cloud computing environment. In Proceedings of the 2020 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 22–24 January 2020; pp. 1–6.
36. Moubayed, A.; Refaey, A.; Shami, A. Software-defined perimeter (sdp): State of the art secure solution for modern networks. *IEEE Netw.* **2019**, *33*, 226–233. [CrossRef]
37. Ahmed, M.; Petrova, K. A Zero-Trust Federated Identity and Access Management Framework for Cloud and Cloud-based Computing Environments. Available online: https://aisel.aisnet.org/wisp2020/4/ (accessed on 30 July 2022).
38. de Weever, C.; Andreou, M. *Zero Trust Network Security Model in Containerized Environments*; University of Amsterdam: Amsterdam, The Netherlands, 2020.
39. Zhu, G.; Yin, Y.; Cai, R.; Li, K. Detecting virtualization specific vulnerabilities in cloud computing environment. In Proceedings of the 2017 IEEE 10th International Conference on Cloud Computing (CLOUD), Honolulu, HI, USA, 25–30 June 2017; pp. 743–748.
40. Pék, G.; Buttyán, L.; Bencsáth, B. A survey of security issues in hardware virtualization. *ACM Comput. Surv.* **2013**, *45*, 1–34. [CrossRef]
41. Crowd Research Partners, Cloud Security Report 2018. Available online: https://crowdresearchpartners.com/portfolio/cloud-security-report/ (accessed on 30 June 2022).
42. Kandias, M.; Virvilis, N.; Gritzalis, D. The insider threat in cloud computing. In Proceedings of the International Workshop on Critical Information Infrastructures Security, Lucerne, Switzerland, 8–9 September 2011; Springer: Berlin/Heidelberg, Germany, 2011; pp. 93–103.
43. Info Sheet: Embracing a Zero Trust Security Model (February 2021). Available online: https://media.defense.gov/2021/Feb/25/2002588479/-1/-1/0/CSI_EMBRACING_ZT_SECURITY_MODEL_UOO115131-21.PDF (accessed on 30 June 2022).
44. Samaniego, M.; Deters, R. Zero-trust hierarchical management in IoT. In Proceedings of the 2018 IEEE international congress on Internet of Things (ICIOT), San Francisco, CA, USA, 2–7 July 2018; pp. 88–95.
45. Dhar, S.; Bose, I. Securing IoT devices using zero trust and blockchain. *J. Organ. Comput. Electron. Commer.* **2021**, *31*, 18–34. [CrossRef]
46. Zhang, B.; Huang, Z.; Xiang, Y. A novel multiple-level trust management framework for wireless sensor networks. *Comput. Netw.* **2014**, *72*, 45–61. [CrossRef]
47. Luo, J.; Liu, X.; Fan, M. A trust model based on fuzzy recommendation for mobile ad-hoc networks. *Comput. Netw.* **2009**, *53*, 2396–2407. [CrossRef]
48. Singh, S.; Choudhary, G.; Shandilya, S.K.; Sihag, V.; Choudhary, A. Counterfeited Product Identification in a Supply Chain using Blockchain Technology. *Res. Briefs Inf. Commun. Technol. Evol.* **2021**, *7*, 3.
49. Lopez, J.; Rubio, J.E. Access control for cyber-physical systems interconnected to the cloud. *Comput. Netw.* **2018**, *134*, 46–54. [CrossRef]
50. Dadheech, K.; Choudhary, A.; Bhatia, G. De-militarized zone: A next level to network security. In Proceedings of the 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), Coimbatore, India, 20–21 April 2018; pp. 595–600.
51. Tovarnák, D.; Vaekova, A.; Novák, S.; Pitner, T. Structured and interoperable logging for the cloud computing Era: The pitfalls and benefits. In Proceedings of the 2013 IEEE/ACM 6th International Conference on Utility and Cloud Computing, Dresden, Germany, 9–12 December 2013; pp. 91–98.
52. Jeuk, S.; Salgueiro, G.; Baker, F.; Zhou, S. Network segmentation in the cloud a novel architecture based on UCC and IID. In Proceedings of the 2015 IEEE 4th International Conference on Cloud Networking (CloudNet), Niagara Falls, ON, Canada, 5–7 October 2015; pp. 58–63.
53. Du, R.; Zhao, C.; Li, S.; Li, J. A strategy of network coding against wiretapping attack based on network segmentation. In Proceedings of the Second International Conference on Communications, Signal Processing, and Systems, Tianjin, China, 1–2 September 2013; Springer: Cham, Switzerland, 2014; pp. 1137–1144.
54. Wagner, N.; Şahin, C.Ş.; Winterrose, M.; Riordan, J.; Pena, J.; Hanson, D.; Streilein, W.W. Towards automated cyber decision support: A case study on network segmentation for security. In Proceedings of the 2016 IEEE Symposium Series on Computational Intelligence (SSCI), Athens, Greece, 6–9 December 2016; pp. 1–10.
55. Wagner, N.; Şahin, C.Ş.; Pena, J.; Riordan, J.; Neumayer, S. Capturing the security effects of network segmentation via a continuous-time markov chain model. In Proceedings of the 50th Annual Simulation Symposium, Virginia Beach, VA, USA, 23–26 April 2017; pp. 1–12.
56. Raj, J.S. Efficient information maintenance using computational intelligence in the multi-cloud architecture. *J. Soft Comput. Paradig.* **2019**, *1*, 113–124. [CrossRef]

57. Sulochana, M.; Dubey, O. Preserving data confidentiality using multi-cloud architecture. *Procedia Comput. Sci.* **2015**, *50*, 357–362. [CrossRef]
58. Kovács, J.; Kacsuk, P. Occopus: A multi-cloud orchestrator to deploy and manage complex scientific infrastructures. *J. Grid Comput.* **2018**, *16*, 19–37. [CrossRef]
59. Gundu, S.R.; Panem, C.A.; Thimmapuram, A. Hybrid IT and Multi Cloud an Emerging Trend and Improved Performance in Cloud Computing. *SN Comput. Sci.* **2020**, *1*, 256. [CrossRef]
60. Alshammari, M.M.; Alwan, A.A.; Nordin, A.; Al-Shaikhli, I.F. Disaster recovery in single-cloud and multi-cloud environments: Issues and challenges. In Proceedings of the 2017 4th IEEE International Conference on Engineering Technologies and Applied Sciences (ICETAS), Salmabad, Bahrain, 29 November–1 December 2017; pp. 1–7.
61. Endo, P.T.; de Almeida Palhares, A.V.; Pereira, N.N.; Goncalves, G.E.; Sadok, D.; Kelner, J.; Melander, B.; Mangs, J.E. Resource allocation for distributed cloud: Concepts and research challenges. *IEEE Netw.* **2011**, *25*, 42–46. [CrossRef]
62. Hogade, N.; Pasricha, S.; Siegel, H.J. Energy and network aware workload management for geographically distributed data centers. *IEEE Trans. Sustain. Comput.* **2021**, *7*, 400–413. [CrossRef]
63. Rodriguez-Martinez, M.; Seguel, J.; Greer, M. Open source cloud computing tools: A case study with a weather application. In Proceedings of the 2010 IEEE 3rd International Conference on Cloud Computing, Miami, FL, USA, 5–10 July 2010; pp. 443–449.
64. Huang, Q.; Yang, C.; Liu, K.; Xia, J.; Xu, C.; Li, J.; Gui, Z.; Sun, M.; Li, Z. Evaluating open-source cloud computing solutions for geosciences. *Comput. Geosci.* **2013**, *59*, 41–52. [CrossRef]
65. Voras, I.; Mihaljević, B.; Orlić, M.; Pletikosa, M.; Žagar, M.; Pavić, T.; Zimmer, K.; Čavrak, I.; Paunović, V.; Bosnić, I.; et al. Evaluating open-source cloud computing solutions. In Proceedings of the 34th International Convention MIPRO, Opatija, Croatia, 23–27 May 2011; pp. 209–214.
66. Esposito, C.; Castiglione, A.; Tudorica, C.A.; Pop, F. Security and privacy for cloud-based data management in the health network service chain: A microservice approach. *IEEE Commun. Mag.* **2017**, *55*, 102–108. [CrossRef]
67. Lakhan, A.; Memon, M.S.; Elhoseny, M.; Mohammed, M.A.; Qabulio, M.; Abdel-Basset, M. Cost-efficient mobility offloading and task scheduling for microservices IoVT applications in container-based fog cloud network. *Clust. Comput.* **2022**, *25*, 2061–2083. [CrossRef]
68. Amaral, M.; Polo, J.; Carrera, D.; Mohomed, I.; Unuvar, M.; Steinder, M. Performance evaluation of microservices architectures using containers. In Proceedings of the 2015 IEEE 14th International Symposium on Network Computing and Applications, Cambridge, MA, USA, 28–30 September 201; pp. 27–34.
69. Kyryk, M.; Pleskanka, N.; Pleskanka, M.; Kyryk, V. Infrastructure as Code and Microservices for Intent-Based Cloud Networking. In *Future Intent-Based Networking*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 51–68.
70. State of the Cloud Report, (March 2022). Available online: https://info.flexera.com/CM-REPORT-State-of-the-Cloud (accessed on 30 June 2022).
71. Mujib, M.; Sari, R.F. Performance evaluation of data center network with network micro-segmentation. In Proceedings of the 2020 12th International Conference on Information Technology and Electrical Engineering (ICITEE), Yogyakarta, Indonesia, 6–8 October 2020; pp. 27–32.
72. Dzogovic, B.; Santos, B.; Hassan, I.; Feng, B.; Jacot, N.; Van Do, T. Zero-Trust Cybersecurity Approach for Dynamic 5G Network Slicing with Network Service Mesh and Segment-Routing over IPv6. In Proceedings of the 2022 International Conference on Development and Application Systems (DAS), Suceava, Romania, 26–28 May 2022; pp. 105–114.
73. Ramezanpour, K.; Jagannath, J. Intelligent Zero Trust Architecture for 5G/6G Networks: Principles, Challenges, and the Role of Machine Learning in the context of O-RAN. *arXiv* **2021**, arXiv:2105.
74. Bello, Y.; Hussein, A.R.; Ulema, M.; Koilpillai, J. On Sustained Zero Trust Conceptualization Security for Mobile Core Networks in 5G and Beyond. *IEEE Trans. Netw. Serv. Manag.* **2022**, *19*, 1876–1889. [CrossRef]
75. Stewart, A. Three Emerging Innovative Technologies Required for Cyber Operations to Execute Commander's Intent at Machine Speed. *Mil. Cyber Aff.* **2020**, *4*, 3. [CrossRef]
76. Chen, T.; Marques, A.G.; Giannakis, G.B. DGLB: Distributed stochastic geographical load balancing over cloud networks. *IEEE Trans. Parallel Distrib. Syst.* **2016**, *28*, 1866–1880. [CrossRef]

Introducing Zero Trust by Design: Principles and Practice Beyond the Zero Trust Hype

Dwight Horne* and Suku Nair
AT&T Center for Virtualization, Southern Methodist University, Dallas, TX 75275, USA
{rhorne, nair}@smu.edu

Abstract. The zero trust (ZT) threat model is one in which no implicit trust exists. Companies across industries, governments, and more are at various stages of evaluation, planning, and implementation of re-architected networks that deviate from legacy perimeter-centric security models in favor of ZT approaches. A lack of open standards, or even consensus on ZT details and priorities in some cases, has potentially slowed adoption. But recent standardization attempts have taken strides toward filling that gap. ZT research and practice to date have focused predominantly on the network architecture and operations. However, to fully realize the potential benefits of the ZT model, we must expand our conception of ZT. In this paper, we introduce Zero Trust by Design (ZTBD), a set of guiding principles that extend beyond network architecture to also foster ZT software engineering and ZT protocol design, thereby maximizing synergies and potential benefits. Beyond the ZTBD principles, a set of ZT practices connect theory and guidelines with practical applications, while the introduction of ZT patterns provides reusable solutions to common challenges. Moreover, just as ZT research and practice are evolving, so too will ZTBD. We invite the community to further contribute to these efforts and support continued alignment of this growing body of knowledge with the rapid evolution and expansion of zero trust.

Keywords: zero trust by design, ZTBD, zero trust architecture, zero trust software engineering, zero trust protocol design, zero trust patterns, zero trust model.

1 Introduction

Trust no one. That was a recurring mantra from the well-known television show called The X-Files. The Zero Trust (ZT) threat model extends that idea beyond *trust no one [user]* to further include *trust no device* and *trust no network flow*. In an era of global pandemic accelerated pivots to remote work and major supply chain compromises like the SolarWinds attacks of 2019-2020 [1], ZT concepts are everywhere. ZT ranges from myriad vendor marketing campaigns and a recent USA National Institute of Standards and Technology (NIST) standard [2] to ZT conferences [3] and guidance from the National Security Agency of the USA [4]. Yet not all ZT practitioners fully agree on the critical characteristics of zero trust as evidenced by the results from a survey of technology professionals in 2020 [5]. That study found that 85% of respondents reported ongoing pursuit of defined ZT initiatives within their organizations. Meanwhile, only 12% of the technologists responding indicated that "location and method of access do not confer trust" was a defining characteristic of zero trust networking, despite that statement aligning with the second "basic tenet" of ZT as defined in the recently finalized NIST SP800-207 [2], a standard for Zero Trust Architectures. Consequently, one of the goals of Zero Trust by Design (ZTBD) is to harmonize the tenets of ZT from various resources by distilling them down to overarching principles that can be applied to varying contexts beyond the network architecture.

Moreover, just as it is widely recognized that full lifecycle software and systems security is more successful than when it is treated as an afterthought [6], we argue that benefits of the ZT model cannot be fully realized without extending the notion of ZT beyond the network architecture to include ZT protocol design and full lifecycle ZT software engineering. After distinguishing these separate but complementary notions, we extend the ZTBD principles with good practices for applying the core principles. We further launch an initiative for identifying ZT patterns – a collection of named solutions to commonly recurring problems as an additional enabler for ZT initiatives.

The remainder of this paper is organized as follows. Section 2 provides background information contrasting ZT architectures, ZT software engineering, and ZT protocols. Section 3 then highlights the ZTBD principles and augments them with good ZT practices. Next, Section 4 provides sample ZT patterns that accompany the launch of an initiative to build a large, openly available body of valuable enablers. Lastly, we conclude with an open invitation to the community to further contribute to these efforts for the benefit of all.

2 Background - A Contextual Understanding of Zero Trust

2.1 Zero Trust Security Model

In a ZT security model, no implicit trust exists. That is, all users, devices, network traffic, program inputs/outputs, and more might be considered hostile until proven otherwise depending on the context. In this paper, we consider the abstract idea of the ZT model and its application in certain contexts as an aid for applied researchers and practitioners. For theory related to definitions of trust across disciplines or trust modeling formalizations, please see related resources such as [7] and the works cited therein.

The general notion behind zero trust likely dates back to a time much earlier than when ZT terminology became a regular part of the technology lexicon. Technical solutions specifically targeting a ZT security model by that name date back to at least 2002 with an approach to zero trust intrusion tolerant systems that assumed compromise was inevitable [8]. That approach described self-cleansing systems that periodically restore themselves from a trusted source, a technique that resembles frequent infrastructure churn with trusted image restoration in modern networks and systems. The ZT model was later popularized with application to network architectures, as discussed in Section 2.2. But applicability of the model goes much further than the traditional network architecture centric view of ZT. ZT software engineering and ZT protocol design can be characterized as disparate but complementary applications of the ZT model in related domains. To fully realize the benefits of the ZT model, this comprehensive view of ZT is a necessity.

2.2 Zero Trust Network Architecture

The zero trust model was popularized in the context of network architectures as a complete re-thinking of the traditional, perimeter-centric approach to network security through the writing of Kindervag [9][10]. The early work of [8] described zero trust as "the extreme assumption that all intrusion detection techniques will eventually fail". Over time, that assumption is no longer viewed as "extreme" by many as companies deal with adversaries that have penetrated networks, very real insider threats, and perimeters that have dissolved with migration to cloud-based infrastructure and applications, along with significant remote work.

A ZT network architecture has been described as having Internet security everywhere, assuming the network is already compromised, and lacking implicit trust that might be common in legacy network architectures. The main logical components of the ZT architecture as defined by NIST SP800-207 include trust and policy engines composing the policy decision point (PDP) where access decisions are made. The PDP then communicates the access determinations in some way to the policy enforcement point (PEP) where the access controls are enforced. A sample trust evaluation process that might be used with a ZT network appears in Fig. 1. For more details regarding ZT network architectures, a number of useful resources exist such as [2], [11], and [12].

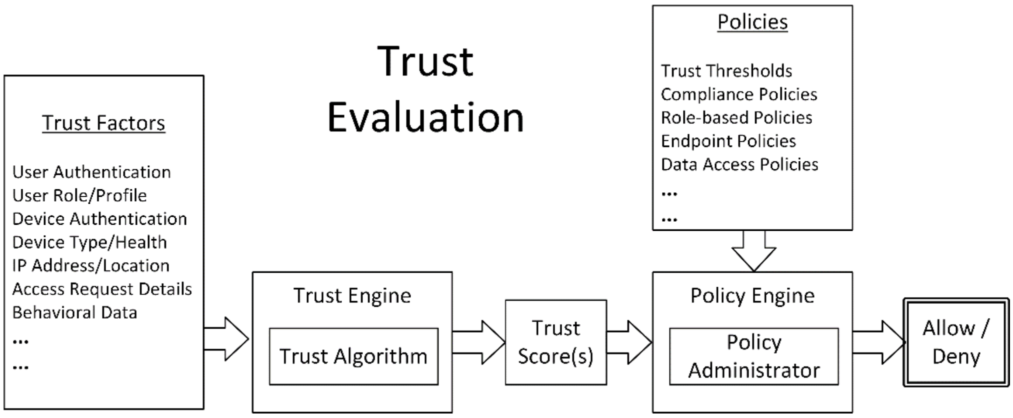

Fig. 1. An example trust evaluation process with zero trust network architectures.

2.3 Zero Trust Software Engineering

While ZT network architectures have received significant attention, ZT software engineering deserves designation as a separate but complementary domain. Viewing software engineering through the lens of the ZT security model can help facilitate a more complete realization of potential security and privacy benefits. ZT software engineering is not a major deviation from traditional software engineering, but rather a tailoring of existing requirements determination, design/development/deployment processes, and an overall change in mindset.

ZT software engineering explicitly recognizes the evolving nature of the threat landscape, the adaptations necessary to fully realize the potential of ZT networks, and features required to function properly within the context of ZT frameworks. For example, software applications must support the goals of the ZT network architecture in areas such as authentication, authorization, appropriate logging levels, and more in an environment with Internet security everywhere. Software lacking support for key features of the network architecture could inhibit ZT ambitions. In fact, Gilman and Barth devoted an entire chapter of their book to "Trusting Applications" in the context of ZT networks [11].

Moreover, insecure software could serve as the weakest link introducing unwanted vulnerabilities that might serve to circumvent controls. In an increasingly API driven world, consider that Gartner suggested "API abuse" would be the most common attack category by 2022 [13]. Perhaps more alarming, a report from Salt Security suggested that 91% of enterprise security professionals reported incidents in 2020 involving API security and that the rate of growth of malicious API traffic exceeded that of non-malicious API traffic [14]. Clearly, software requirements driven by the needs of a ZT network and secure software development practices, particularly for interfaces such as APIs, are critical elements of ZT software engineering.

2.4 Zero Trust Protocols

Protocol design plays a foundational role in support of both ZT architectures and ZT software engineering. Even the most secure networks and applications would fail to afford adequate protections with the use of insecure protocols. For this reason, it is important to recognize ZT protocol design as an essential enabler and contributor to ZT success. ZT protocols can be divided into at least two broad categories including ZT networking protocols and domain specific ZT protocols.

Zero Trust Networking Protocols. Gilman and Barth claim that ZT networks use existing technologies in novel ways, but do not require new protocols or libraries [11]. While that may be true in some sense, ZT networking can potentially be strengthened by new protocols, particularly if standardized with a focus on interoperability. Indeed, some researchers have already recognized this. For instance, [15] presents autonomic security for ZT networks with protocols and software to facilitate log parsing, orchestration, and threat response support. Meanwhile, others have worked to develop a generic firewall policy language to support protocols for dynamic access control policy enforcement [16]. Moreover, other lines of research have aimed to demonstrate how transport access control and first packet authentication protocols can be effectively used to implement ZT cloud networks [17]. Secure protocol design is expected to continue to play a foundational role in the evolution of ZT software and networks.

Domain Specific Zero Trust Protocols. The ZT model can also be used in the design of novel, domain specific protocols. There are a number of examples of protocols that embody one or more ZT principles, although they do not necessarily use ZT terminology directly. For example, the Bitcoin electronic cash protocols utilize a distributed ledger and hash-based proof-of-work to avoid reliance on a trusted third party and to offer a level of pseudo-anonymity [18]. The Monero cryptocurrency uses a different protocol with an obfuscated ledger attempting to achieve a greater level of anonymity, again assuming a lack of trust in third parties [19]. Privacy enhanced matchmaking technologies such as [20] and [21] exhibit various ZT principles applied to novel matchmaking protocols. Yet another example established protocols for ZT hierarchical trust management in IoT networks [22]. Domain specific protocol designs inspired by a ZT model are expected to continue to be active areas of research, likely with increasing adoption of practical technologies leveraging such protocols going forward.

1. **Trust no one (and no thing)**
   Assume compromise. No user, device, packet, or input is implicitly trusted.
2. **What perimeter?**
   Secure perimeters have morphed or disappeared in a cloud-centric world with remote work and insider threats. There are no implicit trust zones.
3. **Apply the principle of least privilege**
   Limit access to the minimum required; deny by default where feasible.
4. **Dynamic, risk-based policies**
   Use dynamic, contextual risk assessment and strict policy enforcement.
5. **Require strong authentication and authorization**
   All resource access requests and network flows must be authenticated and authorized.
6. **Log, monitor, inspect, and adapt**
   Traffic monitoring, input validation, status logging, analysis, alerts, adaptive trust levels, issue prevention and recovery actions.
7. **Employ defense in depth**
   Secure devices and communications; secure the weak links; think like an adversary and increase work factor for lateral movement.
8. **Full lifecycle, zero trust security**
   Full lifecycle security under a zero trust model; security is not an afterthought.
9. **Confidentiality and integrity by default**
   Encrypt all communications; protect data at rest; verify integrity.
10. **Strike the right balance with tradeoffs**
    Cost vs. benefit; security vs. usability; cost proportional to risk/impact.

Fig. 2. Foundational principles of Zero Trust by Design (ZTBD) v1.0

3 Introducing Zero Trust by Design

3.1 Core Principles of Zero Trust by Design

The principles of Zero Trust by Design (ZTBD) aggregate and harmonize core elements of the approach and communicate the essence of the ZT model, whether applied to network architecture, software engineering, protocol design, or another domain. The ZTBD principles codify fundamental considerations that are critical to the ZT mindset. The principles for ZTBD v1.0 appear in Fig. 2.

3.2 Beyond the Core Principles – Zero Trust Practices

Use of the ZTBD principles can facilitate application of the ZT model in specific contexts. Beyond the core principles, this work further attempts to capture ZT practices that have proven to be beneficial for translating the principles into practice. Notice that these are presented as good practices rather than best practices. This is because they have been shown to be good practices in many contexts, but claiming a universal best practice can be risky as it only takes one counter-example to prove otherwise.

Good Practices for ZT Networks. Application of the ZTBD principles can translate to a number of good practices for transforming high level principles of ZT network architectures to practical security controls. Some of such good practices include the following.

Strongly and securely identify users with multi-factor authentication. Strong identity validation is essential for trust evaluation and access control. Leverage multiple factors for authentication such as something the user knows (e.g., password), is (e.g., biometrics), and possesses (e.g., token).

Securely provision and strongly identify devices. Managed devices should be securely configured with trusted images, securely provisioned, and strongly identified, which may for example include use of a Trusted Platform Module (TPM) with certificate based authentication schemes like X.509.

Enforce device security. Devices should be in the most secure state possible at all times. Enforce timely patching of OS, firmware, and applications. Consider patch level and security posture as a factor in trust evaluation. Leverage appropriate security tools for the context such as antivirus, host-based firewalls, intrusion prevention software, and activity monitoring. Consider tailored access privileges and custom policies based on device types or functional characteristics. For example, one might treat a provisioned company laptop differently than an IoT network device or personal phone in a BYOD scenario.

Adopt an image refresh strategy. Somewhat akin to the original ZT work of [8], develop a strategy to restore from trusted base images where feasible. Consider the age of a device image to be a risk factor under the assumption that more usage time since original image installation implies greater probability of compromise.

Use dynamic policy enforcement and trust scoring. Evaluate trust based on as many pertinent data points as possible. Trust evaluation can potentially provide the greatest benefit if considering a variety of factors that may contribute to the level of trust such as user ID, device ID, sensitivity of target resource, temporal data, geographic data, and real-time threat intelligence data. While a core tenet of ZT security is that location alone should not confer trust, note that location could potentially be one of a number of factors considered when assessing risk depending on the context.

Employ re-authentication policies for users and devices, but balance security with usability in the process. There are risks associated with one-time authentication and permanent trust thereafter. It is common to adopt a periodic re-authentication strategy based on temporal constraints or continual risk assessments. However, when developing a re-authentication strategy, balance risk management with usability impact.

Leverage micro-segmentation and micro-perimeterization. Segregate networked resources strategically to facilitate least privilege access control policies and reduce risks. This can be accomplished in different ways depending on the context. For example, security gateway appliances such as firewalls, hypervisors, virtual network overlays, software-defined micro-perimeters, routing zones, Layer 2 subnetworks, and more. Micro-segmentation and software defined perimeters have become commonplace in complex cloud environments [23][24] and efforts to apply micro-segmentation to new areas such as 5G cellular networks [25] will also continue to be explored. Regularly monitor for the latest research trends and vendor offerings in this category for future enhancements.

Continually monitor network traffic, configure alerts, and adapt policies as needed. Ensure that continuous monitoring is part of the network security strategy. Additionally, log network traffic for post analysis, investigation support, training data sets for machine learning driven security solutions, and more. Make use of risk-based alerting features for risk detection and rapid response.

For more complex application scenarios, adopt a cross-functional ZT working group. The involvement of a cross-functional group, whether for actual decision making or advisory consultation purposes, can help to ensure that all stakeholders' interests are considered with ZT initiatives in complex environments.

Good Practices for ZT Software Engineering and ZT Protocol Design. Application of the ZTBD principles can also translate to good practices for software engineers and protocol designers. Some of such good practices include the following.

Encrypt all communications. There is no longer a "safe" internal network inside a secured perimeter (perhaps there never was!). Perimeters are dissolving. Even the best perimeter defenses can be circumvented, and the insider threat is real. Securely encrypt communications to the fullest extent practicable.

Authenticate/authorize network connections. Here again, there is no safe internal network guarded by a perfectly secured perimeter. Assume connection requests and packets are hostile until proven otherwise. One example of this might be via mutually authenticated Transport Layer Security (TLS).

Secure the software supply chain. The SolarWinds attack of 2020 significantly impacted many government agencies and Fortune 500 companies, and was a stark reminder of the criticality of securing the software supply chain [1]. The entire supply chain must be secured including code/artifact repositories, dependencies, build systems, DevOps continuous integration pipelines, and more.

Maximize automated use of security analysis tools. Beyond human security reviews, maximize use of value-added tools like static code analysis / static application security testing (SAST), dynamic application security testing (DAST), vulnerability scanners, and fuzzing tools.

Secure design and coding practices. Use appropriate practices like threat modeling, mandatory secure coding guidelines, security focused design and code reviews, and code signing.

Take API security very seriously. Despite broad encouragement to adopt secure coding practices, vulnerability counts continue to rise as seen in Fig. 3, which reflects annual totals from the NIST National Vulnerability Database (NVD). Since APIs have become a cornerstone of modern application development, ensuring API security is essential. More information on API security can be found in resources such as [26], [13], and [27].

Adopt privacy by design. Be mindful of applicable privacy regulations and adopt privacy by design as part of the full lifecycle security and privacy practices. Avoid the use of personally identifiable information in protocol messages or permanent storage to the fullest extent possible.

Design protocols to support access proxies for externalized applications and services. Access proxies can add a level of indirection and provide features such as enforcement of access controls, traffic anomaly detection, denial-of-service protection, and more. A good example of this can be seen with Google's BeyondCorp implementation [28].

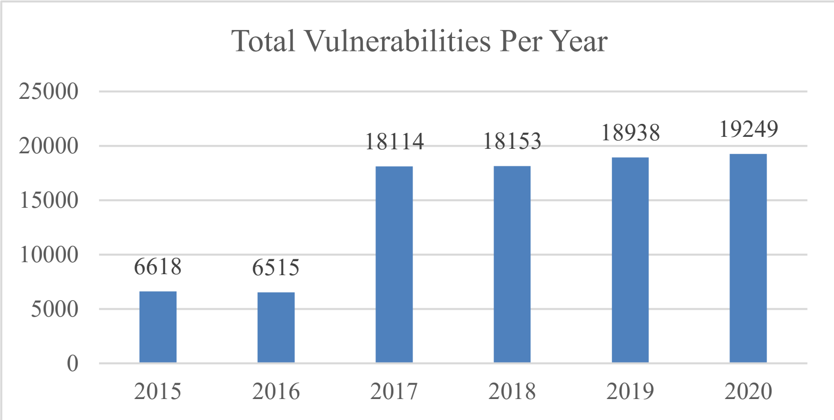

Fig. 3. Total number of vulnerabilities per year recorded in the NIST National Vulnerability Database (https://nvd.nist.gov).

4 Enhancing Zero Trust by Design with Zero Trust Patterns

4.1 Zero Trust Patterns – A Key Enabler for Practitioners

The 23 software design patterns from [29] have had a remarkable impact on software engineering through the re-use of effective solutions to commonly recurring problems. That success has fostered the identification of additional software design patterns, as well as patterns in other areas like security [30]. Beyond establishment of the ZTBD principles and a growing set of good ZT practices, current efforts include further aid to practitioners in the form of ZT patterns. A few examples of such ZT patterns are given here, while many more will continue to be added to the growing ZT body of knowledge (ZTBOK) accompanying ZTBD.

Zero Trust Software Design Patterns. This category of ZT patterns facilitate software engineering success with identified solutions to common problems in software system design. The ZTBOK includes a growing set of patterns in areas such as API security, authentication and authorization, effective multi-level logging, and more. For instance, consider the following example.

Pattern Title: Application Layer for Confidentiality

Problem: I need to ensure confidentiality of information exchanged.

Solution: Use application layer encryption for end-to-end confidentiality guarantees.

Details: Applications should not rely on confidentiality mechanisms at lower layers. While approaches such as link-level point-to-point encryption and VPNs can afford some advantages, they can leave traffic vulnerable outside of the encryption context. Use encryption between endpoints at the application layer for confidentiality. Also, be sure to use strong cipher suites.

Zero Trust Networking Patterns. Similarly, the ZTBOK includes a growing list of patterns targeting success in ZT networks. The following example represents a pattern in this category.

Pattern Title: Prefer a Private PKI

Problem: I need to choose the right PKI for my certificate based ZT authentication schemes.

Solution: Prefer a private PKI system over public PKI entities.

Details: While public PKIs are a good fit for many situations, the implementation of ZT networks tends to result in a large growth in the number of certificates and private PKIs generally have more benefits and fewer risks. Private PKIs can offer clear cost advantages and more control over the certificate creation and rotation processes as well as more flexible automation opportunities. Moreover, with the very large number of public PKI systems, it brings into question the extent of the trust that can confidently be afforded to third parties. For more on the public vs. private PKI discussion for ZT networks, refer to [11].

Zero Trust Protocol Patterns. The third category of patterns in the ZTBOK assist ZT protocol designers, although some could also help with ZT domains beyond protocol design. The following examples represent two sample ZT protocol design patterns.

Pattern Title: Hybrid Post-Quantum Confidentiality

Problem: My protocol uses key exchange algorithms like RSA and ECDH that are vulnerable to compromise by quantum computing enabled adversaries with Shor's algorithm.

Solution: Use post-quantum cryptography. But full confidence cannot be placed in candidate quantum-resistant cryptographic algorithms prior to standardization, or possibly for some period of time thereafter. Combine classical and quantum-resistant algorithms in a hybrid scheme that leverages both approaches to retain the soundness of the strongest approach in the event that the other is weakened or compromised.

Details: NIST in the USA has invited global cooperation for development, testing, and standardization of quantum-resistant public-key cryptographic algorithms. At the time of writing, draft standards were planned for the 2022-2024 timeframe [31]. In the meantime, hybrid post-quantum protocols are being developed for research, testing, and preparation for security against quantum enabled adversaries. The post-quantum hybrid Transport Layer Security (TLS) [32][33], a hybrid post-quantum protocol for fair and privacy-enhanced matchmaking [34], and X.509 compliant hybrid certificates [35] are three examples reflecting the main idea embodied by this pattern.

Pattern Title: Single Packet Authorization (Alt: First / Single Packet Authentication)

Problem: How can I limit my service to authorized packet flows only as part of my zero trust initiatives and thereby reduce my attack surface?

Solution: Use Single Packet Authorization (SPA) where the payload of the first packet includes authentication/authorization information. Packets are passively monitored and dropped by default if not appropriately pre-authenticated as an authorized flow.

Details: SPA packets are encrypted and typically include fields such as username, date/time stamp, protocol version, message type, access request related information, random data to prevent replay attacks, and a message digest. One example of practical SPA tools is fwknop [36][37]. Although SPA was distinguished from port knocking and analyzed quite some time ago [38], attempts to further improve SPA [39][40][41] and to integrate first packet authorization with ZT protocols and cloud networks such as [17] continue. Note in the title and aliases that there is a clear difference between authentication and authorization, but that topic is beyond the scope of this paper and is addressed at [36].

5 Conclusion and Future Work

5.1 Fostering the Next Generation Zero Trust Ecosystem

ZT is not a security appliance to buy or a one-time project. ZT is a realistic security model with a new mindset appropriate for the modern threat landscape with borderless networks, insider threats, remote workers, cloud applications, and advanced persistent threats. In this paper, we aggregated and harmonized guidance from disparate ZT resources, distilling them down in the form of the Zero Trust by Design (ZTBD) core principles. We further provided a set of good practices to foster successful ZT initiatives and extended ZT concepts beyond the network architecture to encompass ZT protocol design and ZT software engineering, which are separate but complementary domains that further facilitate ZT success. The sample ZT patterns represent additional enablers accompanying ZTBD v1.0. Given modern security threats and perimeter-less networks, adoption of the ZT model continues to accelerate.

5.2 An Open Invitation to the Community

Lastly, we extend an open invitation to the community to visit ZeroTrustByDesign.com and contribute to the growing ZTBD body of knowledge. The site also includes links to many useful ZT references to facilitate further ZT learning. We invite feedback on the core principles, additional good practices, and supplementary ZT patterns as we evolve toward ZTBD v2.0 and work together to foster the next generation of the zero trust ecosystem.

References

1. Peisert, S., Schneier, B., Okhravi, H., Massacci, F., Benzel, T., Landwehr, C., Mannan, M., Mirkovic, J., Prakash, A. and Michael, J.B.: Perspectives on the SolarWinds incident. IEEE Security & Privacy, 19(02), pp. 7-13 (2021).
2. Rose, S., Borchert, O., Mitchell, S., Connelly, S.: Zero trust architecture. NIST Special Publication 800-207, National Institute of Standards and Technology, U.S. Department of Commerce (2020).
3. DevSecOps and zero trust architecture (ZTA) for multi-cloud environments, January 27, 2021, https://www.nist.gov/news-events/events/2021/01/devsecops-and-zero-trust-architecture-zta-multi-cloud-environments, last accessed 2021/04/24.
4. Embracing a zero trust security model, United States (US) National Security Agency (NSA), February 25, 2021, https://media.defense.gov/2021/Feb/25/2002588479/-1/-1/0/CSI_EMBRACING_ZT_SECURITY_MODEL_UOO115131-21.PDF, last accessed 2021/04/24.
5. McGillicuddy, S.: Enterprise Zero Trust Networking Strategies: Secure Remote Access and Network Segmentation, Enterprise Management Associates, Inc., pp. 1-32 (2020).
6. McGraw, G.: Software Security: Building security in. 1st edn. Addison-Wesley Professional (2006).
7. Cho, J. H., Chan, K., and Adali, S.: A survey on trust modeling. ACM Computing Surveys (CSUR), *48*(2), pp. 1-40 (2015).
8. Sood, A.K., Huang, Y., Simon, R., White, E., and Cleary, K.: Zero trust intrusion containment for telemedicine, George Mason University, Fairfax, VA, USA (2002).
9. Kindervag, J.: Build security into your network's DNA: The zero trust network architecture. Forrester Research Inc., pp. 1-26 (2010).
10. Kindervag, J. and Balaouras, S.: No more chewy centers: Introducing the zero trust model of information security. Forrester Research, pp. 1-13 (2010).
11. Gilman, E. and Barth, D.: Zero Trust Networks, O'Reilly Media, Incorporated (2017).
12. Kerman, A., Borchert, O., Rose, S., and Tan, A.: Implementing a zero trust architecture, The MITRE Corporation, Tech. Rep. (2020).
13. O’Neill, M., Zumerle, D., and D’Hoinne, J.: How to build an effective API security strategy. Gartner, Inc. (2017).
14. State of API Security Q1 2021. Salt Security, 2021, https://salt.security/api-security-trends, last accessed 2021/04/26.
15. Eidle, D., Ni, S. Y., DeCusatis, C., and Sager, A.: Autonomic security for zero trust networks. In 2017 IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON), pp. 288-293, IEEE (2017).
16. Vanickis, R., Jacob, P., Dehghanzadeh, S., and Lee, B.: Access control policy enforcement for zero-trust-networking. In 2018 29th Irish Signals and Systems Conference (ISSC), pp. 1-6, IEEE (2018).

17. DeCusatis, C., Liengtiraphan, P., Sager, A., and Pinelli, M.: Implementing zero trust cloud networks with transport access control and first packet authentication. In 2016 IEEE International Conference on Smart Cloud (SmartCloud), pp. 5-10, IEEE (2016).
18. Nakamoto, S.: Bitcoin: A peer-to-peer electronic cash system, https://bitcoin.org/bitcoin.pdf (2008).
19. Noether, S.: Ring signature confidential transactions for monero, IACR Cryptol. ePrint Arch., 1098 (2015).
20. Shin, J. S. and Gligor, V. D.: A new privacy-enhanced matchmaking protocol, IEICE Trans. Commun, 96(8), pp. 2049-2059 (2013).
21. Horne, D. and Nair, S.: A new privacy-enhanced technology for fair matchmaking with identity linked wishes, IEEE Systems Journal, 14(1), pp. 298-309 (2019).
22. Samaniego, M. and Deters, R.: Zero-trust hierarchical management in IoT. In 2018 IEEE international congress on Internet of Things (ICIOT), pp. 88-95, IEEE (2018).
23. Klein, D.: Micro-segmentation: securing complex cloud environments, Network Security, 2019(3), pp. 6-10 (2019).
24. Moubayed, A., Refaey, A., and Shami, A.: Software-defined perimeter (sdp): State of the art secure solution for modern networks, IEEE Network, 33(5), pp. 226-233 (2019).
25. Mämmelä, O., Hiltunen, J., Suomalainen, J., Ahola, K., Mannersalo, P., and Vehkaperä, J.: Towards micro-segmentation in 5G network security. In European Conference on Networks and Communications (EuCNC 2016) Workshop on Network Management, Quality of Service and Security for 5G Networks (2016).
26. Siriwardena, P.: Advanced API security: OAuth 2.0 and beyond, second edition, Apress (2020).
27. OWASP API Security - Top 10 | OWASP, https://owasp.org/www-project-api-security/, last accessed 2021/04/29.
28. Ward, R., and Beyer, B.: Beyondcorp: a new approach to enterprise security, ;login, USENIX (2014).
29. Gamma, E., Helm, R., Johnson, R., and Vlissides, J.: Design patterns: abstraction and reuse of object-oriented design. In European Conference on Object-Oriented Programming, pp. 406-431, Springer, Berlin, Heidelberg (1993).
30. Schumacher, M., Fernandez-Buglioni, E., Hybertson, D., Buschmann, F., and Sommerlad, P.: Security Patterns: Integrating security and systems engineering, John Wiley & Sons (2013).
31. Post-Quantum Cryptography | CSRC, National Institute of Standards and Technology (NIST), United States Department of Commerce, https://csrc.nist.gov/Projects/post-quantum-cryptography/workshops-and-timeline, last accessed 2021/04/29.
32. Crockett, E., Paquin, C., and Stebila, D.: Prototyping post-quantum and hybrid key exchange and authentication in TLS and SSH. *IACR Cryptol. ePrint Arch.,* 858 (2019).
33. Paquin, C., Stebila, D., and Tamvada, G.: Benchmarking post-quantum cryptography in TLS. In *International Conference on Post-Quantum Cryptography* (pp. 72-91). Springer, Cham (2020).
34. Horne, D. and Nair, S.: Toward a quantum resistant secretmatch privacy enhanced technology–a case study. In Proceedings of the International Conference on Security and Management (SAM), pp. 147-153 (2019).
35. Bindel, N., Braun, J., Gladiator, L., Stöckert, T., and Wirth, J.: X.509-compliant hybrid certificates for the post-quantum transition, Journal of Open Source Software, 4(40), 1606 (2019).
36. Rash, M.: Single packet authorization with fwknop, https://www.cipherdyne.org/fwknop/docs/SPA.html, last accessed 2021/04/29.
37. Rash, M.: Single packet authorization – a comprehensive guide to strong service concealment with fwknop, https://www.cipherdyne.org/fwknop/docs/fwknop-tutorial.html#design, last accessed 2021/04/29.
38. Jeanquier, S.: An analysis of port knocking and single packet authorization, MSc Thesis, University of London (2006).
39. Srivastava, V., Keshri, A. K., Roy, A. D., Chaurasiya, V. K., and Gupta, R.: Advanced port knocking authentication scheme with QRC using AES, In 2011 International Conference on Emerging Trends in Networks and Computer Communications (ETNCC), pp. 159-163, IEEE (2011).
40. Mehran, P., Reza, E. A., and Laleh, B.: SPKT: secure port knock-tunneling, an enhanced port security authentication mechanism. In 2012 IEEE Symposium on Computers & Informatics (ISCI), pp. 145-149, IEEE (2012).
41. Lucion, E. L. R. and Nunes, R. C.: Software defined perimeter: improvements in the security of single packet authorization and user authentication. In 2018 XLIV Latin American Computer Conference (CLEI), pp. 708-717, IEEE (2018).

PROJECT DESCRIPTION

IMPLEMENTING A ZERO TRUST ARCHITECTURE

Alper Kerman

National Cybersecurity Center of Excellence
National Institute of Standards and Technology

Oliver Borchert
Scott Rose

Advanced Network Technologies Division
National Institute of Standards and Technology

Eileen Division
Allen Tan

The MITRE Corporation

October 2020

nccoe-zta-project@list.nist.gov

This revision incorporates comments from the public.

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

NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

The National Cybersecurity Center of Excellence (NCCoE), a part of the National Institute of Standards and Technology (NIST), is a collaborative hub where industry organizations, government agencies, and academic institutions work together to address businesses' most pressing cybersecurity challenges. Through this collaboration, the NCCoE develops modular, adaptable example cybersecurity solutions demonstrating how to apply standards and best practices by using commercially available technology. To learn more about the NCCoE, visit https://www.nccoe.nist.gov. To learn more about NIST, visit https://www.nist.gov.

This document describes a challenge that is relevant to many industry sectors. NCCoE cybersecurity experts will address this challenge through collaboration with a Community of Interest, including vendors of cybersecurity solutions. The resulting reference design will detail one or more approaches that can be incorporated across multiple industry sectors.

Abstract

The proliferation of cloud computing, mobile device use, and the Internet of Things has dissolved conventional network boundaries. The workforce is more distributed, with remote workers who need access to resources anytime, anywhere, and on any device, to support the mission. Enterprises must evolve to provide secure access to company resources from any location and asset, protect interactions with business partners, and shield client-server as well as inter-server communications.

A zero trust cybersecurity approach removes the assumption of trust typically given to devices, subjects (i.e., the people and things that request information from resources), and networks. It focuses on accessing resources in a secure manner, regardless of network location, subject, and asset, and enforcing risk-based access controls while continually inspecting, monitoring, and logging interactions. This requires device health attestation, data-level protections, a robust identity architecture, and strategic micro-segmentation to create granular trust zones around an organization's digital resources. Zero trust evaluates access requests and communication behaviors in real time over the length of open connections, while continually and consistently recalibrating access to the organization's resources. Designing for zero trust enables enterprises to securely accommodate the complexity of a diverse set of business cases by informing virtually all access decisions and interactions between systems and resources.

This NCCoE project will show a standards-based implementation of a zero trust architecture (ZTA). Publication of this project description begins a process that will further identify project requirements and scope, as well as the hardware and software components to develop demonstrations. The NCCoE will build a modular, end-to-end example ZTA(s) using commercially available technology that will address a set of cybersecurity challenges aligned to the NIST Cybersecurity Framework. This project will result in a freely available NIST Cybersecurity Practice Guide.

Keywords

cybersecurity; enterprise; identity and access management; network security; remote access; zero trust; zero trust architecture

Disclaimer

Certain commercial entities, equipment, products, 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 or NCCoE, nor is it intended to imply that the entities, equipment, products, or materials are necessarily the best available for the purpose.

Table of Contents

1 Executive Summary

Purpose

Conventional network security has focused on perimeter defenses—once inside the network perimeter, subjects (i.e., end users, applications, and other non-person entities that request information from resources) are often given broad access to multiple corporate resources. If the subjects are compromised, malicious actors—through impersonation and escalation—can gain access to the resources from inside or outside the network. Moreover, the growth in cloud computing, Internet of Things (IoT), business partners, and the growing number of remote workers raises the complexity of protecting an organization's digital resources, because more points of entry, exit, and data access exist than ever before.

Organizations are rethinking the conventional network security perimeter. A zero trust architecture (ZTA) addresses this trend by focusing on protecting resources, not network perimeters, as the network location is no longer viewed as the prime component to the security posture necessary for a resource.

Zero trust is a set of cybersecurity principles used to create a strategy that focuses on moving network defenses from wide, static network perimeters to focusing more narrowly on subjects, enterprise assets (i.e., devices, infrastructure components, applications, virtual and cloud components), and individual or small groups of resources. A ZTA uses zero trust principles to plan and protect an enterprise infrastructure and workflows. By design, a ZTA environment embraces the notion of no implicit trust toward assets and subjects, regardless of their physical or network locations (i.e., local area networks versus the internet). Hence, a ZTA never grants access to resources until a subject, asset, or workload are verified by reliable authentication and authorization.

This document defines a National Cybersecurity Center of Excellence (NCCoE) project to help organizations design for zero trust. This project will produce an example implementation(s) of a ZTA, using commercially available technology designed and deployed according to the concepts and tenets documented in National Institute of Standards and Technology (NIST) Special Publication (SP) 800-207, *Zero Trust Architecture* [1]. The primary objective of this project is to demonstrate a proposed architecture(s) that brings into play different enterprise resources (e.g., data sources, computing services, and IoT devices) that are spread across on-premises and cloud environments that inherit the ZTA solution characteristics outlined in NIST SP 800-207.

Another objective of this project is to document the impacts on administrator and end-user experience because of employing a ZTA strategy.

This project will result in a publicly available NIST Cybersecurity Practice Guide, a detailed implementation guide of the practical steps needed to implement a cybersecurity reference design that addresses the project goals.

Scope

The scope of this project is limited to implementing a ZTA for a conventional, general purpose enterprise information technology (IT) infrastructure that combines users (including employees, contractors, guests, and non-person entities); assets; and enterprise resources. Resources could be hosted and managed—by the corporation itself or a third-party provider—on premise, in the cloud, at the edge, or some combination of these. There may also be branch or partner offices, teleworkers, and bring-your-own-device (BYOD) usage.

This project will focus primarily on access to enterprise resources. More specifically, the focus will be on behaviors of enterprise employees, contractors, and guests accessing enterprise resources while connected from the corporate (or enterprise headquarters) network, a branch office, or the public internet. Access requests can occur over both the enterprise-owned part of the infrastructure as well as the public/non-enterprise-owned part of the infrastructure. This requires that all access requests be secure, authorized, and verified before access is enforced, regardless of where the request is initiated or where the resources are located.

ZTAs for industrial control systems and operational technology (OT) environments are explicitly out of scope for this project. However, the project seeks to provide an approach and security principles for a ZTA that could potentially be extended to OT environments.

Challenges

Many organizations are looking to build for zero trust, but challenges exist. Current challenges to implementing a ZTA include:

- Maturity of vendor products to support a ZTA.
- Organization's ability/willingness to migrate to a ZTA because of:
  - heavy investment in other (legacy) technologies
  - absence of, or deficiency in, identity governance
  - lack of ability/resources to develop a transition plan, pilot, or proof of concept
- Security concerns such as:
  - compromise of the zero trust control plane
  - ability to recognize attacks and detect malicious insiders
- Interoperability considerations of ZTA products/solutions with legacy technologies such as:
  - standard versus proprietary interfaces
  - ability to interact with enterprise and cloud services
- User experience. To date, there has been no detailed examination of how a ZTA would or could affect end-user experience and behavior. The goal of a ZTA should be to enhance security in a way that is transparent to the end user.

This practice guide aims to mitigate these challenges, using the solutions and collaborators selected for the demonstration project.

Background

Historically, the perimeter-based network security model has been the dominant model for information security. It assumes users inside the corporate network perimeter are "trusted" and anyone on the outside is "untrusted." For several decades, this view of trust has served as the basis for determining what resources a subject/asset can access.

Several high-profile cyber attacks in recent years, including the Office of Personnel Management breach in 2015, have undermined the case for the perimeter-based model [2]. Moreover, the perimeter is becoming less relevant due to several factors, including the growth of cloud computing, mobility, and changes in the modern workforce. It is with this backdrop that the Federal Chief Information Officer (CIO) Council [3] engaged the NIST NCCoE in 2018 to help federal agencies coalesce around a definition for ZTA and understand the benefits and limitations of a ZTA. The interagency collaboration resulted in publication of NIST SP 800-207, *Zero Trust Architecture*.

This NCCoE project description builds on the work with federal agencies and the Federal CIO Council as we seek to build and document one or more demonstrable ZTAs, using commercially available products that align to the concepts and principles in NIST SP 800-207.

2 Scenarios

Responses from industry organizations that express interest in taking part in this project will affect the potential scenario-set in terms of the composition and number of scenarios demonstrated. These scenarios encapsulate the notion of providing subjects access to corporate resources hosted on premise or in the cloud. Access requests may come from within the enterprise network or the public internet, in the case of teleworkers. It is assumed the enterprise is implementing a ZTA within an existing typical corporate environment.

Scenario 1: Employee Access to Corporate Resources

An employee is looking for easy and secure access to corporate resources, from any work location. This scenario will demonstrate a specific user experience where an employee attempts to access corporate services such as the corporate intranet, a time-and-attendance system, and other human resources systems by using either an enterprise-managed device or a personally owned device. The ZTA solution implemented in this project will enforce the associated access request, dynamically and in near real-time. The employee will be able to perform the following:

- Access on-premise corporate resources while connected from the corporate intranet.
- Access corporate resources in the cloud while connected directly from the corporate intranet.
- Access on-premise corporate resources while connected from a branch office.
- Access corporate resources in the cloud while connected from a branch office.
- Access on-premise corporate resources from the public internet while teleworking.
- Access corporate resources in the cloud from the public internet while teleworking.

Scenario 2: Employee Access to Internet Resources

An employee is trying to access the public internet to accomplish some tasks. This scenario will show a specific user experience where an employee attempts to access an enterprise-sanctioned, web-based service on the internet by using an enterprise-managed device. Although the web-based service is not owned and managed by the enterprise, the associated access request for that resource will still be enforced, dynamically and in real time, by a ZTA solution implemented in this project. The solution will manage the employee's access, regardless of location. That is, the employee can access the internet while connected inside the corporate intranet, a branch office, or the public internet by using an enterprise-managed device.

If an employee is allowed by corporate policy to access non-enterprise-managed resources and services in the public internet by using enterprise-managed devices, the ZTA solution will allow the enterprise to determine the extent of this access.

Examples of access restrictions in the above paragraph could include:

- Access to social media sites is not sanctioned.

- Access to an internet search engine is permitted, and the associated access request for this resource does not need to be granted in real time through the corporate network when an employee is working at a branch office or while teleworking (e.g., coffee shop or airport).
- Mission-critical services on the public internet (e.g., GitHub) can be accessed directly by the employee.

Scenario 3: Contractor Access to Corporate and Internet Resources

A contractor is trying to access certain corporate resources and the internet. This scenario will show a specific user experience where a contractor attempts to access certain corporate resources and the internet to perform the planned service for the organization. The corporate resources can be on premise or in the cloud, and the contractor will be able to access corporate resources while on premise or from the public internet, using an enterprise-managed device given to the contractor, a contractor-owned and managed device, or a BYOD scenario. The ZTA solution implemented in this project will enforce, dynamically and in near real time, the associated access requests for resources by the contractor.

Scenario 4: Inter-server Communication Within the Enterprise

Corporate services often have different servers communicating with each other. For example, a web server communicates with an application server. The application server communicates with a database to retrieve data back to the web server. This scenario will demonstrate examples of inter-server interactions within the enterprise, which will include servers that are on premise, in the cloud, or between servers that are on premise and in the cloud. The ZTA solution implemented in this project will enforce, dynamically and in near real time, the associated network communications among designated servers that interact with one another.

Scenario 5: Cross-Enterprise Collaboration with Business Partners

Two enterprises (Enterprise A and Enterprise B) may collaborate on a project where resources are shared. In this scenario, the ZTA solution implemented in this project will enable users from one enterprise to securely access specific resources from the other enterprise, and vice versa. For example, Enterprise A users will be able to access a specific application from Enterprise B, while Enterprise B users will be able to access a specific database from Enterprise A.

Scenario 6: Develop Trust Score/Confidence Level with Corporate Resources

Enterprises have monitoring systems, security information and event management (SIEM) systems, and other resources that can provide data to support security analytics to a policy engine to create a more granular trust score/confidence level for access to corporate resources and promote strict access based on the confidence level. In this scenario, a ZTA solution will integrate these monitoring and SIEM systems with the policy engine to produce more precise calculation of trust scores/confidence levels in near real time.

Note: The scenarios above may be created and demonstrated in different phases throughout the project.

3 High-Level Architecture

Figure 1 illustrates a high-level, notional architecture of the logical and functional components that could make up a ZTA for a typical IT enterprise.

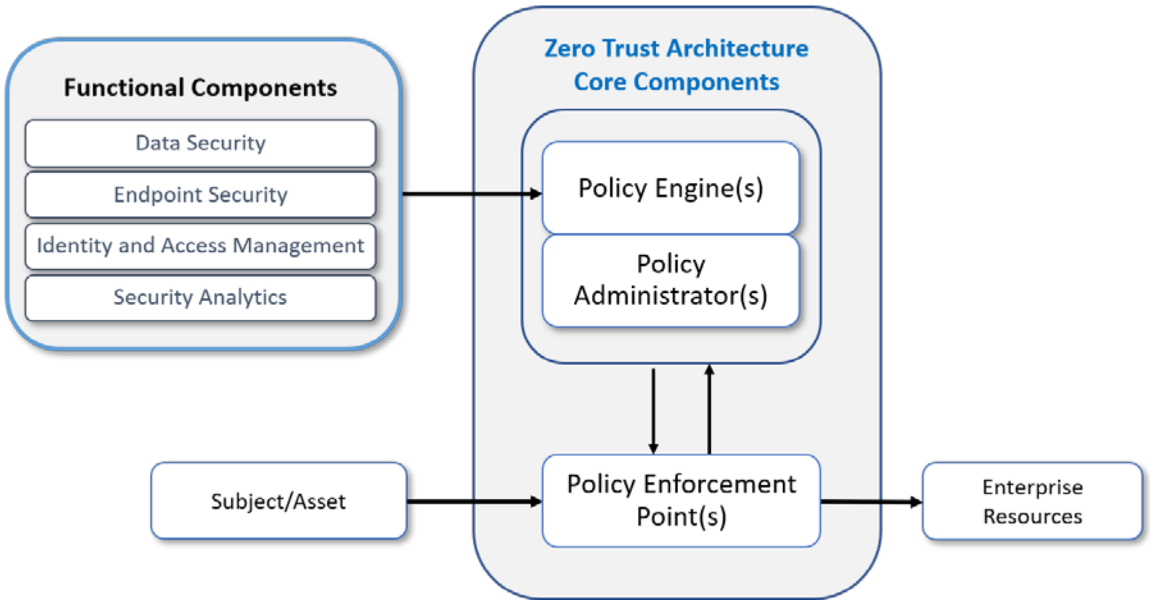

Figure 1. ZTA High-Level Architecture

Component List

The technical components required of the ZTA solution(s) for this project include but are not limited to:

Core Components:

- The policy engine handles the ultimate decision to grant, deny, or revoke access to a resource for a given subject. The policy engine calculates the trust scores/confidence levels and ultimate access decisions.
- The policy administrator is responsible for establishing/terminating the transaction between a subject and a resource. It generates any session-specific authentication and 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 a session.
- The policy enforcement point handles enabling, monitoring, and eventually terminating connections between a subject and an enterprise resource.

Functional Components:

- The data security component includes all the data access policies and rules that an enterprise develops to secure its information, and the means to protect data at rest and in transit.
- The endpoint security component encompasses the strategy, technology, and governance to protect endpoints (e.g., servers, desktops, mobile phones, IoT devices) from threats and attacks, as well as protect the enterprise from threats from managed and unmanaged devices.
- The identity and access management component includes the strategy, technology, and governance for creating, storing, and managing enterprise user (i.e., subject) accounts and identity records and their access to enterprise resources.

- The security analytics component encompasses all the threat intelligence feeds and traffic/activity monitoring for an IT enterprise. It gathers security and behavior analytics about the current state of enterprise assets and continuously monitors those assets to actively respond to threats or malicious activity. This information could feed the policy engine to help make dynamic access decisions.

Devices and Network Infrastructure Components:

- Assets include the devices/endpoints, such as laptops, tablets, and other mobile or IoT devices, that connect to the enterprise.
- Enterprise resources include data and compute resources as well as applications/services hosted and managed on premise, in the cloud, at the edge, or some combination of these.
- Network infrastructure components encompass network resources a medium or large enterprise might typically deploy in its environment. It is assumed that the ZTA core and functional components and devices are connected via, or integrated into, the network infrastructure. Note: The network infrastructure is not depicted in Figure 1. The NCCoE will provide these components as part of its internal lab infrastructure.

Desired Security Characteristics and Properties

This project seeks to develop a reference design and implementation, using commercially available technology that meets the following characteristics:

- All interactions throughout the proposed architecture are achieved in the most secure manner available, with emphasis on protecting confidentiality and integrity through a consistent identification, authentication, and authorization scheme.
- All interactions throughout the proposed architecture are continually reassessed with possible reauthentication and reauthorization as necessary to mitigate unauthorized access to enterprise resources.
- Access to an enterprise resource is assessed on a per-session basis and authorized specifically for that enterprise resource.
- Access requests are evaluated dynamically based on organizational policies and rules for accessing enterprise resources, including the observable state of:
    - subject identity (e.g., user account or service identity with associated attributes)
    - requesting asset (e.g., laptop, mobile device, server) device characteristics such as the software version installed, security posture, network location, time/date of request, previously observed behavior, and installed credentials
    - requested resource (e.g., server, application, service) characteristics
- Enterprise assets and resources are continuously monitored and reassessed to maintain them in their most secure states possible.
- Log and event data generated about the current state of enterprise assets, resources, and interactions throughout the proposed architecture are collected and leveraged for better policy alignment and enforcement to increase the enterprise's overall security posture.
- Secure access to corporate resources, hosted either on premise or within a cloud environment, as well as to non-corporate resources on the internet are provided

without the use of conventional network and network perimeter access and security solutions.

- Integration with various directory protocols and identity management services (e.g., Lightweight Directory Access Protocol [LDAP], OAuth 2.0, Active Directory, OpenLDAP, Security Assertion Markup Language) is demonstrated.
- Integration with SIEM tools through common application programming interfaces is demonstrated.
- Desired enterprise device security characteristics are demonstrated, including:
  - maintaining data protection at rest and in transit
  - remediating device vulnerabilities that could result in unauthorized access to data stored on or accessed by the device, and misuse of the device
  - mitigating malware execution on the device that could result in unauthorized access to data stored on or accessed by the device, and misuse of the device
  - mitigating the risk of data loss through accidental, deliberate, or malicious deletion or obfuscation of data stored on the device
  - maintaining awareness of and responding to suspicious or malicious activities within and against the device to prevent or detect a compromise of the device

4 Relevant Standards and Guidance

The references, standards, and guidelines that apply to this project are listed below.

- NIST Cybersecurity Framework v.1.1, Framework for Improving Critical Infrastructure Cybersecurity
  https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.04162018.pdf
- NIST SP 800-30 Revision 1, Guide for Conducting Risk Assessments
  https://doi.org/10.6028/NIST.SP.800-30r1
- NIST SP 800-37 Revision 2, Risk Management Framework for Information Systems and Organizations: A System Life Cycle Approach For Security and Privacy
  https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-37r2.pdf
- NIST SP 800-40 Revision 3, Guide to Enterprise Patch Management Technologies
  https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-40r3.pdf
- NIST SP 800-46 Revision 2, Guide to Enterprise Telework, Remote Access, and Bring Your Own Device (BYOD) Security
  https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-46r2.pdf
- NIST SP 800-53 Revision 4, Security and Privacy Controls for Federal Information Systems and Organizations
  https://csrc.nist.gov/csrc/media/publications/sp/800-53/rev-4/archive/2013-04-30/documents/sp800-53-rev4-ipd.pdf
- NIST SP 800-57 Part 1 Revision 4, Recommendation for Key Management: Part 1: General
  https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-57pt1r4.pdf
- NIST SP 800-61 Revision 2, Computer Security Incident Handling Guide
  http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-61r2.pdf

- NIST SP 800-63 Revision 3, Digital Identity Guidelines
https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-63-3.pdf
- NIST SP 800-92, Guide to Computer Security Log Management
http://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-92.pdf
- NIST SP 800-114 Revision 1, User's Guide to Telework and Bring Your Own Device (BYOD) Security
https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-114r1.pdf
- NIST SP 800-122, Guide to Protecting the Confidentiality of Personally Identifiable Information (PII)
https://nvlpubs.nist.gov/nistpubs/legacy/sp/nistspecialpublication800-122.pdf
- NIST SP 800-124 Revision 2 (Draft), Guidelines for Managing the Security of Mobile Devices in the Enterprise
https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-124r2-draft.pdf
- NIST SP 800-160 Vol. 2, Developing Cyber Resilient Systems: A Systems Security Engineering Approach
https://csrc.nist.gov/publications/detail/sp/800-160/vol-2/final
- NIST SP 800-162, Guide to Attribute Based Access Control (ABAC) Definition and Considerations
https://nvlpubs.nist.gov/nistpubs/specialpublications/NIST.sp.800-162.pdf
- NIST SP 800-175B, Guideline for Using Cryptographic Standards in the Federal Government: Cryptographic Mechanisms
https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-175b.pdf
- NIST SP 800-171 Revision 2, Protecting Controlled Unclassified Information in Nonfederal Information Systems and Organizations
https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-171r2.pdf
- NIST SP 800-205, Attribute Considerations for Access Control Systems
https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-205.pdf
- NIST SP 800-207 (Second Draft), Zero Trust Architecture
https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-207-draft2.pdf
- NIST SP 1800-3, Attribute Based Access Control
https://www.nccoe.nist.gov/sites/default/files/library/sp1800/abac-nist-sp1800-3-draft-v2.pdf
- Cloud Security Alliance, Software Defined Perimeter Working Group, SDP Specification 1.0
https://downloads.cloudsecurityalliance.org/initiatives/sdp/SDP_Specification_1.0.pdf
- ISO/IEC 27001, Information Technology–Security Techniques–Information Security Management Systems
- American Council for Technology-Industry Advisory Council, Zero Trust Cybersecurity Current Trends
https://www.actiac.org/system/files/ACT-IAC%20Zero%20Trust%20Project%20Report%2004182019.pdf
- Federal Information Processing Standards 140-3, Security Requirements for Cryptographic Modules
https://nvlpubs.nist.gov/nistpubs/FIPS/NIST.FIPS.140-3.pdf

5 Security Control Map

This table maps the characteristics of the commercial products that the NCCoE will apply to this cybersecurity challenge to the applicable standards and recommended practices described in the *Framework for Improving Critical Infrastructure Cybersecurity* and to other NIST guidance. This information represents an approach to document the applicability of standards, guidelines, and recommended practices to the security characteristics of the solution, but it does not imply that products and services will meet an industry's requirements for regulatory approval or accreditation.

Table 1: Security Control Map

Cybersecurity Framework v1.1		
Function	**Category**	**Subcategory**
IDENTIFY (ID)	Asset Management (ID.AM)	ID.AM-1: Physical devices and systems within the organization are inventoried.
		ID.AM-2: Software platforms and applications within the organization are inventoried.
		ID.AM-5: Resources (e.g., hardware, devices, data, time, personnel, and software) are prioritized based on their classification, criticality, and business value.
	Risk Assessment (ID.RA)	ID.RA-1: Asset vulnerabilities are identified and documented.
		ID.RA-3: Threats, both internal and external, are identified and documented.
PROTECT (PR)	Identity Management, Authentication, and Access Control (PR.AC)	PR.AC-1 Identities and credentials are issued, managed, verified, revoked, and audited for authorized devices, users, and processes.
		PR.AC-3 Remote access is managed.
		PR.AC-4 Access permissions and authorizations are managed, incorporating the principles of least privilege and separation of duties.
		PR.AC-5: Network integrity is protected (e.g., network segregation, network segmentation).
		PR.AC-6: Identities are proofed and bound to credentials and asserted in interactions.

Cybersecurity Framework v1.1		
Function	Category	Subcategory
		PR.AC-7 Users, devices, and other assets are authenticated (e.g., single-factor, multifactor) commensurate with the risk of the transaction (e.g., individuals' security and privacy risks and other organizational risks).
	Data Security (PR.DS)	PR.DS-2 Data in transit is protected.
		PR.DS-5: Protections against data leaks are implemented.
		PR.DS-6 Integrity-checking mechanisms are used to verify software, firmware, and information integrity.
		PR.DS-8: Integrity-checking mechanisms are used to verify hardware integrity.
	Information Protection Processes and Procedures (PR.IP)	PR.IP-1: A baseline configuration of IT/industrial control systems is created and maintained, incorporating security principles (e.g., concept of least functionality).
		PR.IP-3: Configuration change control processes are in place.
	Protective Technology (PR.PT)	PR.PT-3 The principle of least functionality is incorporated by configuring systems to provide only essential capabilities.
		PR.PT-4: Communications and control networks are protected.

Cybersecurity Framework v1.1		
Function	**Category**	**Subcategory**
DETECT	Anomalies and Events (DE.AE)	DE.AE-2: Detected events are analyzed to understand attack targets and methods.
		DE.AE-3: Event data are collected and correlated from multiple sources and sensors.
		DE.AE-5: Incident alert thresholds are established.
	Security Continuous Monitoring (DE.CM)	DE.CM-1: The network is monitored to detect potential cybersecurity events.
		DE.CM-2: The physical environment is monitored to detect potential cybersecurity events.
		DE.CM-4: Malicious code is detected.
		DE.CM-5: Unauthorized mobile code is detected.
		DE.CM-6: External service provider activity is monitored to detect potential cybersecurity events.
		DE.CM-7 Monitoring for unauthorized personnel, connections, devices, and software is performed.
		DE.CM-8: Vulnerability scans are performed.
	Detection Processes (DE.DP)	DE.DP-5: Detection processes are continuously improved.
RESPOND	Mitigation (RS.MI)	RS.MI-1: Incidents are contained.
		RS.MI-2: Incidents are mitigated.

Appendix A References

[1] S. Rose et al., *Zero Trust Architecture,* National Institute of Standards and Technology (NIST) Special Publication 800-207, Gaithersburg, Md., August 2020. Available: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-207.pdf

[2] J. Chaffettz, *Adopting a zero trust cyber model in government*, Federal News Network, September 19, 2016. Available: https://federalnewsnetwork.com/commentary/2016/09/adopting-zero-trust-cyber-model-government/

[3] The Federal Chief Information Officer Council: https://www.cio.gov/