SANDIA REPORT
SAND20XX-XXXX
Printed Click to enter a date

Sandia National Laboratories

Zero Trust Architectures in Nuclear Control Systems

Benjamin Karch
Andrew Hahn
Alex Haddad
Christopher C. Lamb

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico
87185 and Livermore,
California 94550

Issued by Sandia National Laboratories, operated for the United States Department of Energy by National Technology & Engineering Solutions of Sandia, LLC.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@osti.gov
Online ordering: http://www.osti.gov/scitech

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5301 Shawnee Rd
Alexandria, VA 22312

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.gov
Online order: https://classic.ntis.gov/help/order-methods/

ABSTRACT

In this report, we look at Zero Trust Architecture (ZTA) principles and outline where and which tenets are applicable to nuclear power control systems, both for current generation systems and potential future Small/Modular and Advanced systems. ZTA approaches are becoming more popular in IT systems and are recommended approaches for building new systems. We have also seen some partial ZTA solutions in place for industrial systems, but nothing with the rigor required of nuclear power systems. We first define ZTA and discuss multiple current implementations in IT systems, cloud computing systems, and finally industrial systems. With this context, we then discuss where ZTA techniques can be applied in current and future systems based on current standards and regulatory guidance. We close the report with a summary of technical challenges that need to be addressed for ZTA to be useful, and where in nuclear systems ZTA can have the most impact on system security.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

This page left blank

EXECUTIVE SUMMARY

Sandia National Laboratories developed this report under guidance with funding from the Department of Energy (DOE) Nuclear Engineering (NE) Cybersecurity branch. This report aligns with goals from the DOE NE strategic plan. Overall, the goals of DOE NE cybersecurity efforts are to demonstrate and document cybersecurity approaches to designing and implementing modern control-system architectures within nuclear power systems that enhance cybersecurity at a lower cost than the systems in place today. The overall milestone addressing this work is to develop and document an approach for cybersecurity by design in advanced reactor systems, delivering guides, functional requirements, and implementation guidance that would shorten implementation time and costs for these systems.

This report opens with detailed descriptions of the history of Zero Trust Architecture (ZTA) and current modern guidance on how to develop ZTA systems. We specifically discuss Google's implementations within their IT systems, cloud computing implementations in the New York State, and operational technology (OT) implementations lead by Deloitte consulting in industrial systems. Unsurprisingly, these systems are not the same and implement key functions in different ways. Furthermore, one area that is commonly addressed incorrectly is authentication and authorization.

Current modern guidance defines ZTA as following a group of tenets, as summarized by the National Institute of Standards and Technology (NIST):

1. All data sources and computing services are considered resources
2. All communication is secured regardless of network location
3. Access to individual enterprise resources is granted on a per-session basis
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
5. The enterprise monitors and measures the integrity and security posture of all owned and associated assets
6. All resource authentication and authorization are dynamic and strictly enforced before access is allowed
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

Of these tenets, (3), (4), and (6) are essentially impossible to implement in industrial systems today for system-to-system authentication and authorization. Typical implementation used in IT systems are simply not suitable for highly available, hard real-time control systems like those in nuclear power systems. These current implementations introduce new systems that increase potential latency in unexpected ways and introduce new potential single-point and common cause failure vectors.

ACRONYMS AND TERMS

Acronym/Term	Definition
CISO	Chief Information Security Officer
DHCP	Dynamic Host Configuration Protocol
DiD	Defense in Depth
DOE	Department of Energy
HMI	Human Machine Interface
HR	Human Resources
I&C	Instrumentation and Control
IAEA	International Atomic Energy Agency
IAM	Identity Access Management
ICS	Industrial Control System
IEC	International Electrochemical Commission
IEEE	Institute of Electrical and Electronics Engineers
IoT	Internet of Things
IPSEC	Internet Protocol Security
IT	Information Technology
LAN	Local Area Network
LWR	Light Water Reactors
NE	Nuclear Engineering
NGFW	Next-Generation Firewall
NIST	National Institute of Standards and Technology
NPP	Nuclear Power Plant
NRC	Nuclear Regulatory Commission
OS	Operating System
OT	Operational Technology
PA	Policy Administrator
PDP	Policy Decision Point
PE	Policy Engine
PEP	Policy Enforcement Point
PKI	Public Key Infrastructure
PLC	Programmable Logic Controller
RTOS	Real-Time Operating Systems
SCADA	Supervisory Control and Data Acquisition
SEP	Secure Enclave Processor
SIEM	Security Information and Event Management

Acronym/Term	Definition
SSO	Single Sign-On
TAC	Transport Access Control
TLS	Transport Layer Security
TPM	Trusted Platform Module
VA	Vital Area
VPN	Virtual Private Network
ZTA	Zero Trust Architecture
ZTNA	Zero Trust Network Architecture

1. INTRODUCTION

In the past decade, the world has witnessed an explosion of cyber-attacks on industrial systems and critical infrastructure. Starting in the 2010s, we have had attacks on critical infrastructure ranging from water treatment plants in Florida, to power systems in Ukraine, to industrial furnace facilities in Germany. Due to the deteriorating global political climate today, these attacks show no sign of becoming anything other than more frequent.

We have seen two attacks on or adjacent to nuclear systems. One, in the United States, was malware installed on a laptop in a nuclear facility business local area network (LAN) via spearfishing, watering hole attacks and exploit kits. The other, in India, did not breach control system protections, but did show a deep understanding of the attacked systems and resulted in large amounts of data being successfully exfiltrated from the facility. In fact, the first attack more broadly, and successfully, compromised facilities and personnel across the energy sector in the United States, including at federal agencies. The threat to our energy and nuclear systems is here, today, and likely to become stronger in the coming years.

Furthermore, costs associated with implementing cybersecurity controls is escalating across the nuclear sector. In an energy sector where we depend on carbon-free energy production that operates on razor thin margins, this increase in cybersecurity costs leads to an increase on cost per unit of power generated, making nuclear less competitive with other carbon-emitting energy production methods.

This report is an initial examination of using Zero Trust Architecture (ZTA) techniques to secure nuclear control systems. We briefly examine the history of ZTA to set the context and understand the motivations around creating it, look at the state of ZTA today via a group of case studies, and examine how we could apply ZTA to control systems in nuclear plants. When looking at potential nuclear plant application, we look over the regulatory and standards landscape to see how ZTA needs to adapt to this environment. We also look at key differences between Information Technology (IT) and operational technology (OT) systems that an OT flavor of ZTA would need to accommodate to be applied in nuclear systems.

2. ZERO TRUST ARCHITECTURE

2.1. Early Zero Trust

Before a formulation for a network architecture based on a "Zero Trust" paradigm, the concept of trust had to first be formalized for use in computation. This problem is discussed in detail by [1], where Marsh provides a formalism used to define trust between agents and their situation. Trust between agents $T(a_1, a_2)$ is defined on an interval, $T \in [-1, +1]$. Here, a value of $+1$ indicates a complete, blind, trust between agents. A value where, $0 < T \leq 1$, indicates some trust between agents, and $-1 \geq T > 0$ indicates some distrust between agents. The term zero trust comes from the case where $T = 0$. This allows for the differentiation between zero trust and distrust. When designing a ZTA network architecture, this is used to force a no trust relationship between network resources and an agent, such that the network assumes it has no knowledge of the agent and that the agent must authenticate itself and prove its trustworthiness at each request for access.

Officially founded in 2004, The Jericho Forum (later merged with The Open Group [2]), was founded by like-minded CISO's with the goal of "de-perimeterized" IT infrastructure; the ideas developed and discussed eventually became ZTA. Up until this time, the prevailing approach to network security was to confine all network assets to a secure network [3]. In this secure network, assets blindly trust each other. The Jericho Forum promoted a deviation from this approach, towards one where data is protected outside of the network boundary, rather than attempting to restrict all data to reside within the network boundary. Both approaches are depicted in Figure 1.

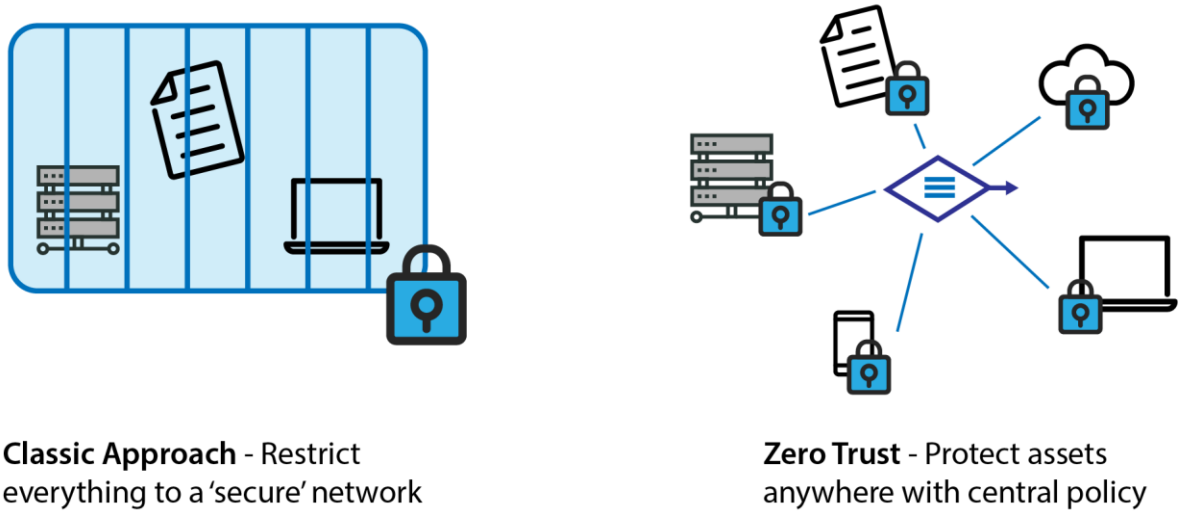

Figure 1. Classic vs. ZTA Network Architecture [3]

This change in paradigm promotes a more granular, centralized security model. A central policy makes an access determination for each session and between every network asset. Previously, policy decisions were essentially constrained between absolute access and no access within the network. In the zero-trust model, any asset can be allowed or disallowed dynamically from accessing other assets in the network.

The Jericho Forum lays out a set of eleven "Commandments" that must be followed for de-perimeterization [4]. The commandments are summarized below in Table 1.

Table 1. Jericho Forum Commandments [4]

Number	Summary
1.	The scope and level of protection must be appropriate and specific to each individual asset (not applied to groups of assets)
2.	Security mechanisms must be pervasive, simple, scalable, and easy to manage
3.	Ideal security procedures are not static across environments, each needs its own procedure
4.	Devices and applications must communicate over open and secure protocols that provide Confidentiality, Integrity, and Availability
5.	All devices must be capable of maintaining their security posture even when placed in an unknown/insecure network
6.	All devices, applications, and people must have declared and transparent levels of trust for all procedures
7.	Mutual authentication is required, and authentication and authorization procedures must support mutual authentication
8.	Access, authentication, and authorization must operate outside of the defined network; trust and trust information must be exchanged between organizations
9.	Access to data should be controlled by security attributes of the data itself, i.e., encryption
10.	Data privacy requires segregation of duties; privileges and key management should be done by individuals to avoid total compromise if the top of the chain is compromised
11.	Data must be protected by default in storage, use, and transit

Much of the early discussions on ZTA and de-perimeterization revolves around trust and cooperation and identity management. Trust and cooperation must require that two entities have some measurable reputation [5]. There are two possible mechanisms for this. The first is that organizations exchange information about reputations of individuals, such that one can learn from the experience of another. The other is that an initial, very low-level of trust be given to an individual, and the organization slowly grows the level of access as the individual's reputation is improved through good behavior. The concept of sharing information between organizations regarding trust of individuals and other organizations is referred to as federation.

Another core consideration laid out by the Jericho Forum for zero trust is that of identity management [5]. Identifiers are used to trust that the underlying entity is what or who they claim to be. These identifiers must provide a secure, verifiable level of confidence. The identifiers, along with

other attributes, are used to establish a contextual trust in the entity. For example, an identifier can be verified to establish a linkage to another entity, such as a government organization. Entities must be able to maintain multiple identities related to separate identifiers that are generated from the entity's core identity and do not reveal any compromising information about the core identifier used by the entity. These separate identities can then be used in separate contexts for separate tasks. Access to resources must be granted or denied based on rules validating these identifiers. The responsibility for protecting identifiers from use by an unauthorized entity relies on the identity owner, but authorized entities should also be able to seize another entity's identity. Identity management often assumes that the underlying entity is a person. This becomes a unique problem in OT in which entities operate constantly and autonomously without having a person who holds a secret such as a password when authenticating to a service. These problem and potential solutions or workarounds are discussed in further detail in Section 4.

ZTA or Zero Trust Network Access (ZTNA) was later coined by John Kindervag at Forrester Research [6] in 2010 [7]. This was roughly a year following the start of the implementation of BeyondCorp at Google. ZTA was an attempt to codify the initial work of the Jericho Forum and focused on how to replace implicit trust built into computer systems and protocol designs with explicit, provable trust.

Kindervag's primary idea was that an organization should not extend trust to anything outside or inside a given perimeter [8]. This effectively eliminates the need for a perimeter, essentially shrinking a given perimeter to a single system that requires an explicit trust relationship to be established prior to access. This led to the catchphrase "Never trust, always verify" that has become associated with ZTA. This is also one of the key tenets of ZTA that was later adopted by the National Institute of Standards and Technology (NIST), along with other system abstractions and authorization techniques.

This initial work inspired initial IT implementations, most notably Google's BeyondCorp [9] efforts, discussed in Section 3.

2.2. Zero Trust Definition

Most recently (2020), NIST provides abstract definitions for ZTA in Special Publication 800-207 [10]. This document defines a set of tenets that can be used to describe a network implementing ZTA, like those commandments set forth by the Jericho Forum and discussed in Section 2.1. The tenets defined by NIST are as follows:

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

These tenets are purposely vague, so that determinations and design decisions can be made abstractly before assessing specifics for procedures such as authentication, log gathering, and identity management. However, some of these tenets prove to be difficult to implement due to their ambiguity. For example, one organization's implementation of a dynamic access policy will look quite different to another's, and there may be discrepancies in their efficacy (see Section 3 for further discussion on implementation variations).

NIST improves on the general understanding of ZTA provided by the Jericho Forum by defining the logical components (depicted in Figure 2) found in a ZTA deployment. These components are used in determining access to resources by subjects. Each request must always pass through the Policy Enforcement Point (PEP), which will receive the request and negotiate with the Policy Decision Point (PDP). The PDP determines whether the system can be considered trusted and allowed to access the resource in question. The PEP is responsible for carrying out the decisions made at the PDP, enabling, monitoring, and terminating connections between systems and resources. The PDP is composed of two logical components: the Policy Administrator (PA) and Policy Engine (PE). The PE is the component which will utilize access policies, identity information, trust algorithms, etc., to decide if the system is allowed access. The PA will then establish the communications path (e.g., session-specific authentication tokens) between the system and resource, and issue relevant commands to the PEP. These components are distinct in their logical responsibilities, but may be implemented in the same physical component.

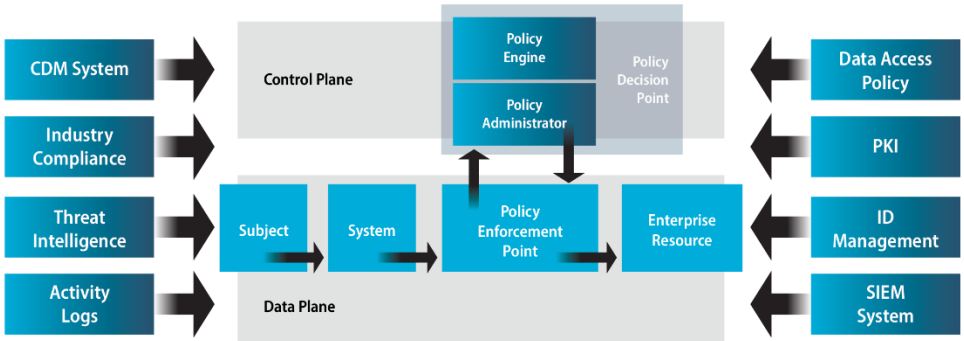

Figure 2. ZTA Components [10]

The PE contains the trust algorithm [10]. The trust algorithm has several inputs to consider when making a decision regarding an access request. These include the access request itself, subject database, asset database, resource policy requirements, and any related threat intelligence and logs. It is important to draw a distinction between the subject and the asset in this process. The subject can be considered the person or process that is requesting access to the service, and the asset is the device that the subject's request originates from. The subject database contains a set of identity information / attributes that can be used to establish a level of confidence in the subject's purported identity. The asset database contains known status of the device: operating system (OS), location, installed software, etc. Establishing this type of information has been explored in IT-based systems but lacks research in OT/ICS systems. Instrumentation and Control (I&C) devices operate without

a human present most of the time, meaning that the requestor for many access requests would be a process running on the I&C device. This process will typically be running with a high level of access to the asset's OS, making establishing trust in the subject and asset a challenge at the PE. See Section 4.2 for further discussion.

NIST also goes on to describe a few use cases and scenarios for which an organization would benefit from a ZTA implementation [10]. These scenarios are useful for outlining the benefits that are gained from a ZTA implementation. These include situations such as an enterprise with satellite facilities, cloud-to-cloud enterprises, and enterprises with contracted services and/or nonemployee access. These highlight the benefits of ZTA, mainly showing that ZTA offers organizations a cybersecurity solution in which data can stream between organizations, services, devices, etc. without the need for overbearing restrictions on whether that data flow may be visible to untrusted individuals or devices. This is because all data is protected by default and each endpoint is validated at the beginning of each session. The scenario most like a Nuclear Power Plant (NPP) would be the enterprise with contracted services. Vendors or maintenance personnel may be required to access NPPs physically, which poses a cybersecurity threat given their current cybersecurity practices. ZTA in this case allows those contractors to have access to resources and devices that they need to and be denied access by default to any other part of the network. Additionally, identity management components of ZTA mean that the contractors can be verified to be members of the organization that they represent and have their access levels immediately and automatically assigned.

2.3. NPP Cybersecurity Landscape

Typically, innovations in cybersecurity and computation in general are adopted first in Enterprise environments, and later may be integrated into OT environments. ZTA implementations (discussed in Section 3) are growing in popularity in the Enterprise sector, but just beginning to emerge for OT applications. For these OT applications, because the nuclear industry is highly risk averse, standard practices tend to stay commonplace for very long periods of time, and the process for implementing modern techniques may put a NPP in violation of standards.

Currently, the common practice for cybersecurity in a NPP involves strict perimeterization, shown in Figure 1. The current security approach involves network segmentation according to the security posture of the physical area, shown in Figure 3. Each boundary implies a division in the network, often many individual networks exist within the boundaries of these areas. Within these networks, there is often very little or no cybersecurity measures, and cybersecurity is assumed because of stringent physical access requirements.

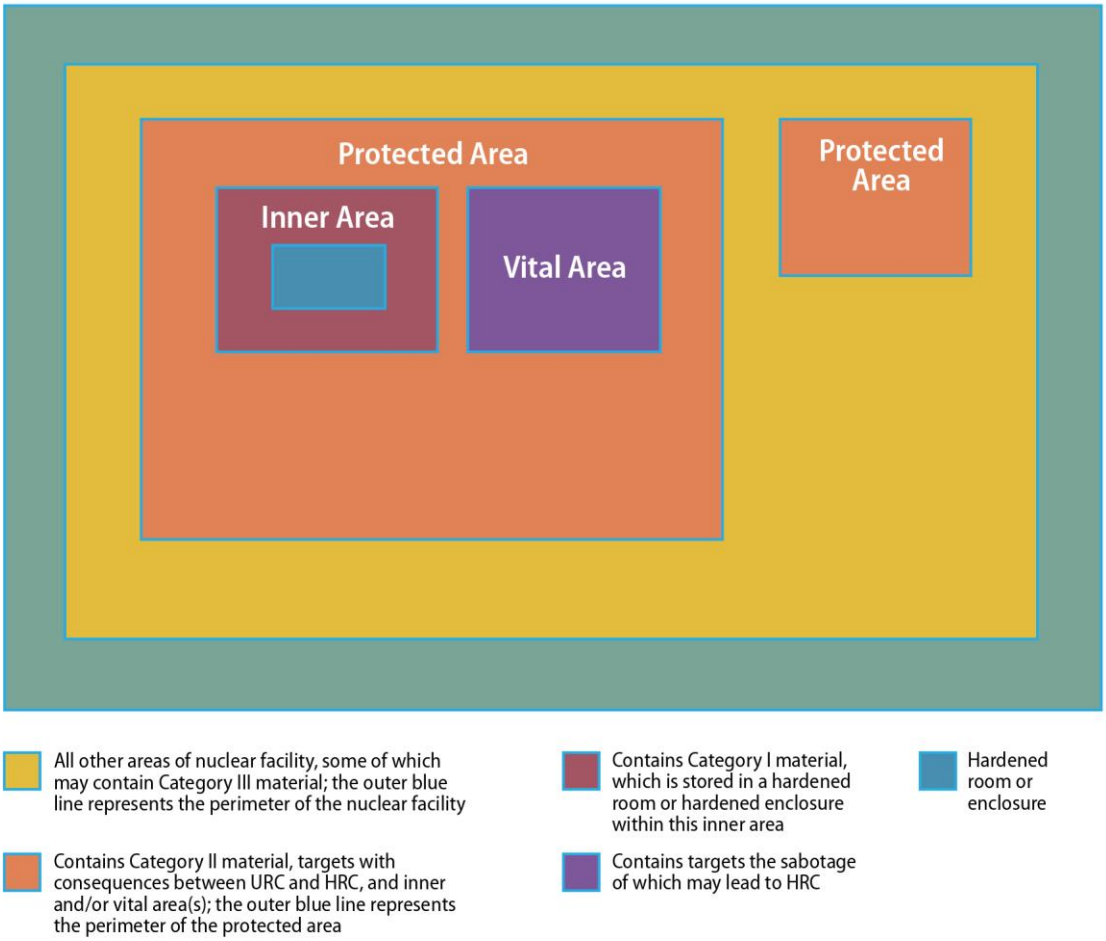

Figure 3. IAEA NSS 27-G Description of NPP security areas.

2.4. Summary

Core concepts to ZTA have matured over the years, growing from a small consortium of industry partners to being described in a NIST Special Publication. Concepts such as the commandments set forth by the Jericho Forum have been expanded upon, and NIST now provides logical component listings to guide implementations for ZTA. The concepts and logical components have not changed in the main goal of the architecture, which is to enable a network in which data can safely be transferred in the open. Two entities must explicitly form a trusted relationship for the data exchanges to be accessed and modified. The application of such an architecture to a NPP may allow for data to be transferred securely across boundaries between the areas depicted in Figure 3.

3. ZTA IMPLEMENTATIONS

3.1. IT Implementations

Organizations may choose to implement ZTA into their enterprise environment in several ways [10]. For the approach to be considered ZTA, it must adhere to all the defined tenets of ZTA (Section 2.2), but one or two logical components (Figure 2) can be chosen to be the driving force of the architecture. For example, an enhanced identity governance-driven architecture makes ID management the main logical component used for developing access policies. Other examples of implementation variations include logical micro-segmentation or network-based segmentation. Certain approaches may lend themselves to a more streamlined implementation for an organization, depending on their current practices and on-hand hardware. Today, common implementations that are called ZTA systems do not apply all ZTA tenets, most commonly omitting dynamic authentication techniques.

3.1.1. BeyondCorp

BeyondCorp [9] is an initiative created by Google to implement ZTA in their production environment; the process for this implementation is well documented [11]. The goal of the BeyondCorp initiative is to implement a de-perimeterized approach for access to internal resources by employees and devices. ZTA is implemented to change access policies from their typical basis in location and originating network to information regarding devices' states and their associated user. The BeyondCorp model considers both internal and external networks to be untrusted, and access is granted to resources by dynamically asserting a computed level (tier) of access.

BeyondCorp implements several components that cooperate with each other to ensure that no unauthenticated or unauthorized users or devices access internal resources. These components and their information flows are depicted below in Figure 4.

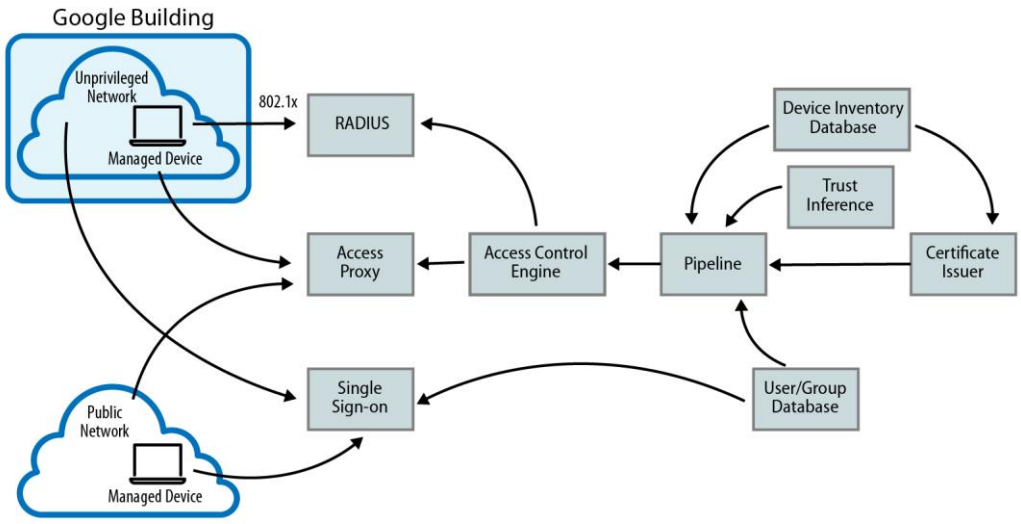

Figure 4. BeyondCorp Architecture [12]

An important aspect of this architecture to note is that devices both internal and external to the network are managed devices. The implication of this is that only devices that are procured and actively managed by the organization are allowed to access resources. These devices use either hardware or software-based Trusted Platform Modules (TPMs). The TPM is a tamper-resistant module used for secure key storage, which can be used to offer secure attestations to a device's identity and state. Google maintains multiple databases for storage of information regarding devices (Device Inventory Database), which satisfies the asset database component of the NIST ZTA paradigm.

BeyondCorp also implements the subject database component through what they call a User Database and Group Database. These databases exchange information with Human Resources (HR) processes ensuring that the Access Control Engine has access to updates information regarding employees' roles, responsibilities, access levels, etc. This information is used to securely identify users using a single sign-on (SSO) system that is externally accessible. The SSO system is the centralized point at which users obtain tokens that are presented to the Access Proxy, which serves as the gateway to a network resource. The Access Proxy and the Access Control Engine serve as the PEP and the PDP, respectively.

The Access Control Engine relies on what Google refers to as the Trust Inferer, which is responsible for creating an evaluation and issuing the final say on the tier of access given to a device or user. The Trust Inferer uses real-time information from the requestor as well as information from the User/Group Database and Device Inventory Database in this determination. Access tiers are based on the level of assurance that the device can provide; these assurances include things such as disk encryption, proof of management/configuration agents, and OS patch number. BeyondCorp implements several exceptions to the Trust Inferer, one which is most notable for this research is that IoT devices may be handled by exceptions and placed into their own tier. This is because IoT devices may not be able to maintain certificates that are essential to BeyondCorp's methods for proving asset identities. This is notable because almost all safety critical assets in a NPP are similar to IoT devices in this way; this problem and potential solutions are discussed in further detail in Section 4.

All internal resources are public facing, ensuring that any user or device with appropriate access levels can access resources. This also means that BeyondCorp's intranet must be treated as an unsecure/public network. This is accomplished by implementing a network that has no access to any resources other than the internet (and necessary networking services like DHCP) and ensuring that access to any resource is granted through the Access Control Engine. By establishing this type of network, all external and internal devices are treated the same way: initially untrusted.

BeyondCorp is an example of a real-world implementation of the principles and goals that have driven forward the concept of ZTA. BeyondCorp intakes real-time data about users and devices and uses that data for an intelligent selection of access levels. This allows Google to treat employees who are on-site and off-site in the same manner without affecting their productivity or causing delays. They state that they have not experienced issues with latency even when a policy or device state changes in real time.

However, there are some aspects of this implementation that are not ideal. For example, this implementation only allows devices which are owned and maintained by the corporation to access the internal resources. This relieves a large engineering burden by removing the need to prove trustworthiness on completely unknown devices, but also severely limits the accessibility and flexibility of the network. In the case of Google, it may be that there are very seldom contractors or

nonemployees who have a legitimate need to access resources without a device owned by the organization, but this is not the case with NPPs. NPPs may sometimes require contractor work where contractors must supply their own devices for maintenance or integration efforts, though this would not be the case in safety critical systems. The applicability of mitigations such as disallowing all non-plant owned devices from joining the network is discussed in more detail in Section 3.3.

The SSO used in this case is a centralized authenticator and does serve as a single point of failure. As described, a successful attack on the SSO would result in the ability to issue tokens to any machine that allow access to any resource. BeyondCorp has also not completely removed use of virtual private network (VPN), though it is stated that this is done in a very limited capacity and planned to be removed in the future [12]. Use of a VPN for access to resources does not coincide with the ZTA paradigm because it is a way of virtually placing the user into the trusted environment rather than individually proving trust at each request for access.

3.1.2. NY State Cloud Computing Center

BeyondCorp documents the overall architecture and design of a ZTA network, however the specifics of the component's implementations (e.g., the trust inferer) are somewhat lacking in detail. NY State Cloud Computing Center and Blackridge Technologies implement and explore one potential implementation of a PEP in [13]. The approach combines transport access control (TAC) and first packet authentication to verify identities of devices and users for TCP sessions.

This architecture means that each individual network session is authenticated at the transport layer. The authentication occurs before the user or device requesting access is allowed to access network resources, and all network resources are placed behind a gateway that implements the authentication procedure.

When a user or device requests access to the network resource, the first organization owned device that receives that network traffic is a gateway which applies identity information, and the second is the TAC gateway. This gateway will inspect the first TCP packet of the session, which must include an identity token. For the purposes of [13], that token is described as a 32-bit cryptographically secure token, which will expire after 4 seconds. Those tokens are associated with existing Identity Access Management (IAM) services, for example Windows Active Directory. The TAC gateway can then serve as a PEP. The included identity will be associated in an internal database with a certain level of access and either allow or disallow the subsequent session from being established. An example of the information flow in this scheme is shown in Figure 5.

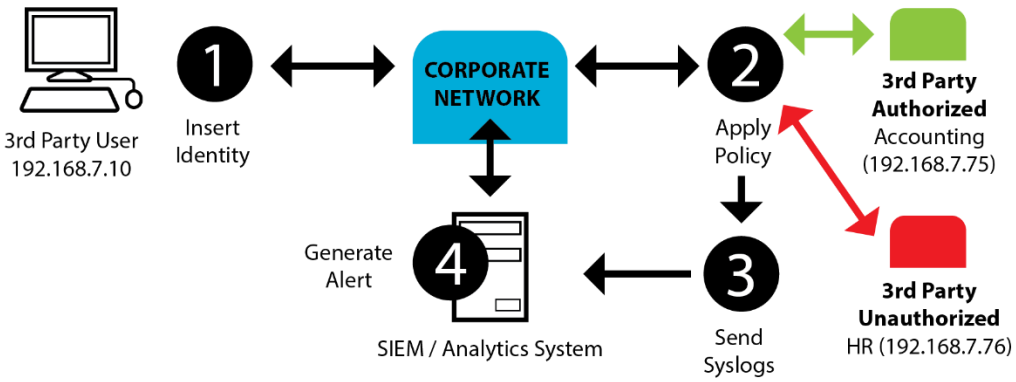

Figure 5. ZTA Proposed Network Flow [13]

In the above example, a user is allowed access to the organization's accounting services, but not internal HR information. Upon initiating a request, the user's identity information is pulled from the existing IAM service and inserted into the first TCP packet of the session, at 1. The connection then continues to 2, where the policy is enforced. Because the user is not allowed to access HR, if the TCP packet is set to route to 192.168.7.76, the connection is refused. If the request is for accounting, 192.168.7.75, a channel is then established. Information about identity tokens and access requests are then forwarded to a SIEM for continuous threat intelligence monitoring. Because in the case of a disallowed connection, the user's connection request is simply dropped, there is no information sent back to the user regarding access. This prevents any potentially sensitive information from leaking throughout the process. The user also should not be able to recover any of the access tokens used in the process because those take place in the organization's infrastructure, which the user is not allowed to access without explicit permissions.

This paper has provided a well described method for authenticating new network requests that are received in a semi-autonomous manner. There is no need for the user to interact with a new application because identity management information is embedded into the request at the transport layer. However, there is a significant amount of work that must be done to implement this in an OT environment. It is assumed that the environment will utilize an already established device and user identity management service such as Active Directory from Microsoft. ICS devices are not set up to utilize such services, and therefore there needs to be a service with parallel functionality developed for ICS devices in order to utilize this process in an NPP. In short, this implementation is an example of ZTA where the driving force is the ID management component, but NPP current practice lacks a suitable foundation of ID management in its OT devices, and therefore a large amount of work must be completed for a similar implementation in an NPP.

3.2. OT Implementation

Very few documented implementations of ZTA for OT networks exist. In 2021 Deloitte released a document describing a project that claims to have established a ZTA in a large ICS spread across multiple facilities of a chemical manufacturer [14]. Using the Purdue model of ICS architecture as a basis, Deloitte integrated principals of ZTA into the design of this ICS network. Implementing micro-segmentation, data flow restrictions and controls, access control, VPNs and more secure jump servers, and monitoring capabilities with a SIEM has certainly made a more secure environment. These security principals laid over the Purdue model are proposed as a reference architecture for ZTA in OT networks.

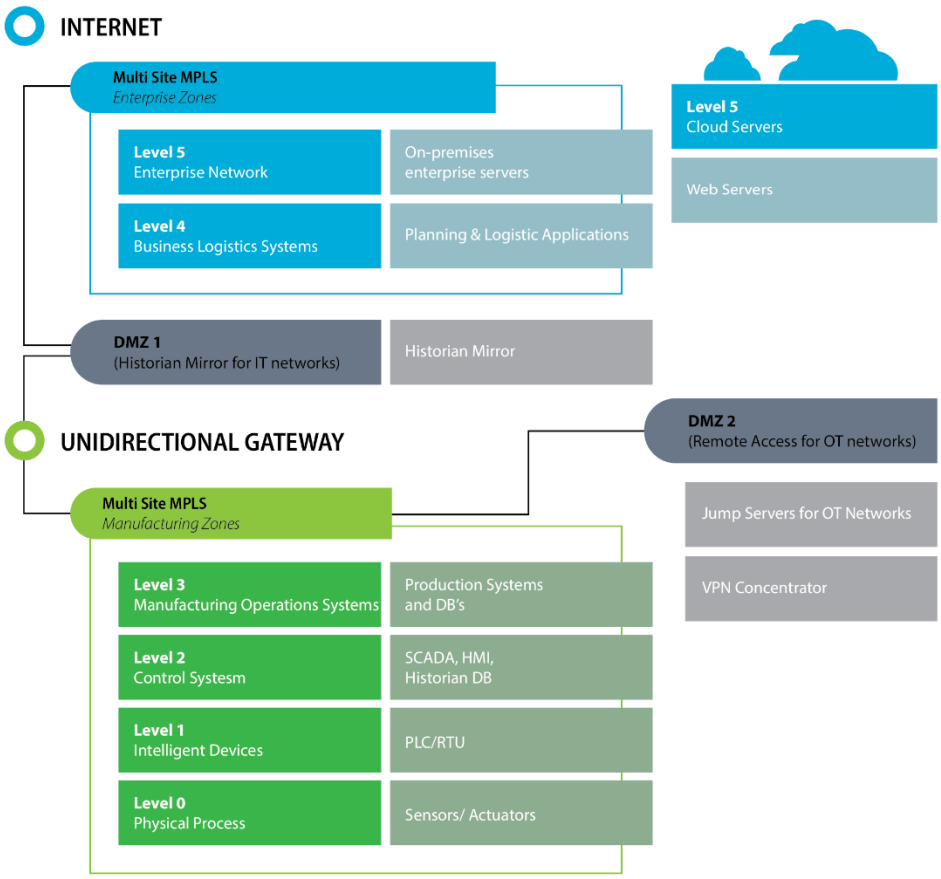

Figure 6. Deloitte OT ZTA implementation diagram [14].

While the Deloitte implementation does have some of the qualities of a ZTA, it falls short of the tenets described by NIST 800-207 [10]. The described implementation still contains many areas that have implicit trust between devices and is highly reliant on perimeterization of these implicit trust zones. This architecture is more focused on the interconnections to and above level 3 of the Purdue model, leaving the entire OT system from production servers and databases down to the PLCs and physical process controls as implicit trust areas of the network. This is depicted in Figure 6. Though the network is micro-segmented, individual resource access and communications in the OT network are not secured or controlled.

The architecture that Deloitte developed is certainly a step in the right direction for OT network security, but it does not qualify as a ZTA by the fundamental tenets of Zero Trust. Some aspects of the network security principals of Zero Trust are integrated into this implementation, but it is the things that are left out that are indicative of the current state of ZTA and its application to OT. The level of control over communications and resources within OT networks that ZTA demands is difficult to retrofit into systems that were never designed to provide these resources. The reasons that there are so few attempts at ZTA implementation in OT could be explained by the current state of the technology available and a cost benefit imbalance for system owners.

3.3. Analysis of ZTA for ICS/OT

The IT implementations reviewed do not include custom identity management; it is generally scoped out and something like Active Directory is used. This vastly constrains the type of devices that can be used, and compatible devices do not include ICS devices like PLCs, HMIs, or other typical industrial systems. Given currently available literature, ZTA can be implemented in the IT environment that exists at the plant before the historian. In other words, a NPP can implement ZTA for the business portion of their network, but there is no available literature or tools for extending that environment into the operational portion of the network. If a NPP was to extend this capability into the OT network, it would violate regulation by setting an accessible path to safety-related ICS devices from outside of the network, i.e. deperimeterized.

Furthermore, industrial systems must be viewed through an operational lens, where availability and integrity are more important than confidentiality. This is in direct contrast to business environments, where confidentiality is of great concern and brief disturbances in availability do not result in potential safety impacts (and are often expected). Recognizing all computing services and data sources and sinks as resources, we can re-interpret ZTA guidance regarding secure communications. Specifically, communication must be secured by focusing on availability and integrity first and foremost. This specifically impacts NIST ZTA tenets (1) and (2) and is a distinctly different approach to how communications are secured in IT systems, where confidentiality and integrity are typically more important attributes than availability. Not to imply that IT communications are not concerned about system availability, rather that availability is much more important to industrial systems and NPP as a lack of data availability at the wrong time can lead to significant physical consequences.

In addition, granting access to resources within a control system is more frequently system-to-system use, not user-to-system use; this renders modern techniques used to strongly authenticate users like multi-factor authentication much more difficult. Authenticating a system based on something that system has, like a certificate, is possible, but evaluating something a system knows is not, as that is trivially accessible if a system is compromised. This makes controlling system access on a per-session basis more difficult. Certainly, systems can re-present certificates, but this may be of little value as those certificates are present on compromised systems as well as uncompromised ones. In ZTA implementations, systems are identified using a device database, like the device inventory database used in Google's BeyondCorp implementation. Modern mobile devices like iPhones use a Secure Enclave Processor (SEP), separate and inaccessible to the application processor to store sensitive information like certificates or biometric data. Google uses trusted computing support to identify systems in tandem with their device database. Other approaches use weaker methods to identify systems like MAC addresses to authenticate systems for network access, but these kinds of methods can result in compromise through MAC spoofing attacks. Overall, ICS systems were not

historically designed to provide these kinds of strong authentication services. This impacts NIST ZTA tenet (3).

NIST ZTA tenets (4) and (6) require dynamic authentication policy application after authentication material is presented. This requires fusing state information with policies describing system-to-system access. Typically, this is applied at PEPs and PDPs within authentication systems and require policy storage and access. Even if this kind of evaluation is engineered to be extraordinarily performant, it still creates dependencies on authenticating systems. This leads to additional potential points of failure (which must be very carefully managed), as well as some unavoidable additional latency. As connections between resources is managed on a per-session basis, the amount of introduced latency is heavily variable, but the system dependencies will always exist in some form.

NPP licensees typically apply NIST ZTA tenets (5) and (7), already. This is done by closely monitoring system behavior for pending failures and overall performance, as well as system security, when designed and built following international standards and national regulatory guidance.

NIST ZTA tenets (1) and (2) have a different focus in NPP systems. Most protocols used to secure communication provide confidentiality and integrity protections, but not necessarily availability protections. Typically, availability is provided by other protocols in a given network stack. TLS and IPSEC are both common communication protection protocols, but session control, including retransmission, is provided by TCP. ZTA tenets (3), (4), and (6) have significant barriers to implementation, potentially increasing system risk by increasing the risks of single point and common cause failures as well as increasing communication latency in unpredictable ways. An additional consideration for these tenets is the development, purchase, and integration of OT-focused solutions for operators. Finally, tenets (5) and (7) are typically adhered to within NPPs today.

4. ZTA FOR OT/I&C NPP – LANDSCAPE

4.1. Advantages

Many advantages can be achieved through implementing ZTA in an organization, and those advantages also apply to NPPs. Current cybersecurity practices in NPPs (Section 2.3) are lacking in maturity; these practices have not kept pace with advancements in cybersecurity that have become commonplace in IT and enterprise environments and make many assumptions about access to resources that may not hold true. If a NPP were to implement ZTA across its entire network architecture, its cybersecurity posture would be much more hardened than that of the typical NPP. This is due to the various tenets and requirements of ZTA implementations, for example that data should be protected in use, transfer, and storage. Many OT environments utilize legacy serial communications protocols that do not offer modern cybersecurity protections like encryption, message authentication, or non-repudiation.

ZTA puts a strong emphasis on federation and identity management for threat intelligence purposes, as defined by tenet 4 of NIST SP 800-207 [10]. Policy must be set not only according to identities of assets as users, but also environmental factors and behavioral attributes. A proper ZTA implementation takes measures to record information relevant to these attributes, which can be used for threat intelligence and continuous monitoring, adhering to tenet 5. Federation and Identity Management are further defined in NIST SP 800-63c [15]. These requirements will result in any ZTA-implementing NPP's maintenance of rigorous databases on relevant information to circumstances surrounding a user's access to important NPP operational data. A Licensee may utilize a Federation Authority to normalize data across multiple plants or sites. It is also possible for a Federation Authority to be developed such that multiple separate Licensees may benefit from the recorded behavioral characteristics from another Licensee's plant. This federation would provide a strong basis for real-time threat detection, as the Federation Authority can revoke access levels of a party to all its managed environments after malicious or abnormal activity is detected at one. This directly lowers risk for Licensees by reducing the potential impact of a malicious actor.

ZTA may also provide an advantage by increasing the predictability of traffic found on the network. This is because when a user or machine on the network accesses a resource, it must follow a specific procedure to gain access to its desired network resource. For example, a network implementing an access policy similar to [13] would expect to see the first packet of the network traffic include some key used to identify the subject. Should a network analyst (either human or automated) observe network traffic that is not consistent with this procedure, there can be a somewhat high degree of confidence that the traffic originates from an attempt to gain unauthorized access. This information is then able to be processed for ongoing threat intelligence and shared within and beyond the organization, as appropriate.

[16] scenario 4 details how the flexibility gained with a ZTA implementation could offer advantages to an NPP. NPPs operate according to standards and guidance from a multitude of organizations including standards bodies like the Institute of Electrical and Electronics Engineers (IEEE), International Electrochemical Commission (IEC), and International Atomic Energy Agency (IAEA), and national regulators like the US Nuclear Regulatory Commission (NRC). They maintain product supply chains originating from many different countries as well. A NPP which implements ZTA will have an easier time maintaining compliance and compatibility by instituting an automated audit system for continuous monitoring. Core principle 6, Alignment and Automation, is applied to the plant to enable real-time log capture, storage, and analysis, meaning that automated analysis rules can provide Operators with immediate notification of standards compliance. Upon updates to standards

or guidance, the rules are updated, and all non-compliant assets or procedures are detected for remediation.

ZTA must also provide an asset-centric security approach to the network, as stated in core principle 8 [16]. This means security practices are tailored to the assets within the network rather than an approach that is application-centric, for example. This reduces the complexity of the network and enables an easier exchange of data between interfaces. Approaches such as format preserving encryption and tokenization reduce the complexity of interfaces that must be maintained between different types of assets. ZTA also helps to secure high-value systems by adapting to changes in the environment quickly and autonomously. ZTA specified secured zones, policy-driven access control, and context-specific data security. This means that policies can be adaptive and provide data security depending on situational risk at any time to systems without the need to network-wide changes or updates.

4.2. Challenges

There are many challenges that must be overcome to implement ZTA in an NPP; most of these challenges arise from the inherent lack of cybersecurity features in most ICS and SCADA systems and components. For example, generally the control system for the NPP will be implemented over a serial communication protocol. Common serial communication protocols like DNP3 or modbus do not implement encryption, authentication, or other security measures by default [17]. ZTA requires data to be protected at all points during its lifecycle, so there must be some hurdles overcome to enable encrypted communications within the plant's operating environment. There are some protocols which implement modern security standards, such as Modbus/TCP support for Transport Layer Security (TLS). An overview of the process for Modbus/TCP security is depicted in Figure 7 [18].

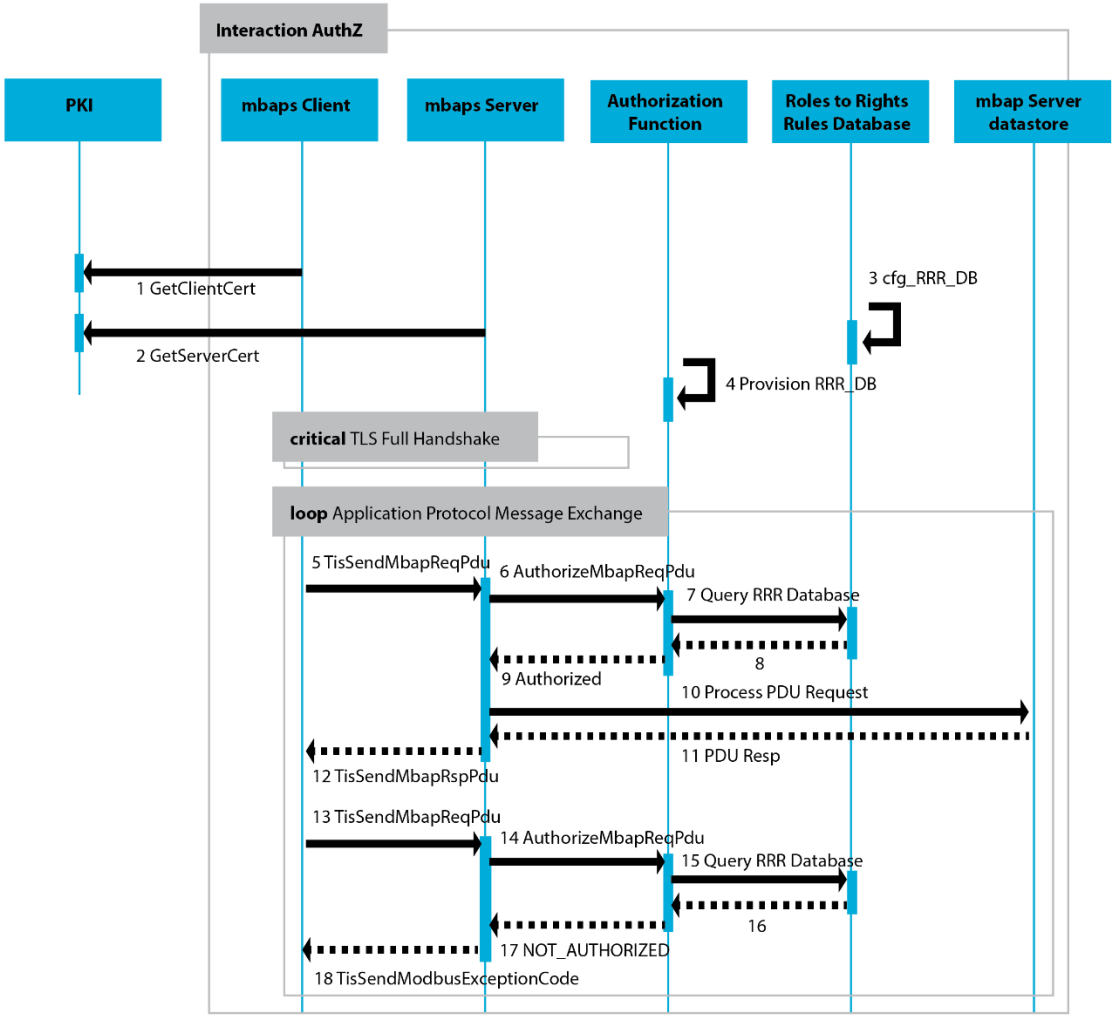

Figure 7. Modbus/TCP Security [18]

A critical component of the security for Modbus/TCP and TLS in general is labels 1 and 2 in Figure 7, where the client and server verify each other's identity using X.509 certificates. This requires a robust Public Key Infrastructure (PKI) be developed and maintained for the devices, which is currently not in practice. A PKI internal to the plant could be operated, but the storage of keys must also be secure. BeyondCorp can relieve itself of many challenges associated with this process by requiring that all assets on the internal network include a TPM that is managed by the organization [12]. This is infeasible using current ICS devices because manufacturers either do not include any TPM or TPM-like component or do not allow for operators to access these components through any provided programming software.

The challenge of maintaining secure private keys and certificates for TLS encryption gives rise to the larger issue of identity maintenance in general for ICS devices. In Enterprise environments, there has been decades of research and development invested in the creation of protocols and architectures for identity management of both users and their devices, such as Active Directory. Active Directory uses Kerberos to issue and keep track of access requests and "tickets" that are served to users based on successful requests for access to assets [19]. There is no such architecture in place for ICS, so some identity and access management system must either be developed and tailored to ICS devices or manufacturers of ICS devices must make major changes to both included hardware and device firmware to make them compatible with existing solutions in the IT space. It may also prove to be challenging to maintain required real-time constraints on ICS devices when implementing modern encryption and authentication protocols.

ZTA implementations must also maintain robust databases for users, not just digital assets. This poses a unique problem with ICS devices because there is not generally any user space on the device. The concept of separate address spaces for user and kernel space (shown in Figure 8) is implemented in Operating Systems that are used in IT environments. This logical separation ensures that a typical user on a device is not able to access fragile components of the system's kernel or directly interface with the system's hardware. The kernel implements many security measures to provide this assurance. Real-Time Operating Systems (RTOSs) operate exclusively in the kernel address space or do not protect kernel address space from processes running in user space. This means that any user who gains access to an ICS device running an RTOS can masquerade requests as coming from the digital asset rather than the user. Additionally, kernel space processes that are used for reporting accesses by users or other important security information that is required to be logged and analyzed constantly by ZTA can be subverted relatively easily.

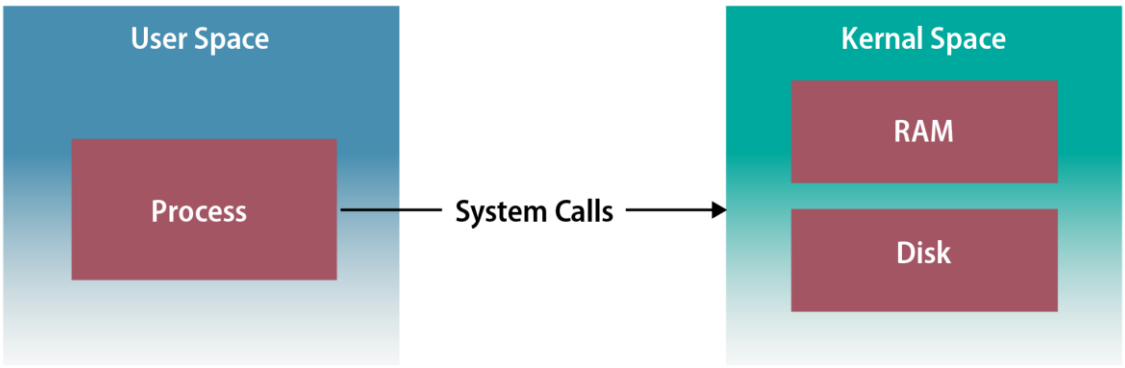

Figure 8. User and Kernel Space Segregation [20]

Significant challenges for ZTA stem from the needs of the operational environment present in NPPs. The communications in OT are far more important to the safety of a facility than communications in an IT network. A PLC reporting a safety value, or a stop command, must reach its destination to ensure safe operation. It is also vital in OT systems that information is timely, and communications cannot be encumbered by encryption or other security controls that delay safety systems or stresses the limited computational resources on controllers, especially in situations where performance of these PLCs could be degraded. The security of the information communicated is far less important than ensuring the information's integrity and validity, which presents a very different paradigm from IT security.

For continued safe operations of a process control system, some controllers may need to be allowed to communicate and operate even if de-authenticated. For many systems in an NPP the loss of authentication on a PLC and subsequent communications halt could blind operators and reliant control systems from critical process data. A ZTA implementation will need flexibility to allow some contingencies for scenarios where a device critical to operations can be untrusted but allowed to operate on the network. This is a unique requirement of OT for security, and therefore some method of retaining operational safety while ensuring security must be implemented.

A method of remote authentication of devices may need to be developed. Should a device fail authentication from network faults or environmentally induced noise, it may need to be reauthenticated remotely. Demanding a full stop of production or a reactor shut down for a device to be re-authenticated would be an undue burden on operators and reduce the likelihood of adoption and retention of ZTA security practices. This could be aggravated in emergency conditions where areas of the control system may be inaccessible to workers, but control over the system must be guaranteed. This presents a major issue in ensuring the security of a ZTA network, potentially introducing a backdoor to later exploit. But operation outside of normal parameters and expectations must be taken into account to ensure that potential implementations of security measures do not hinder or limit emergency responses or compromise system safety.

As mentioned previously, Licensees must adhere to regulations and standards from multiple organizations and ensure compliance with information security guidelines. Implementation of ZTA would not contradict the cybersecurity regulations according to CFR 73.54 [21] and would likely improve cybersecurity posture of plants. However, a partial or complete implementation of ZTA would likely result in systemic changes to a plant's network structure, cybersecurity and incident response plans, and communication flows. This means that a new Cyber Security Plan must be developed in accordance with NEI 08-09 [22].

Many nations derive NPP cybersecurity regulations and guidance from IAEA documents on the subject. For example, Section 5.2 (Cybersecurity Measures) of Canadian Nuclear Safety Commission DIS-21-03 [23] references three IAEA documents: NSS 17-T [24], NSS 23-G [25], and NSS 33-T [26].

IAEA NSS 17-T [24] recommends establishing logical and physical boundaries for information flows based on associated risk levels of information. Boundaries increase as risk decreases. ZTA implementations are not conducive to logical boundaries of data, as the purpose is for a deperimeterized network. So, while a ZTA implementation can provide rigorous data protections that allow for data to safely leave local network segments, guidance provided by IAEA may not allow for this flexibility. ZTA Tenet 2 states that access requests originating internally and externally must meet the same requirements, and that all communications must be secured regardless of network location [10].

IAEA NSS 23-G [25] provides guidance on establishing confidentiality, integrity, and authenticity of information and communications within an NPP. This guidance allows for information to be shared with outside organizations, such as appropriate state agencies, external states, or the public, when necessary. The guidance also provides the ability to change security measures of information according to regular audits and investigations. The Information Security Plan guidance in 23-G aligns heavily with ZTA principles. For example, the plan should provide for regular monitoring and review to ensure that procedures remain relevant and effective, which aligns with ZTA Tenet 5 [10]. Additionally, for sensitive information, information must be restricted to those who need access to perform their duties, have been granted the authority, and who have undergone a trustworthiness

check commensurate with the classification level of the information [25]. ZTA provides a framework for which this trustworthiness check can be implemented as well as allowing for the requirements of access to be dynamically updated according to new information, and subsequently enforced by the PEP autonomously.

IAEA NSS 33-T's objective is "to provide guidance for the protection of I&C systems at nuclear facilities on computer security against malicious acts that could prevent such systems from performing their safety and security related functions" [26]. A key concept to this guidance is a risk informed approach to cybersecurity. This is in line with other IAEA guidance on cybersecurity, following closely in line with information classification levels, and security measures that line up with those measures. A ZTA implementation is well suited for compliance with NSS 33-T, as a strong relationship between risk management and computer security teams can be made effective and efficient with dynamic policy updates, continuous monitoring, and revocation or adjustments of access privileges based on real time threat intelligence and changing risk factors / consequences. NSS 33-T prescribes a strict access control policy and a minimal number of access points. This is implemented in ZTA with a centralized PDP and PEP.

5. APPLICATION TO NUCLEAR POWER PLANTS

The nuclear industry must be conservative when it comes to the adoption of new technology as introducing systems with low operational experience presents a significant risk. Currently, ZTA for OT is an emerging technology and is not ready for deployment in the ICS environment of an NPP. However, ZTA has the potential to provide a more secure posture for OT systems soon and allow modes of operation not previously possible that would provide major economic benefits to new designs. The benefits are highly contextual, and we will have to analyze three different scenarios to understand the application cost-to-benefit ratio.

Over the years, NPP designs have changed, and with those changes come different network layouts, security considerations, and requirements for data confidentiality, integrity, and availability. This means that ZTA implementations for different NPP designs must also take on very different designs. However, ZTA standards and guidance leave room for this and can look very different. For example, NIST SP 800-207 [10] mentions a ZTA design based on network micro-segmentation. This involves a network architecture based on intelligent switches or routers or Next-Generation Firewalls (NGFWs) to act as PEPs and protect related groups of resources rather than each resource independently. For NPP designs with different data requirements (discussed in more detail in this section), micro-segmentation can be used to isolate Level 1 or safety data that may not leave the protected area from other resources associated with resource groups from higher levels. This would allow for a legacy design NPP to implement ZTA and maintain a high level of deperimeterization for many functions, while still maintaining strong physical protection as the required barrier of entry to interact with safety related communications.

5.1. Current Fleet

The reactors operating today are large units, constructed to provide gigawatt scale power to achieve some economy of scale. This has led the implementation of large, complex inter-dependent systems with multiple diverse redundant safety systems. Through decades of staggered system upgrades, repairs, and replacements, their networks are highly complex and frequently have mixed communication protocols. Today, these networks are separated into levels that correspond to the criticality of their functions as related to their necessary function and reliability to maintain safety and plant operation. Guidance from a variety of standards and governance organizations, including the IEC and the IAEA specifically address using a risk-based and graded approach to system design for nuclear systems, so that the most critical systems are afforded the strongest protection.

TABLE III–1. LIST OF SYSTEMS: EXAMPLE OF APPLICATION OF COMPUTER SECURITY LEVELS AND ZONES

System	Most significant function	CSL	Logical boundary	Physical boundary
I&C reactor protection system	Prevent accident conditions	1	Dedicated internal network decoupled using data diode No external network connectivity	Equipment located in a single VA only Computer security measure (data diode) located in VA
I&C reactor limitation system	Control reactivity	2	Dedicated networks, decoupled using data diode, firewall or other security devices	Equipment located in one or more VAs Network cables, equipment, or routing outside of VAs are physically hardened (e.g. secured conduit, panels)
I&C process information system	Provide alarms and notifications to operator on facility environment and status	3	Interconnected networks with HMI Note: This may be a separate or additional MCR HMI console	Equipment and networks located in PA and/or VAs

TABLE III–1. LIST OF SYSTEMS: EXAMPLE OF APPLICATION OF COMPUTER SECURITY LEVELS AND ZONES (cont.)

System	Most significant function	CSL	Logical boundary	Physical boundary
I&C operational automation systems	Control BOP systems	3	Interconnected networks with HMI Note: This may be a separate or additional MCR HMI console or combined with an I&C process information system	Equipment and networks located in PA and/or VAs
Office IT	Perform personnel functions	4	No logical connection (wired, wireless or portable interface) allowed with any level 1, 2 or 3 zone (system)	Allowed in LAA, PA and VAs
Telecommunication systems	Call to response forces or other external agencies as required	4	No logical connection (wired, wireless or portable interface) allowed with any zone assigned to level 1, 2 or 3	Allowed in all locations necessary for operator objectives
Personal mobile IT devices	None required — exemption only	5	Only allowed on level 5 networks No proximity with any zone assigned to level 1, 2 or 3	Not allowed in VAs

Note: BOP — balance of plant; CSL — computer security level; HMI — human–machine interface; I&C — instrumentation and control; IT — information technology; LAA — limited access area; MCR — main control room; PA — protected area; VA — vital area.

Figure 9. IAEA NSS 17-T Computer Security Zones and Levels [27]

The security levels above 1 will have some sort of network, though they may be limited, and thus could potentially benefit from ZTA. Level 1 systems may use networking to allow systems to communicate, but those networks are typically secured within vital areas and are isolated from external systems by hardware that will only physically allow outgoing communication.

This provides extensive isolation of these systems, with associated increases in overall system security. These systems however are heavily reliant on physical security controls to guarantee integrity, and communication protocols are generally simple and transmitted with no information protections supporting integrity or confidentiality.

Physical security controls may be sufficient to ensure that all systems communication is trustworthy, but nevertheless the inability to digitally identify systems and guarantee the integrity of communications is a possible vulnerability, though likely a vulnerability that is very difficult to exploit due to the physical security posture surrounding these systems and areas.

Furthermore, even if the supporting hardware to implement systems identification and associated authorization and authentication did exist, the time required to validate systems and communication from those systems may very well break hard real-time limitations for system control, leading to compromised overall system safety. Likewise, this kind of system authentication through a ZTA system introduces additional points of potential failure that need to be managed. For example, in an implementation like BeyondCorp, authentication of a system requires input from the Device Inventory Database as well as the Trust Inferer to implement a given policy. This introduces at least two points of compromise that could be potentially exploited. If the system is not designed to handle this possible exploitation, this attack could lead to a loss of function as authentication is arbitrarily denied or requests simply do not receive a response. Generally, this is an increase in system interdependence and a lack of system isolation which may introduce additional common cause failures.

Frequently, information from these systems is sent to an HMI in a remote-control room outside the VA. This communication is essentially outside the VA, even though the data communicated is generated by the equipment within the VA. As such, this communication is via a hardware limited data diode and is in fact at Level 3.

System-to-system access under ZTA principles may increase system risk dramatically. User-to-system access is not constrained by such real-time limits. User access could very well be managed via ZTA principles, but it will potentially create the same kinds of system dependency issues as authentication and would require policy and current state analysis prior to the grant of access, leading to potential common cause and single-point failures.

Overall, using designs currently implemented in IT systems, ZTA approaches at this level may lead to excessive new risk with little gain, assuming strong physical security controls are in place and that systems are inaccessible external to a given VA. From a ZTA perspective, while these systems are certainly monitored and information is collected describing the current state of assets and resources, applying dynamic authentication and authorization and per-session access determination is, currently, an unacceptably high risk.

Level 2 systems have similar isolation requirements as Level 1 systems but require command and control input from a main control room and as a result cannot be completely isolated from external influence. Arguments that apply to Level 1 systems apply equally to Level 2 systems with respect to system authentication and authorization within a set of systems at this level. Likewise, communication from a Level 2 enclave can be assumed to be at Level 3, and adheres to safety, security, and performance requirements of that level as it is external to the Level 2 enclave. Incoming data from a control room is a different issue, however.

Current guidance from the IAEA suggests using logical and physical partitioning to separate command systems in a control room from other systems in that control room at higher levels [24]. In this case, the systems that issue commands to Level 2 systems could still be considered Level 2 systems. However, this requires complete logical and physical separation. As a result, these systems would not be able to share physical communication infrastructure like cabling or switching or logical infrastructure such as authentication or printing systems. This would increase the cost of a given control room implementation, but would essentially allow these systems to operate at Level 2 within an environment containing Level 3 and higher systems. This would, however, imply that the control room would be secured physically in compliance with guidelines established for Level 2 systems.

ZTA would likely prove most useful in the systems in Levels 3-5 where more interconnections happen and contact with the systems from contractors and personnel are more frequent. These networks begin to resemble IT networks as they become more interconnected and less rigorously managed from a safety perspective. That said, systems at Level 3 still require extensive system redundancy and must be designed to avoid common-cause failure conditions. Adhering to the current recommendations of NPP OT network design and security results in a segmented, diverse, and redundant control system [28] [29] [30]. This can make implementing ZTA very difficult, as it would need to be deployed across many small networks with a large variety of vendors and operational needs. This diverse topology also lends some inherent protection as it requires a far larger investment from an adversary to defeat many different types of equipment with different vendors, on different networks, at the same time to disable one redundantly provided plant function.

The implementation of ZTA on legacy plants does not balance out favorably between its benefits and its cost with one exception, remote access. Studies reveal that employees prefer flexible work environments, which allow for hybrid or full-time remote work [31]. Additionally, during the COVID-19 pandemic, high education individuals were more likely to experience hybrid or remote work environments [32]. With more high-value employees preferring not to commute for work, and with many companies yielding to this request, vendors and operators may face heavy economic pressures to seek methods to allow for a higher volume of remote work. ZTA's deperimeterized and high security approach to network architecture offers a possible solution for NPPs to maintain an enthusiastic and productive workforce during this shift.

5.2. Advanced Reactors

Many of the new generation of reactors are designed around the concepts of inherent and passive safety. This major shift in design principles moves away from active safety systems that must intervene to return the plant to a safe condition. Safety is central to the design so that no operator or control system intervention is required to ensure that the reactor is within a safe operational envelope. For some designs this removes the need for the critical safety systems in Level 1, which should reduce cost of advanced reactor designs and may make ZTA far more relevant to these plants.

Inherent and passive safety are fundamentally different, and the distinction is important for the operational profile of a facility. Inherent safety is the removal or exclusion of a hazard inherent in the system [33]. Light Water Reactors (LWRs) have high-pressure primary coolant systems, moving to a coolant that operates at atmospheric pressures is an inherent safety feature. By eliminating the hazard completely, a reactor can have an inherently safe operational characteristics. However, power reactors have inherent hazards that cannot be eliminated completely. Advanced reactor designers have therefore taken approaches to reduce the hazards and attempt to mitigate them with passive safety features as much as possible.

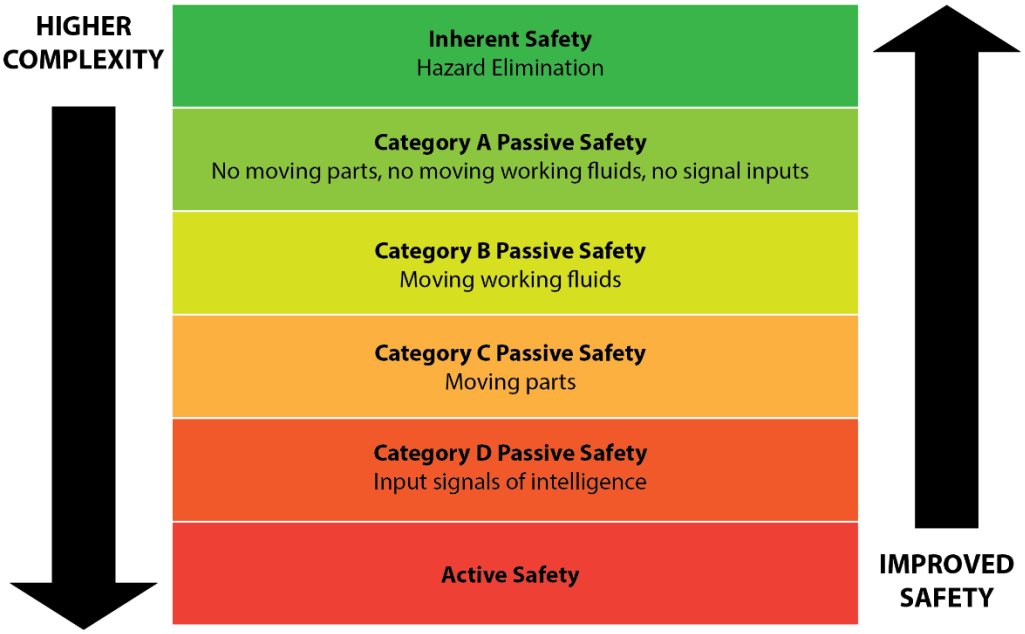

Figure 10. Categorization of safety features as described by the IAEA.

Passive safety features are broken into categories A through D based on their reliance on mechanisms and systems to provide the safety effect (Figure 10) [33]. The strongest passive safety features (category A) are based on the physics of the system, that the hazardous conditions are made impossible by leveraging natural physics. An example would be a reactor that is heavily controlled by negative temperature coefficients and the materials of the fuel, its cladding or encapsulation, and the vessel can withstand the highest temperature of the fuel with minimal heat rejection via the vessel walls.

Passive safety that requires an additional system or mechanism to ensure the safe operation of the plant are inherently less reliable and thus occupies the B-D categories of passive safety systems. An example would be an external heat sink or heat rejection system like DRACS [34] and RVACS [34] [35]. These passive safety systems rely on mechanisms that can still be defeated through external forces, such as collapse of cooling vent paths. While not as robust as inherent safety or category A passive safety features, these provide safety far beyond the capabilities of conventional active safety systems.

Advanced Reactor design's investment in passive and inherent safety mechanisms has sweeping effects in the design decisions and possibilities for control systems. Simpler, streamlined, and interconnected systems allow sophisticated plant control, predictive maintenance, Digital Twin integration, multi-unit control rooms, and remote operation. These come at the cost of potentially reduced Defense in Depth (DiD) as a consequence of smaller, more streamlined control systems, leading to an inherently more vulnerable attack surface [36]. With a more streamlined control system, the inherent protection of diversity, brought on through the sheer size of control systems in

current generation plans, is lost. This can be devastating to the security posture of a plant if their main control system vendor had a major vulnerability across their product line.

CVE-2022-38465 [37] demonstrates the danger of a low diversity control system. This vulnerability was discovered by Team82[1] and effects the entire Siemens S7-1200/1500 PLC product line, which have a global hardcoded cryptographic encryption key. This vulnerability could be used to irreparably compromise these PLCs [38]. If advanced reactor designers intend to streamline and interconnect their control systems, improved and robust cybersecurity measures are imperative. In low diversity systems such a vulnerability could produce a Common Cause Failure (CCF) through adversary action.

Overall, in these systems, Level 1 systems have been eliminated, potentially completely. However, this does not eliminate all concerns with respect to using ZTA techniques and approaches in SMR-oriented NPP. NPP operators do monitor the security posture of assets, but they do not secure all communication, nor do they grant access to resources on per-session basis within security levels.

As the Level 1 systems are reduced or eliminated in advanced reactors, Levels 2-3 will not necessarily expand but may become larger with respect to a single control room with the introduction of the multi-unit control room concept. NuScale intends to operate multiple reactor units, up to 12, in one control room [39]. These reactor units will share some systems including control and protection systems [40] and require advanced automation to reduce operator load across 12 units. In comparison to the current fleet of reactors, this is a significant increase in integration across units [41] and indicates an expansion of Levels 2-3 in principle for a single site. The requirements on automation for balancing operator load and shared systems point to some manner of conformity and interoperation across the control system that could correspond to an overall reduction in diversity and segmentation.

It becomes necessary to consider the possible failure of the system in situations where low diversity produces a CCF through a vulnerability on an entire system. For a multi-unit installation with shared Ultimate Heatsinks, if an adversary gained full control over the highly integrated control and protection system, would they be able to drive the reactors to overwhelm their safety critical heatsink? Adversarial operation must be investigated to fully appreciate the impact of cybersecurity concerns. Implementing ZTA in advanced control systems for multi-unit installments may provide a critical improvement in the cyber security posture to cover the inherent deficit created.

5.3. Micro Reactors

Micro Reactors are a subset of Small Modular Reactors (SMRs) that fill the remote location and portable power niche that requires special consideration and has unique modes of operation. These reactors will have significant economic and operational environmental incentives to operate remotely and autonomously. As such, most microreactor designs are heavily focused on inherent safety [42], their size and limited power production lends well to inherently safe features. However, there are limits to the inherent safety of these reactors [42] and designers often do not assume malicious operation while also centralizing control systems.

[1] Team82 is the cybersecurity research team for Claroty, a cyber-physical protection company. https://claroty.com/team82

Current micro reactor technology has significant uncertainty. There are many conceptual designs and theoretical control systems, but no mature data on control system function or design criteria that could be used to make detailed analysis of their cybersecurity posture. Generalized observations on the operational and market niche requirements can be made and infer some of the potential concerns and respective potential applications for ZTA.

Their size and market demand for remote locations, staffing minimization, and cost reduction drive design features for micro reactors that have cyber security implications. Inherent safety, compact footprints, and economic requirements logically demand system simplification, micro reactor designers seek to eliminate complexity throughout their systems [43]. This system simplification does not exclude the control system, which will inherently simplify with a shrinking control surface. The market and operational environment demand for automated control systems and remote operation [44] in coincidence with a smaller control surface imply that micro reactor control systems will be highly centralized.

If control systems are small and centralized there are significantly less possibilities to improve cybersecurity protection through diversity and segmentation. This cybersecurity posture is further degraded if an external connection for remote operations is necessary. Remote and autonomous operation represents a significant potential application for ZTA. If remote operation is ever to be feasible, secure communications will be the keystone to enabling this mode of operation. ZTA is a strong candidate to ensure that these communications are secure and confidential as remote access to secure environments is a major driver for current commercial ZTA development.

6. CONCLUSION

For the current fleet of reactors, implementing ZTA would come at a high cost of equipment replacement. Because these reactors have some inherent cybersecurity from their diversity and redundancy, there would be little benefit to implementing ZTA for the excessive cost of a full system replacement. Staged deployment as systems are upgraded in their natural replacement cycles could be a reasonable path if the equipment cost differential was not excessive. ZTA in the current fleet would be better suited to IT systems and improving the cybersecurity posture of administrative systems and systems in higher levels of the nuclear control system hierarchy as outlined by the IAEA.

For reactors that are yet to be constructed the implementation costs are significantly lower. With designs that intend to have centralized control rooms and streamlined control systems it will be necessary to consider the importance of secure communications. Designers will need to evaluate the potential for malicious operation and if the control systems have the capability to put any part of the plant in dangerous operational modes. Since the cost of implementation would lower than the current fleet, and some protection via diversity is lost, if the technology is available for ZTA implementation it may provide significant improvement to the cybersecurity posture of future reactors. Furthermore, these systems promise to have significantly smaller numbers of Level 1 systems. ZTA is most applicable to Level 2 systems and above, becoming more applicable and promising more of an impact the higher the level considered.

If remote operation is viable, it would require some method of highly secure communication architecture like ZTA. The cost to implement ZTA for remote operations would be less concerning than the critical need of cybersecurity in a remote operation application. The cost of R&D would be the critical factor for enabling this mode of operation.

Overall, the most significant technical hurdle to ZTA implementation is dynamic session-based system-to-system authentication. Incorporating this kind of authentication using techniques typically used today via remote device databases and certificate authorities creates additional unacceptable layers of complexity in control systems. This additional complexity imposes new single-point and common-cause failure risks and creates unpredictable increases in latency. These new failure risks and latency increases could potentially be eliminated with new engineering approaches, but this is a currently unexplored area of research.

ZTA approaches can increase system security overall but require careful and thoughtful application to be cost and functionally effective.

REFERENCES

[1] S. P. Marsh, "Formalising Trust as a Computational Concept," 1994.

[2] "The Open Group," [Online]. Available: https://www.opengroup.org/.

[3] J. Davis, M. Simos and J. Skoniecki, "Back to the Future: What the Jericho Forum Taught us about Modern Security," Microsoft, [Online]. Available: https://www.microsoft.com/en-us/security/blog/2020/10/28/back-to-the-future-what-the-jericho-forum-taught-us-about-modern-security/.

[4] "Jericho Forum Commandments," 2007.

[5] J. Forum, ""Identity" Commandments".

[6] "Forrester Research," [Online]. Available: https://www.forrester.com/bold.

[7] "The History and Evolution of Zero Trust," Securityweek, 11 July 2022. [Online]. Available: https://www.securityweek.com/history-and-evolution-zero-trust. [Accessed 27 September 2022].

[8] J. Kindervag, "No More Chewy Centers: Introducing The Zero Trust Model of Information Security," 2010.

[9] "BeyondCorp," Google, [Online]. Available: https://cloud.google.com/beyondcorp.

[10] S. Rose, O. Borchert, S. Mitchell and S. Connelly, "NIST Special Publication 800-207: Zero Trust Architecture," 2020.

[11] B. Osborn, J. McWilliams, B. Beyer and M. Saltonstall, "BeyondCorp: Design to Deployment at Google," Google, 2016.

[12] R. Ward and B. Beyer, "BeyondCorp: A New Approach to Enterprise Security," Google, 2014.

[13] C. DeCusatis, P. Liengtiraphan, A. Sager and M. Pinelli, "Implementing Zero Trust Cloud Networks with Transport Access Control and First Packet Authentication," 2016.

[14] Deloitte, "Achieving a Zero Trust Architecture in an Industrial Environment with Multiple Facilites," Deloitte, 2021.

[15] P. A. Grassi, J. P. Richer, S. K. Squire, J. L. Fenton, E. M. Nadeau, N. B. Lefkovitz, J. M. Danker, Y.-Y. Choong, K. K. Greene and M. F. Theofanos, "NIST Special Publication 800-63c: Digital Identity Guidelines," 2017.

[16] T. Ghosh, N. Kumar, S. M. Sakuru, P. Shirazi, M. Simos, A. Valani, A. Carrato, S. Whitlock, J. Hietala, J. Linford and A. Szakal, "Zero Trust Core Principles," The Open Group, 2021.

[17] S. East, J. Butts, M. Papa and S. Shenoi, "A Taxonomy of Attacks on the DNP3 Protocol," 2009.

[18] Modbus, "MODBUS/TCP Security," 2018.

[19] Microsoft, "Kerberos Authentication Overview," [Online]. Available: https://learn.microsoft.com/en-us/windows-server/security/kerberos/kerberos-authentication-overview.

[20] S. McCarty, "Architecting Containers Part 1: Why Understanding User Space vs. Kernel Space Matters," [Online]. Available: https://www.redhat.com/en/blog/architecting-containers-part-1-why-understanding-user-space-vs-kernel-space-matters.

[21] U.S. NRC, "Protection of digital computer and communication systems and networks.," [Online]. Available: https://www.nrc.gov/reading-rm/doc-collections/cfr/part073/part073-0054.html.

[22] Nuclear Energy Institute, "NEI 08-09 [Rev. 6]," 2010.

[23] Canadian Nuclear Safety Commission, "DIS-21-03, Cyber Security and the Protection of Digital Information".

[24] "IAEA Nuclear Security Series No. 17-T," International Atomic Energy Agency, Vienna, 2021.

[25] IAEA, "NSS 23-G: Security of Nuclear Information," 2015.

[26] IAEA, "NSS 33-T: Computer Security of Instrumentation and Control Systems at Nuclear Facilities," 2018.

[27] IAEA, "Computer Security at Nuclear Facilities, IAEA Nuclear Security Series No. 17," IAEA, Vienna, 2011.

[28] NIST, "NIST SP 800-53," National Institute of Standards and Technology, 2020.

[29] USNRC, "NUREG/CR-7007: Diversity Strategies for Nuclear Power Plant Instrumentation and Control Systems," Oak Ridge National Laboratory, Oak Ridge, 2008.

[30] USNRC, "NRC REGULATORY ISSUE SUMMARY 2002-22, SUPPLEMENT 1, CLARIFICATION ON ENDORSEMENT OF NUCLEAR ENERGY INSTITUTE GUIDANCE IN DESIGNING DIGITAL UPGRADES IN INSTRUMENTATION AND CONTROL SYSTEMS," USNRC, WASHINGTON, 2018.

[31] A. Dua, K. Ellingrud, P. Kirschner, A. Kwok, R. Luby, R. Palter and S. Pemberton, "Americans are embracing flexible work—and they want more of it," 23 June 2022. [Online]. Available: https://www.mckinsey.com/industries/real-estate/our-insights/americans-are-embracing-flexible-work-and-they-want-more-of-it.

[32] A. Bick, A. Blandin and K. Mertens, "Work from Home After the COVID-19 Outbreak," Federal Reserve Bank of Dallas, Dallas, 2020.

[33] IAEA, "IAEA-TECDOC-626: Safety related terms for advanced nuclear plants," IAEA, Vienna, Austria, 1991.

[34] H. Zhao, H. Zhang and L. Zou, "Use of DRACS to Enhance HTGRs Passive Safety and Economy," in *ANS 2011 Annual Meeting*, Hollywood, FL, 2011.

[35] S. Lee, Y. J. Choi, J. I. Lee and Y. H. Jeong, "Investigation of various reactor vessel auxiliary cooling system geometries for a hybrid micro modular reactor," *Nuclear Engineering and Design,* vol. 379, 2021.

[36] R. Fasano, C. Lamb and M. Rowland, "CYBER RISKS TO THE OPERATIONAL TECHNOLOGY ARCHITETURES OF NEXT GENERATION NUCLEAR REACTORS," OSTI, Albuquerque, 2021.

[37] Claroty Team 82, "CVE-2022-38465," 2022. [Online]. Available: https://claroty.com/team82/disclosure-dashboard/cve-2022-38465.

[38] Team82 Research, "The Race to Native Code Execution in PLCs: Using RCE to Uncover Siemens SIMATIC S7-1200/1500 Hardcoded Cryptographic Keys," 11 October 2022. [Online]. Available: https://claroty.com/team82/research/the-race-to-native-code-execution-in-plcs-using-rce-to-uncover-siemens-simatic-s7-1200-1500-hardcoded-cryptographic-keys.

[39] Nuscale, "NuScale Power, LLC Submittal of "Concept of Operations," RP-0215-10815, Revision 3," 10th May 2019. [Online]. Available: https://www.nrc.gov/docs/ML1913/ML19133A293.pdf.

[40] USNRC, "NuScale FSER Chapter 7: Instrumentation and Controls," USNRC, 2020.

[41] J. Stevens, K. LaFerriere and R. Flamand, "Small Modular Reactor Control Room Workstation Demonstration," in *Human Factors and Ergonomics Society 2019 Annual Meeting*, 2019.

[42] A. Clark, B. A. Beeny, K. C. Wagner and D. L. Luxat, "Technical and Licensing Considerations for Micro-Reactors," Sandia National Laboratories, Albuquerque, 2020.

[43] D. Shropshire, G. Black and K. Araújo, "Global Market Analysis of Microreactors," INL, 2021.

[44] J. Christensen, W. Moe, W. Poore and R. Belles, "Regulatory Research Planning for Microreactor Development," INL, 2021.

[45] S. Rose, O. Borchert, S. Mitchell and S. Connelly, "NIST Special Publication 800-207: Zero Trust Architecture," NIST, 2020.

[46] "Apple Platform Security: Secure Enclave," Apple, Inc., [Online]. Available: https://support.apple.com/guide/security/secure-enclave-sec59b0b31ff/web. [Accessed 30 September 2023].

DISTRIBUTION

Click here, then press delete to remove this guidance statement.
Required

On an odd-numbered page.

SAND Reports submitted through R&A are automatically sent to the Technical Library; however, it still needs to be included on the distribution.
Ensure a blank, odd-numbered page is included prior to the back cover.

Click here, then press delete to remove this guidance statement.
If emailing a copy internally, include the recipient's name, org., and Sandia email address. List in ascending order by org. number, then alphabetize by the recipient's last name. The Technical Library will not be listed by org. and will be the last entry. At a minimum, the Technical Library will be listed.

Email—Internal

Name	Org.	Sandia Email Address
Technical Library	1911	sanddocs@sandia.gov

Click here, then press delete to remove this guidance statement.
If emailing a copy externally, include the recipient's name, company email address, and company name. List by first name, then last name (e.g., John Doe), then alphabetize by the recipient's last name. Delete the table if not emailing externally.

Email—External

Name	Company Email Address	Company Name

Click here, then press delete to remove this guidance statement.
If sending a hardcopy internally, indicate the number of copies being sent and list the recipient's name, org., and mailstop. List mailstops in ascending order, then alphabetize by the recipient's last name. Delete the table if not sending hardcopy.

Hardcopy—Internal

Number of Copies	Name	Org.	Mailstop

Number of Copies	Name	Org.	Mailstop

Click here, then press delete to remove this guidance statement.
If sending hardcopies externally, indicate the number of copies being sent and list the recipient's name, company name, and full company mailing address. List by first name, then last name (e.g., John Doe), then alphabetize by the recipient's last name. Delete the table if not sending hardcopy.

Hardcopy—External

Number of Copies	Name	Company Name and Company Mailing Address

This page left blank

Sandia National Laboratories

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.