



Enhanced Zero Trust
and 5G

TABLE
OF
CONTENTS

Executive Summary	4
Foreword	5
1. Introduction	6
2. Terminology and Acronyms	7
3. Implications of ZTA for 5G	9
4. Evolution of Zero Trust	10
4.1 Relevant U.S. Government Guidance on ZTA	10
4.2 NIST SP 800-207 Zero Trust Architecture	10
4.3 Enduring Security Framework Security Guidance for 5G Cloud Infrastructures	11
4.4 ZTA Security Controls	12
4.5 CISA Zero Trust Maturity Model	12
5. Zero Trust 5G Use Cases	14
5.1 Use Case 1 – 5GS Control Plane (AS, NAS, SBI/SBA)	14
5.2 Use Case 2 – 5G User Plane (DRB, GTP)	15
5.3 Use Case 3 – Authentication and Authorization	16
5.4 Use Case 4 - Management Plane (OAM)	16
5.5 Use Case 5 – Public and Hybrid Cloud Deployments	17
5.6 Use Case 6 – Cloud-Native Workloads	18
5.7 Use Case 7 – Edge Applications	18
5.8 Use Case 8 – O-RAN Open Fronthaul	19
5.9 Use Case 9 – O-RAN SMO, RICs, and Apps	20
5.10 Use Case 10 – Artificial Intelligence (AI)/Machine Learning (ML)	20
5.11 Use Case 11 – Network Slicing	21
5.12 Use Case 12 – 5G Supply Chain	21
6. How Zero Trust Applies to 5G Today	23
6.1 Zero Trust in the 5GC	23
6.2 5GS Security Domains	23
6.2.1 5G Network Access Security (I)	24
6.2.2 5G Network Domain (II)	25
6.2.3 User Domain Security (III)	25
6.2.4 Application Domain Security (IV)	25
6.2.5 Service-Based Architecture (SBA) (V)	25
6.2.6 Visibility and Configurability of Security (VI)	25
6.3 Alignment of 5G to NIST ZT Tenets	25



TABLE
OF
CONTENTS

7.	ZTA for Cloud/Virtualization/Containerization	27
7.1	Cloud Infrastructure	27
7.2	Virtual Machines, Containers, and Kubernetes	27
7.3	Deployment Models (e.g., Public/Private Clouds)	31
8.	Implementation and Operational Considerations	32
8.1	Edge Deployments	32
8.2	Closed-Loop Automation	33
8.3	AI/ML	35
9.	ZT Policy Management	36
9.1	Introduction	36
9.2	Formal Logic in Zero Trust	36
9.3	Explainable AI in Zero Trust	36
9.4	NIST Policy Machine Extensions	37
9.5	Next-Generation Access Controls (NGAC) Extensions to the Policy Machine	37
9.6	Future Areas of Study for Policy Management	38
10.	Assessment of 3GPP SA3 Zero Trust Security Activities	39
10.1	Introduction	39
10.2	Gap Analysis	39
10.3	Consolidated Recommendations for 3GPP	41
10.3.1	General Recommendations	41
10.3.2	Tenet 1 – All Data Sources and Computing Services are Considered Resources	41
10.3.3	Tenet 2 – All Communications is Secured Regardless of Network Location	41
10.3.4	Tenet 3 – Access to Individual Resources is Granted on a Per-Session Basis	41
10.3.5	Tenet 4 – Access to Resources is Determined by Dynamic Policy	41
10.3.6	Tenet 5 – Monitoring/Measuring the Integrity and Security Posture of all Assets	41
10.3.7	Tenet 6 – All Resource Authentications/Authorizations are Dynamic and Strictly Enforced Before Access is Allowed	42
10.3.8	Tenet 7 – Improving the Security Posture of all Assets, Network Infrastructure, and Communications	42
11.	Conclusions and Next Steps	43
	References	44
	Copyright and Disclaimer	46



Zero trust (ZT) is a concept that no digital system or human user, whether external or internal, can be trusted, regardless of ownership and location. ZT architecture (ZTA) is a plan to implement ZT in a digital system or network of digital systems. ZTA is based upon two core principles:

1. No digital system can be implicitly trusted based upon its ownership or location.
2. Perimeter security alone is insufficient. Each digital system, as an asset, must be secured as a micro-perimeter.

U.S. NIST guidance for a ZTA is general to digital systems [1]. The NSA Enduring Security Framework (ESF) and CISA "Security Guidance for 5G Cloud Infrastructures" offer best practices to "bring a Zero Trust mindset into 5G cloud endpoints and growing multi-cloud environments" [2]. Relevant industry bodies for 5G, specifically 3GPP and O-RAN Alliance, are in the process of forming requirements that align with a ZTA. ATIS convened a study group to assess zero trust for 5G with the goals to form relevant requirements, identify potential gaps, and recommend areas for standardization.

The ATIS 5GZT study was informed by the work at NIST, CISA, and 3GPP and subject matter experts on zero trust from organizations that are stakeholders in 5G network security. There are 10 key findings of the ATIS study:

1. ZTA is a plan based upon the concept of zero trust. It is important that 5G Systems (5GS), as critical infrastructure, strive toward the goal of a ZTA.
2. There are multiple use cases for zero trust in 5G as the standards continue to evolve.
3. Multiple U.S. federal agencies are addressing zero trust. The relevant agencies for 5G zero trust are WH ONCD, DoC NIST, DHS CISA, and the NSA ESF
4. Enduring Security Framework's "Security Guidelines for 5G Cloud Infrastructures" provides a playbook for adapting NIST ZTA to 5G. The guidance in this document should be considered by 5G standards bodies.
5. Each of the NIST seven tenets for ZT can be applied to a 5G ZTA. Different industry bodies may have the scope for any of the tenets. The seven tenets are relevant for the end-to-end 5GS, including RAN and Core. 3GPP should expand its consideration of ZTA beyond the 5G Core (5GC) and also encompass the 5G RAN and potentially the UE, for which ZTA is applicable.
6. NIST's ZT logical components include a Policy Decision Point (PDP) and Policy Enforcement Point (PEP) that can be mapped into existing 5G network functions (NFs) and implemented as logical functions within a network function acting as a micro-perimeter.
7. 5G, as specified by 3GPP, is the most secure generation of mobile technology to date. Many security features of 5G align with a ZTA. Further evolution of mobile technologies is expected to evolve toward a ZTA, beginning with 6G.
8. 5G ZTA is characterized by 12 Security Control Groups. This is an opportunity for further standardization. Areas for further study are Continuous Monitoring, Anomalous Behavior Detection, Policy Management, TDR/EDR, and Threat Intelligence.
9. Cloud security best practices are evolving to support the security needs of 5G and other critical infrastructure.
10. Enhanced security capabilities are needed to support security operations in the ZTA environment.

FOREWORD

As a leading technology and solutions development organization, the Alliance for Telecommunications Industry Solutions (ATIS) brings together the top global ICT companies to advance the industry's most pressing business priorities. ATIS' nearly 200 member companies are currently working to address the all-Internet Protocol (IP) transition, 5G, NF virtualization, big data analytics, cloud services, device solutions, emergency services, M2M, cyber security, network evolution, quality of service, billing support, operations, and much more. These priorities follow a fast-track development lifecycle – from design and innovation through standards, specifications, requirements, business use cases, software toolkits, open source solutions, and interoperability testing.

ATIS is accredited by the American National Standards Institute (ANSI). The organization is the North American Organizational Partner for the 3rd Generation Partnership Project (3GPP), a founding Partner of the oneM2M global initiative, a member of and major U.S. contributor to the International Telecommunication Union (ITU), as well as a member of the Inter-American Telecommunication Commission (CITEL). For more information, visit www.atis.org.





1. INTRODUCTION

On March 2, 2023, the U.S. White House Office of National Cyber Director (ONCD) announced the U.S. National Cybersecurity Strategy [3], which includes reference to zero trust architecture (ZTA):

"This Administration is committed to improving Federal cybersecurity through long-term efforts to implement a zero trust architecture strategy and modernize IT and OT networks."

ZT is a concept that no digital system or human user, whether external or internal, can be trusted, regardless of ownership and location. Over the past 2+ years, multiple U.S. government agencies have been addressing ZTA for enterprise, critical infrastructure, and national defense networks. Recently, standards bodies for mobile communications – including ATIS, 3GPP, and the O-RAN Alliance – have been addressing the application of ZTA to 5G networks.

The purpose of this paper is to provide guidance and recommendations to achieve a ZTA in U.S. deployed 5G networks and drive enhanced standardizations at the relevant standards bodies. The scope of this paper is 5G Standalone (SA) networks, including Open RAN. Non-Standalone (NSA) networks are considered out of scope for this paper and are not addressed.

This paper first presents zero trust terminology with focus on "ZT," "ZTA," and "Zero Trust Network Access (ZTNA)." Implications and use cases of ZTA for 5G networks are then discussed. The evolution of guidance on ZT from U.S. government agencies is then discussed to provide context for further analysis provided in the subsequent sections. Alignment of 5G security standards to the NIST seven tenets for ZT is discussed and a gap analysis is provided. Recommendations to achieve a ZTA in 5G and future generations of mobile technologies are provided.

ZERO TRUST ARCHITECTURE



2. TERMINOLOGY AND ACRONYMS

This section focuses on the ZT terminology used in this paper. It is assumed that the reader is knowledgeable about general industry cybersecurity terms.

ZT is a concept that no digital system or human user, whether external or internal, can be trusted, regardless of ownership and location. NIST refers to its ZTA as a plan that enables a digital system and network of digital systems to have built-in ZT. ZTA complements trust domains and traditional perimeter defenses with micro-perimeters at each asset as a foundation for a defense-in-depth strategy to protect from external and internal threats. This is a paradigm shift for securing telecommunications networks, which have traditionally been

secured using perimeter defenses (e.g., packet gateways, SS7 firewalls, GTP firewalls), out-of-band management networks that leverage centralized logins and local accounts, and secure network management protocols (e.g., SNMPv3, SSHv2, TLS 1.2 or 1.3) to protect from external and internal threats. These traditional security controls provide some level of security but do not provide the granularity nor capability needed to truly align with ZTA.

There are many industry terms and acronyms used in discussion of ZT. Table 1 provides a summary of the most accepted terms. This paper uses the terms ZT, ZTA, and ZTNA as defined in Table 1 below.

Terms	Definitions	Sources/Inputs
Zero Trust (ZT)	<p>ZT is a concept that no digital system or human user, whether external or internal, can be trusted, regardless of ownership and location. No network user, packet, interface, or device should be assumed to be trusted.</p> <p>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.</p>	<ul style="list-style-type: none"> > Three-Technical-Innovations-Will-Ignite-Zero-Trust.pdf (omniapartners.com), J. Kindervag and A. Kindness, Forrester, May 2015 > The Tao Of Zero Trust (forrester.com), Forrester, March 2019 > U.S. NIST SP 800-207, August 2020
Zero Trust Architecture (ZTA)	<p>A plan that provides protection from external and internal threats with the assumption that a threat actor has established a foot-hold in the network. ZTA is an evolving set of cybersecurity paradigms that move defenses from static, network-based perimeters to focus on perimeter-less security, or micro-perimeters, for users, assets, and resources. There is no implicit trust granted to an asset based upon ownership, physical location, or network location. A ZTA includes security controls for confidentiality, integrity, and availability at multiple layers.</p>	<ul style="list-style-type: none"> > U.S. NIST, NIST SP 800-207, August 2020 > U.S. NSA ESF and DHS CISA, Security Guidance for 5G Cloud Infrastructures, Oct and Nov, 2021.
Zero Trust Network Access (ZTNA)	<p>A product or service that provides a software-defined perimeter to ensure secure access to a digital resource (asset or application) by establishing an identity and context-based logical access boundary around an application or set of applications. External (remote) access to a digital resource is permitted only for trusted, authenticated, and authorized identities using the principle of least privilege defined by granular policies</p>	<ul style="list-style-type: none"> > Gartner, Gartner Glossary > Zscaler, Security Terms Glossary, "What is ZTNA?" > Palo Alto Networks, Cyberpedia, "What is ZTNA?"
Zero Trust Security (ZTS)	<p>Continuous monitoring and validation of the user identity and privileges to access resources based upon the least privilege principle and that no one is trusted from inside or outside the network. Also used as a general term for the ZT concept, ZTNA product or service, or ZTA plan.</p>	<ul style="list-style-type: none"> > Cloudflare, Glossary, "What is a Zero Trust Network?"
Zero Trust Networking (ZTN)	<p>Continuous authentication and monitoring to access a network or network resources based upon the principle that no one is trusted from inside or outside the network. Also used as a general term for the ZT concept, ZTNA product or service, or ZTA plan.</p>	<ul style="list-style-type: none"> > Cisco, "What is Zero-Trust Networking?"

Table 1. ZT Terminology and Acronyms

More Terminology

This paper focuses on ZTA in the context of 5G networks. A discussion of ZTA requires a clear definition of external threats and internal threats for 5G networks. The definitions for external and internal threats used in this paper are provided below.

External Threats:

Definition: An unauthorized entity outside the security domain that has the potential to harm an information system through destruction, disclosure, modification of data, and/or denial of service. Also referred to as Outsider Threat [adapted from NIST].

Example:

- > False Base Station
- > Botnet
- > Supply Chains (hardware and software)

Internal Threats:

Definition: The threat that an insider will use their authorized access, wittingly or unwittingly, to do harm to the security of organizational operations and assets, individuals, and other organizations, and the nation. This threat can include damage through reconnaissance, unauthorized disclosure of sensitive information, destruction, or loss or degradation of organizational resources or capabilities. Also referred to as Insider Threat [adapted from NIST].

Examples:

- > Host operator gains access to tenant's operator data.
- > Employee of third-party cloud provider gains access to tenant's data and steals it for financial gain.
- > Adversarial nation-state actor within the network can move laterally to perform reconnaissance undetected over an extended period. This is known as an Advanced Persistent Threat (APT) and is a case where an External Threat becomes an Internal Threat.
- > A malicious insider colludes with an external threat actor to misconfigure security to allow external entry to a network.



3. IMPLICATIONS OF ZTA FOR 5G

Creating a ZTA for 5G requires careful consideration of the 5G architecture's unique aspects. First, the 5G architecture has three distinct network traffic planes: user, control, and management. Different types of network interfaces and protocols are used to transmit data across each of these planes. The user and control planes are 3GPP-defined interfaces that use protocols optimized for real-time telecommunications performance. The real-time requirements of these protocols, especially the control interfaces related to signaling, require innovative solutions to achieve ZT. On the other hand, management protocols are similar to standard information technology (IT) management interfaces and can be protected using technologies such as SSH, TLS 1.2/1.3, and strong authentication coupled with role-based access control (RBAC) enforcing least privilege access.

In addition to the three distinct network traffic planes, the 5G architecture is divided into three major domains: Core, Radio Access Network (RAN), and User Equipment (UE). The Core operates from within a data center and a mobile switching office. Traditionally, the Core was considered secure, making the overhead of encrypting the transport within the Core or authenticating non-management intra-Core transactions unnecessary. The "trusted core" is no longer a valid assumption because ZT assumes that adversaries are present inside every part of a network. Instead, authentication, authorization, and confidentiality, along with their overhead, are required to achieve a ZTA.

The RAN has traditionally relied upon perimeter-based security and physical security measures in a castle-and-moat approach. In previous generations, the RAN was closed, meaning that the RAN infrastructure was run on a vendor's proprietary hardware and software. Security controls were placed on management interfaces, and the use of SIM cards in the UE prevented unauthorized device access to the cellular network. The broad global adoption of cloudification in the 5GC is extending to the RAN. Open RAN introduces disaggregation, open interfaces, and deployment models that require a ZTA, as well. ZTA controls that do not adversely affect RAN functions, such as time synchronization, will have to be developed for the new RAN architectures.

Operators have far less control of UEs once these devices are in the end user's possession. Two exceptions are Firmware Over-the-Air (FOTA) updates, where the mobile device's firmware is wirelessly upgraded by its manufacturer and updates to the UE SIM by the mobile carrier. In both cases, the UE must verify the identity of the device manufacturer and/or the mobile carrier.

USER, CONTROL,
& MANAGEMENT
three distinct network traffic planes



4. EVOLUTION OF ZERO TRUST

4.1 Relevant U.S. Government Guidance on ZTA

Over the past 2+ years, multiple U.S. government agencies have been addressing how to apply ZTA to enterprise networks, critical infrastructure, and national defense. U.S. NSA ESF and DHS CISA have adapted NIST's ZTA to form guidelines for securing 5G critical infrastructure. The evolution of U.S. Government guidance on ZTA for 5G critical infrastructure is summarized in Figure 4-1.1. The NIST SP 800-207 is referenced throughout this document and described further in a subsection below. The CISA publications are also described further in the subsections below. Standards bodies for mobile communications – such as ATIS, 3GPP, and the O-RAN Alliance – should consider this guidance as they continue to evolve security standards for ZTA in 5G networks.

US Government Guidance on ZTA in the 5G Cloud

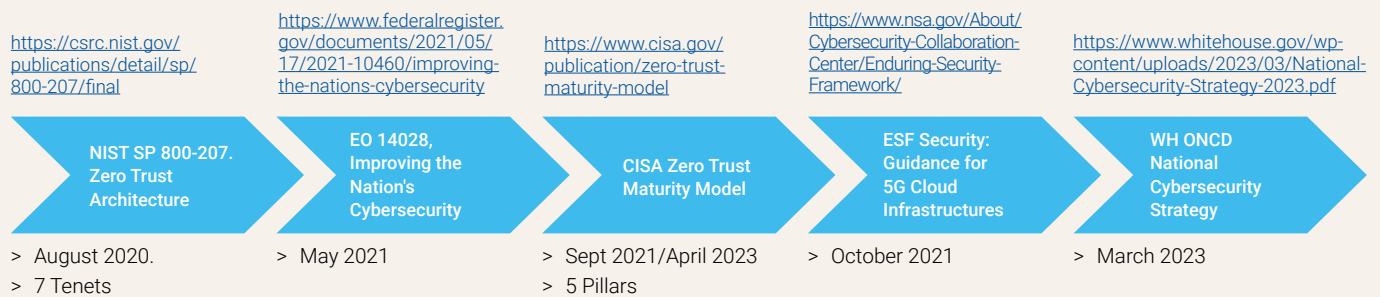


Figure 4.1-1. Evolution of U.S. Government Publications on ZTA Relevant for 5G

4.2 NIST SP 800-207 Zero Trust Architecture

NIST's SP 800-207 Zero Trust Architecture (ZTA) [4] is a plan that enables a digital system and network of digital systems to operate securely based upon the following seven tenets of ZT:

List 4-1. NIST Seven Tenets of ZT [4]

- T1. All data sources and computing services are considered resources.
- T2. All communication is secured regardless of network location.
- T3. Access to individual resources is granted on a per-session basis.
- T4. Access to resources is determined by dynamic policy.
- T5. The operator monitors and measures the integrity and security posture of all owned and associated assets.

- T6. All resource authentication and authorization are dynamic and strictly enforced before access is allowed.
- T7. The operator collects information about the current state of assets, network infrastructure, and communications and uses it to improve its security posture.

These tenets evolve the traditional paradigm of a perimeter-based defense that previously protected against external threats by removing the assumption of trusted assets and trust boundaries. This is similar to the treatment of internal threats. The ZTA treats each asset as a micro-perimeter to protect against internal and external threats. This enables the ZTA to detect and prevent lateral movement. The alignment of 5G security standards to the NIST seven tenets is discussed later in this document. NIST SP 800-207 also provides a logical architecture to implement a ZTA, as shown in Figure 4.2-1.

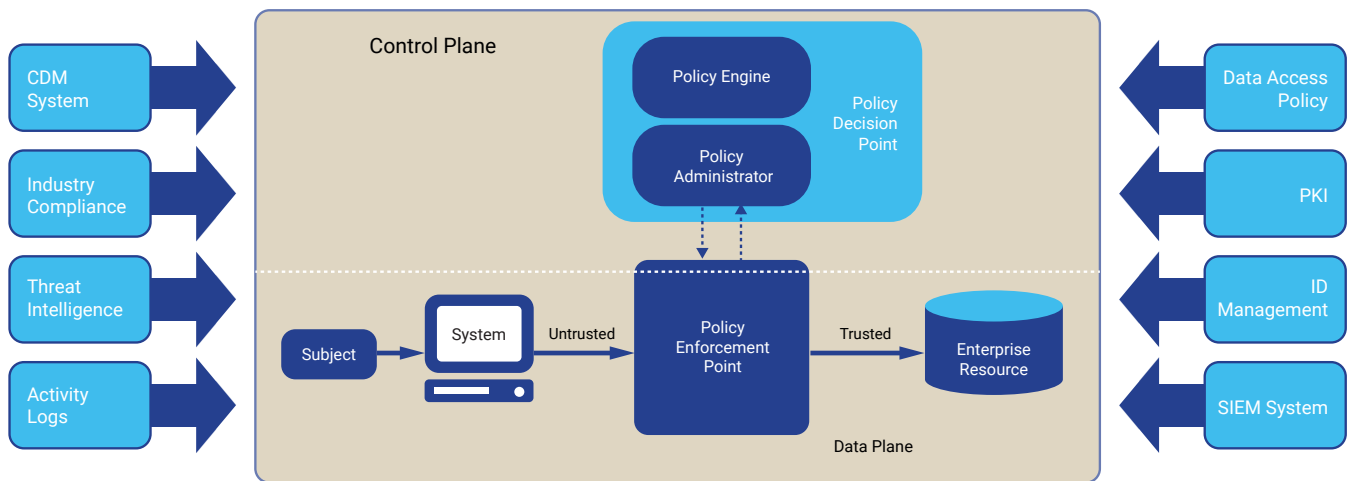


Figure 4.2-1. NIST Logical Architecture for a ZTA [4]

The data plane is where the Subject Actor requests access to a Target Actor using a PEP. The control plane contains the logic necessary to make the access request decision using a PDP. The PEP and PDP are logical entities that may be standalone functions or logical functionality integrated within micro-perimeters. Many different factors are used to decide whether to grant access, such as threat intelligence. Similarly, other tools and applications, such as the management of Actor IDs, are required by the PDP. The “trusted” portion of the system, called a micro-perimeter, is to the right of the PEP and must be as small as possible to enable the decision of the PDP to be as specific as possible.

The Policy Engine (PE) is responsible for the ultimate decision to grant access to a resource for a given subject. The PE uses enterprise policy and input from external sources (e.g., continuous diagnostics and mitigation systems, activity logs, and threat intelligence services) as input to a trust algorithm to grant, deny, or revoke access to a Target Actor. The trust algorithm is a function of information about Subject and Target Actors, including their attributes and roles, historical subject behavior patterns, threat intelligence sources, and other metadata.

The NIST ZTA can also determine the confidence in Subject and Target Actors. For example, at each enforcement point, logs could be sent to the Security Information and Event System and analytics performed on the logs to develop a confidence level for a particular Actor. Confidence levels of different Actors can then be aggregated as appropriate for policy enforcement. The Trust Algorithm can use a set of criteria that all must be met, or a weighted set of criteria that contribute to an overall score, or contextually appropriate criteria for each of the above cases, or other variations.

The PE is paired with the Policy Administrator (PA). The PE makes and logs the decision (as approved or denied), and the PA executes the decision. The PA is responsible for establishing and/or shutting down the communication path between a Subject Actor and a Target Actor via commands to relevant PEPs. The PEP is responsible for

enabling, monitoring, and eventually terminating connections between a Subject and a Target Actor. The PEP is shown as a single logical component in ZTA but may be broken into two different components: the client and a component that provides access to the Target Actor. Access to a Target Actor is determined by dynamic policy.

4.3 Enduring Security Framework Security Guidance for 5G Cloud Infrastructures

In October and November of 2021, the DHS CISA and NSA Enduring Security Framework (ESF) published a four-volume set of documents titled “Security Guidance for 5G Cloud Infrastructures” that offered best practices to “bring a Zero Trust mindset into 5G cloud endpoints and growing multi-cloud environments.” [2] This was the first publication connecting 5G, Cloud, and ZTA. The four volumes, as listed below, provide a playbook for secure deployment of 5G critical infrastructure in the cloud.

Security Guidance for 5G Cloud Infrastructures:

- > Part I: Prevent and Detect Lateral Movement [2]
- > Part II: Securely Isolate Network Resources [5]
- > Part III: Protect Data in Transit, In-Use, and at Rest [6]
- > Part IV: Ensure Integrity of Infrastructure [7]

These documents include some statements that highlight the importance of having a ZTA for 5G critical infrastructure:

- > **“Cloud-native 5G is a lucrative target** for cyber threat actors.”
- > “A characteristic of cloud infrastructure that presents a significant security challenge in 5G is **multitenancy**.”
- > “Zero Trust is the concept that **perimeter defenses are no longer sufficient** to secure a network.”
- > **“Strive to bring a Zero Trust mindset into 5G cloud.”**

4.4 ZTA Security Controls

The ZTA principles discussed in the sections above are consolidated in this paper in the form of a set of twelve fundamental Security Control Groups to achieve a ZTA for 5G networks. These are:

1. Continuous Monitoring, Logging, and Alerting
2. IAM, including dynamic access control policies and the principle of least privilege
3. Multi-Factor Authentication (MFA) for human users
4. Security Information Event Management (SIEM)/ Security Orchestration, Automation, and Response (SOAR) integration
5. Anomalous Behavior Detection, using artificial intelligence/machine learning (AI/ML)
6. Threat and Endpoint Detection and Response (TDR/EDR)
7. PKI-based Mutual Authentication for machine-to-machine communications
8. Secure software development based upon the DevSecOps [8] including continuous integration/continuous deployment (CI/CD) [9], and NIST Secure Software Development Framework (SSDF) [10]
9. Network micro-segmentation and micro-perimeters
10. Sensitive Data Encryption for data in motion, data at rest, and data in use
11. Threat Intelligence
12. Automated Security Testing/Configuration Validation

Table 4.4-1 provides a matrix showing where these Security Controls Groups are covered in relevant U.S. government documents providing guidance for ZTA.

Zero Trust Architecture - Security Controls	1	2	3	4	5	6	References:
Continuous Monitoring, Logging, and Alerting	X	X	X	X	X	X	1. EO 14028
IAM, including dynamic access control policies and Principle of Least Privilege		X	X	X	X	X	2. NIST SP 800-207 Zero Trust Architecture (ZTA)
Multi-Factor Authentication (MA)	X		X	X	X	X	3. CISA Cloud Security Technical Reference Architecture
SIEM/SOAR integration		X			X	X	4. NSA ESF/CISA Security Guidance for 5G Cloud Infrastructures
Anomalous Behavior Detection, using AI/ML and DPI			X		X	X	5. OMB Zero Trust Cybersecurity Principles
Threat and Endpoint Detection and Response (TDR/EDR)	X				X		6. Department of Defense Zero Trust Reference Architecture
PKI-based Mutual Authentication		X		X			
DevSecOps, CI/CD, and NIST SSDF			X			X	
Network micro-segmentation and micro-perimeters				X		X	
Sensitive Data Encryption for data (at-rest, in-motion, in-use)	X					X	
Threat Intelligence		X					
Automated Security Testing/Configuration Validation			X				

Table 4.4-1. Alignment of ZTA Security Controls Groups

4.5 CISA Zero Trust Maturity Model

The U.S. Department of Homeland Security's (DHS) Cybersecurity and Infrastructure Security Agency (CISA) Zero Trust Maturity Model (ZTMM) [11] support U.S. federal civilian agencies in planning and implementing a ZTA as directed in Executive Order 14028 "Improving the Nation's Cybersecurity" [12]. The CISA ZTMM reflects the NIST seven tenets of ZT.

CISA defines five pillars in ZTMM: Identity, Devices, Networks, Applications and Workloads, and Data. Visibility and analytics, automation and orchestration, and governance are three cross-cutting capabilities required across the five pillars.

NOTE: The DoD has also produced a Zero Trust Strategy [13] with seven pillars, adding "Automation and Orchestration" and "Visibility and Analytics." CISA covers the same capabilities across its five pillars and three cross-cutting capabilities.

CISA ZTMM defines four stages of maturity:

0. **Traditional stage**, where manual configuration and static security policies dominate. Policy enforcement is proprietary and inflexible, and incident response and mitigation are manual.
1. **Initial** is the first step from traditional security to a ZTA with the introduction of cross-pillar automation, least privilege access policies, and aggregated visibility of internal systems.
2. **Advanced stage**, where one sees cross-pillar coordination, centralized visibility and identity control, policy enforcement based on some cross-pillar inputs and outputs, and incident response begins to utilize pre-defined mitigations.
3. **Optimal stage** with full automation, dynamic policies with context for decisions coming from across the pillars and based on automated triggers, alignment with open standards for cross-pillar interoperability, and centralized visibility with the ability to look back in time across the enterprise.

The ZTMM intends for each organization to implement incremental changes over time as they advance in their ZTA deployments. The implementation of ZTA-related technology and processes for each pillar can move forward in the maturity model at its own pace until cross-pillar coordination is required. As CISA states, "The path to zero trust is an incremental process that may take years to implement" [14], which is an acknowledgement of complexity, time, and cost incurred to achieve a ZTA.



5. ZERO TRUST 5G USE CASES

ZT is relevant for the entire 5GS, including 5GC, RAN, and UEs. The following use cases for 5G ZTA are discussed further below.

- > Use Case 1 – 5GS Control Plane (AS, NAS, SBI/SBA)
- > Use Case 2 – 5GS User Plane (DRB, GTP)
- > Use Case 3 – Authentication and Authorization
- > Use Case 4 – Management Plane (OAM)
- > Use Case 5 – Public and Hybrid Cloud Deployments
- > Use Case 6 – Cloud-Native Workloads
- > Use Case 7 – Edge Applications
- > Use Case 8 – O-RAN Open Fronthaul
- > Use Case 9 – O-RAN SMO, RICs, and Apps
- > Use Case 10 – Artificial Intelligence (AI)/Machine Learning (ML)
- > Use Case 11 – Network Slicing
- > Use Case 12 – 5G Supply Chain

5.1

Use Case 1 – 5GS Control Plane (AS, NAS, SBI/SBA)

Use Case ID	UC-1
Use Case Name	Control Plane (Access Stratum, Non-Access Stratum, Service Based Interfaces/Service Based Architecture)
Description	The user devices interact directly and indirectly with the RAN and Core Network to establish a wireless session, when handing over to a new cell site, when requesting to setup a data session, etc. This communication goes over the 3GPP-defined Control Plane, which includes the Access Stratum via signaling radio bearers, Non-Access Stratum, and the SBI/SBA in the Core Network. These control plane messages can contain sensitive data relative to the user such as SUPI/SUCI, fine-grain location data, call destination, etc.
Risk Types	Internal and External
Risk Summary	Access to the control plane messages could allow a threat actor to exploit the privacy of the user, allow for the hijacking of a user’s session, and allow for theft of services by allowing a non-customer to potentially establish a data session. For instance, a threat actor could attempt to capture signaling session packets using a passive wireless capture tool. In regard to 5G, 3GPP has defined mandatory security specifications that require Confidentiality and Integrity ciphering on the Access Stratum, Non-Access Stratum, and SBI/SBA communication flows.
Impact	These attacks can impact confidentiality, integrity, and availability. Denial-of-service attacks mainly impact availability. Man-in-the-middle and configuration attacks can affect all three.
Benefits of ZT	The control plane messages and communications meet several of the ZT tenets today, including that all communications must be secured and that all users are securely authenticated and authorized between the user’s device and the Core Network. There are some implementation options within 3GPP that do allow relaxing some of these specifications. Traditional perimeter defenses and system-overload protections should continue to be used to complement ZTA for defense against external attacks on availability.
Existing Mitigations	<ul style="list-style-type: none"> > 3GPP requires 5GS implementations to support confidentiality, integrity, and replay protection of control plane signaling data on AS, NAS, SBI, and N2 interfaces.
Potential Mitigations	<ul style="list-style-type: none"> > The use of confidentiality, integrity, and replay protection on control plane interfaces is required, with the exception of the AS and NAS interfaces, where confidentiality is optional but recommended for use by FCC CSRIC VII [14].

- > Subscriber identity privacy is protected via the use of the SUCI identifier, which is recommended for use by FCC CSRIC VII [14], and the 5G-GUTI identifier.

Additional mitigations may also be implemented:

- > Disabling the null ciphering scheme for both Confidentiality and Integrity algorithms for the Access Stratum and Non-Access Stratum.
- > Network segmentation.
- > Security configuration hardening and configuration management.
- > Security monitoring to detect the use of null ciphers.

5.2

Use Case 2 – 5G User Plane (DRB, GTP)

Use Case ID	UC-2
Use Case Name	5G User Plane (Data Radio Bearers/GPRS Tunneling Protocol)
Description	The user devices interact directly and indirectly with the RAN and data gateways to establish a wireless packet data session. This communication goes over the 3GPP-defined User Plane, which includes the Access Stratum via radio data bearers (DRB) and GPRS Tunneling Protocol (GTP). When a user accesses the internet for any reason, their data is sent via DRBs and over a GTP session. These User Plane messages can contain sensitive data relative to the user.
Risk Types	Internal and External
Risk Summary	Access to the User Plane messages could allow a threat actor to capture credentials, credit card information, sensitive emails, etc. For instance, a threat actor could attempt to capture the data session packets using a passive wireless capture tool. In regard to 5G, 3GPP has defined mandatory security specifications that require Confidentiality and Integrity ciphering on the Access Stratum's DRB messages, but the Integrity protection was only made mandatory in 3GPP Release 16. This integrity protection prevents the data from being manipulated by a man-in-the-middle attack. The data session between the 5G radio and the User Plane Function (UPF) uses GTP, which only encapsulates the data rather than encrypting it. If the user's data session is using end-to-end encryption (e.g., TLS, IPsec, etc.), then that session may be fully encrypted inside of the encapsulated GTP messages. The concern level rises when the user is not using end-to-end encryption accessing Internet-hosted applications and/or their enterprise data network.
Impact	These attacks can impact confidentiality, integrity, and availability. Denial-of-service attacks mainly impact availability. Man-in-the-middle and configuration attacks can affect all three.
Benefits of ZT	The data plane messages and communications meet some of the ZT tenets today, including that all wireless communications must be secured via confidentiality and integrity algorithms. Traditional perimeter defenses and system-overload protections should continue to be used to complement ZTA for defense against external attacks on availability.
Existing Mitigations	<ul style="list-style-type: none"> > 3GPP TS 33.501 User Plane confidentiality protection at the PDCP layer of the DRB [15].
Potential Mitigations	<ul style="list-style-type: none"> > Confidentiality and Integrity algorithms between the 5G radio and the UPF to either eliminate or mitigate the known vulnerabilities of GTP. > Security configuration hardening and configuration management to prevent accidentally disabling the available security controls. > Security monitoring to detect the use of null ciphers as a security anomaly.

5.3 Use Case 3 – Authentication and Authorization

Use Case ID	UC-3
Use Case Name	Authentication and Authorization
Description	Authentication is concerned with an entity proving its identity to another entity to establish a trust relationship. Authorization is concerned with ensuring an authenticated entity gets access only to those services and/or resources they are allowed to use. End users that consume 5GS network services (e.g., voice, data and video) are initially mutually authenticated with the network and subsequently authorized by the network to use the services subscribed to. Similarly, 5GS NFs that consume the services of other NFs mutually authenticate each other and authorize the services allowed to be consumed between each other.
Risk Types	Internal and External
Risk Summary	Entities of a 5GS network, users or NFs, if unauthenticated and unauthorized, may access and consume scarce and valuable capacity and resources of a 5GS network.
Impact	These attacks deprive the network operator of service revenue, reduce available network capacity, potentially degrade service experiences for legitimate network users, and could even lead to denial of service. Trust is eroded, which impacts the network operator’s reputation and business.
Benefits of ZT	ZT brings a paradigm of continuous (re)authentication and (re)authorization of access to and use of 5GS network resources. Continuous monitoring of dynamic information (e.g., location, time of day, and device type) enables (re)evaluation of authentication and authorization policies.
Potential Mitigations	<ul style="list-style-type: none">> Use 5GS Network Access Re-Authentication and Re-Authorization (frequency needs to be balanced against performance and service impacts).> Adopt a least privileges approach to authorize services toward end users and toward NFs controlled, for example, by subscription and policy data.> Use secondary authentication and authorization with external data networks.> Use network-slice-specific authentication and authorization.> Use mutual authentication between 5GS NFs.> Use of monitoring and analytics of user/NF behavior as input to policy decisions to (re)evaluate granting/revoking access to and use of resources.

5.4 Use Case 4 - Management Plane (OAM)

Use Case ID	UC-4
Use Case Name	Management Plane (OAM)
Description	All of the RAN and Core Network assets deployed to deliver 5G services will be managed via dedicated management interfaces and/or a separate management network. These interfaces and networks will support the required Mobile Network Operator (MNO)’s Operations, Administration, and Management (OAM) traffic. The MNOs will use OAM Operational Support Systems (OSS) and other tools from centralized locations to manage the network and the individual assets so that the 5G services and the customer experiences are being delivered as expected. The OAM network(s) are not typically exposed to the internet and have strong perimeter security around them.
Risk Types	Internal
Risk Summary	The OAM networks use traditional perimeter security and have historically used a trust domain between the deployed assets and some of the centralized OSS platforms. The OAM traffic will consist of common protocols such as SSH, TLS, SNMP, Netconf, etc. Misconfigurations could be exploited by an insider.

Impact	These attacks can impact confidentiality, integrity, and availability. Denial-of-service attacks mainly impact availability. Man-in-the-middle and configuration attacks can affect all three.
Benefits of ZT	The OAM network is not unique to the MNO. The same management architectures are used in all enterprises. As a result, ZT was designed for OAM networks. ZT can complement the perimeter security as part of a defense-in-depth cybersecurity strategy. Traditional perimeter defenses and system-overload protections should continue to be used to complement ZTA for defense against external attacks on availability.

Potential Mitigations

- > Developing, implementing, and maintaining a mature IAM program
- > Enforce least privilege using RBAC
- > MFA for management users for OAM
- > Network segmentation
- > Security configuration hardening and configuration management
- > Security monitoring for system-level events (e.g., MITRE's ATT&CK framework)
- > Use only secure management protocols (e.g., SSHv2, Netconf, SNMPv3, SFTP, TLS 1.2 or higher, and deprecate outdated protocols that use insecure algorithms (e.g., HTTPS, TLS 1.0) and/or communicate using clear text (e.g., telnet, HTTP).

5.5 Use Case 5 – Public and Hybrid Cloud Deployments

Use Case ID	UC-5
Use Case Name	Public and Hybrid Cloud Deployments
Description	Cloudification of the 5GC and RAN have enabled deployments of 5G critical infrastructure in hybrid and public clouds.
Risk Types	Internal and External
Risk Summary	Traditional network infrastructure has been deployed within the operator's premises where the operator owns, manages, and controls all assets, including facility, infrastructure, NFs, and data. The cloud introduces third parties that may own, manage, and control facilities, NFs (applications), and/or infrastructure.
Impact	The introduction of third parties increases threats from internal and external attacks on 5G critical infrastructure. External and internal threat actors could gain access to 5G virtual or cloud-native NFs through the infrastructure they own or manage to perform confidentiality, integrity, and availability attacks on the network. The impact of these attacks could include outages for the 5G voice and data services, performance degradation, unauthorized reconnaissance, and data theft.
Benefits of ZT	ZTA ignores legacy implicit trust granted based upon asset ownership, physical location, and network location. Instead, it introduces security controls that protect against internal and external threats introduced by third parties in cloud deployments explicitly.
Potential Mitigations	<ul style="list-style-type: none"> > Incorporate only secure-by-design and secure-by-default software products > Micro-segmentation and isolation > Continuous SBOM(s) delivery for release/patch vulnerability scanning > Configuration management > MFA for management users > OAuth 2.0 for digital systems authorization > mTLS with PKI and X.509 on network interfaces > Encryption of data in transit and data at rest, using NIST-approved cipher suites

- > API Security
- > Vulnerability assessments and penetration testing programs
- > Continuous monitoring, logging, and alerting at the cloud platform and application layers
- > Patch-management program

5.6

Use Case 6 – Cloud-Native Workloads

Use Case ID	UC-6
Use Case Name	Cloud-native workloads
Description	Deployment of 5G network functions as cloud-native workloads
Risk Types	Internal and External
Risk Summary	Cloud-native workloads and applications using Kubernetes expand the threat surface, whether used in private, public, or hybrid cloud. Cloud-native functions (CNFs) provide less isolation than virtual network functions (VNFs).
Impact	Attacks on a CNF can impact confidentiality, integrity, and availability.
Benefits of ZT	Cloud-native computing fits within a ZTA approach in that one of its main functions is to provide compute resource isolation from a security segmentation standpoint.
Potential Mitigations	<ul style="list-style-type: none"> > Cloud security best practices from Open Web Application Security Project (OWASP) [16], Center for Internet Security (CIS) [17] Cloud Security Alliance (CSA) [18] and US DoC National Institute of Standards and Technology NIST [19]. > Kubernetes hardening. > Guidance provided by CISA and NSA Enduring Security Framework [20]: <ul style="list-style-type: none"> > SBOM(s) for initial release/patch vulnerability scanning > Ongoing vulnerability scanning > Isolation > Role-Based Access Control (RBAC) using least privilege > Strong API and user authentication and authorization > Monitoring, logging, and auditing at the application layer > Periodic configuration validation

5.7

Use Case 7 – Edge Applications

Use Case ID	UC-7
Use Case Name	Application deployment at the Edge
Description	With 5G Edge deployments, functions such as processing and analytics are performed closer to where the data is generated, which helps enhance user experience and increase productivity. Applications are deployed and delivered with a holistic architecture that provides seamless application distribution and security at the Edge to provide services to end customers.
Risk Types	Internal and External
Risk Summary	Cloud-native workloads in private, hybrid, and public clouds expand the threat surface from traditional on-premise networks. The new approach of application deployment at the Edge combines

compute, networking, and storage aspects with security and application management. In general, application deployment at the Edge operate closer to the users to provide better overall user experience. At the same time, the Edge can have restricted security capabilities because it is designed for minimized environments without the heavy compute and security typically found in traditional mobile data centers. Furthermore, the flexibility of the standards introduces concerns related to shared hosting in which a mobile service provider might host an Edge function element (e.g., EES) on behalf of multiple enterprise Edge function elements (e.g., ECS) within the 5G network.

Impact

Application deployment at the Edge becomes distributed to more points across different geolocations, essentially creating a distributed attack vector. If no adequate protection is in place, the applications running at the Edge could be vulnerable to the distributed attack vector that can negatively impact the confidentiality, integrity, and availability of the Edge services.

Benefits of ZT

A ZTA approach will reduce risks of attacks.

Potential Mitigations

- > In the untrusted domain, PCF communication should be allowed via a security gateway. This requires further standardization.
- > In the trusted domain, in addition to the use of NEF/SCEF, PCF communication should be allowed via a security gateway. This requires further standardization.
- > In the case of hosted ECS, any communication between third parties, regardless of the location and ownership, must be performed via a security gateway in addition NEF/SCEF. This requires further standardization.

5.8 Use Case 8 – O-RAN Open Fronthaul

Use Case ID

UC-7

Use Case Name

O-RAN Open Fronthaul

Description

Open Radio Access Network (O-RAN) architecture specified by O-RAN Alliance that disaggregates the RAN to enable a multi-vendor environment and cloud-native RAN deployments.

Risk Types

Internal and External

Risk Summary

Open RAN, in general, introduces an expanded attack surface [21] [22]. The O-RAN LLS 7-2x disaggregates the O-RU and O-DU and introduces the Open Fronthaul (OFH) interface between them. External and internal threat actors could gain access to the OFH to perform confidentiality, integrity, and availability attacks on the CUS-Plane and M-Plane.

Impact

The impact of these attacks is outage, performance degradation, reconnaissance, and data theft.

Benefits of ZT

ZTA is a new paradigm for securing RAN. ZTA brings to RAN security controls that protect against internal and external threats for Open RAN cloud deployments. These new threats include Advanced Persistent Threats (APTs) that have penetrated the perimeter and established a beachhead inside the network with a long dwell time to perform reconnaissance through lateral movement.

Existing Mitigations

- > 3GPP 5GS air interface encryption to protect subscriber data from eavesdropping.
- > 3GPP 5GS user plane integrity protection to counter unauthorized subscriber data modification.
- > 3GPP 5GS Subscription Concealed Identifier (SUCI) to keep the mobile subscriber identity private.

Potential Mitigations

- > MFA for management users
- > mTLS with PKI and X.509 on management interfaces
- > Encryption of data in transit and data at rest using NIST-approved Cipher suites
- > IEEE 802.1X for port-based network access control

5.9

Use Case 9 – O-RAN SMO, RICs, and Apps

Use Case ID	UC-9
Use Case Name	O-RAN SMO, RICs, and Apps
Description	O-RAN architecture introduces Service Management and Orchestration, RAN Intelligent Controllers (RICs), and RIC applications (rApps and xApps). Each of these new functions also introduces new interfaces. xApps and rApps can be provided natively or by a third-party software supplier.
Risk Types	Internal and External
Risk Summary	Open RAN, in general, introduces an expanded attack surface [21] [22]. The O-RAN SMO, RICs, and RIC Apps can be exploited for confidentiality, integrity, and availability attacks by external and internal threat actors using both traditional attack techniques, as well as emerging AI/ML attacks.
Impact	The impact of these attacks is outage, performance degradation, reconnaissance, and data theft.
Benefits of ZT	ZTA is a new paradigm for securing RAN. ZTA brings to RAN security controls to protect against internal and external threats for Open RAN cloud deployments. These new threats include Advanced Persistent Threats (APTs) that have penetrated the perimeter and established a beachhead inside the network with a long dwell time to perform reconnaissance through lateral movement.
Potential Mitigations	<ul style="list-style-type: none">> MFA for management users> mTLS with PKI and X.509 on network interfaces> OAuth-based authorization to access resources and information> Encryption of data in transit and data at rest using NIST-approved Cipher suites> Digital signing of images and secure on-boarding> SBOM> Monitoring, logging, and alerting

5.10

Use Case 10 – Artificial Intelligence (AI)/Machine Learning (ML)

Use Case ID	UC-9
Use Case Name	AI/ML
Description	Use of AI/ML in 5G
Risk Types	Internal and External
Risk Summary	According to NIST, AI introduces new risks and increases the severity of some existing risks when compared to traditional software [23]. 5G will increasingly utilize AI/ML for performance and resource optimization in the RAN and Core. While AI/ML brings benefits, it also introduces threats due to risks such as data poisoning, compromised data transfer, untrusted data sources, and corrupted data models.
Impact	These attacks on AI/ML data and models can impact confidentiality, integrity, and availability.
Benefits of ZT	A ZTA approach will reduce risks of attacks on data and data models.
Potential Mitigations	AI/ML security is a rapidly developing topic and should continue to be monitored. Guidance provided by ETSI [24] and NIST [25] includes:

- > Mitigation against training attacks
 - > Poisoning attack mitigations
 - > Backdoor attack mitigations
- > Mitigation against inference attacks
 - > Evasion attack mitigations
 - > Model stealing attack mitigations
 - > Data extraction attack mitigations

5.11

Use Case 11 – Network Slicing

Use Case ID	UC-11
Use Case Name	Network Slicing
Description	Network slicing is the concept of multiple logical customized networks on a shared infrastructure, each complying with agreed SLAs for their function. Slicing architecture runs over different domains (device, access network, core, transport, network management systems) and multiple vendors [26] [27].
Risk Types	Internal and External
Risk Summary	The DHS CISA and NSA Enduring Security Framework “Potential Threats to 5G Network Slicing” lists 20 threat vectors rated as High (3), Medium (9), and Low (8). The High-level threats are denial of service, man-in-the-middle, and configuration attacks [28].
Impact	Attacks on network slicing can impact confidentiality, integrity, and availability. Denial-of-service attacks mainly impact availability, whereas man-in-the-middle and configuration attacks can affect all three.
Benefits of ZT	Network slicing incorporates ZTA concepts by providing slice security isolation for each segment. Proper end-to-end slice isolation reduces lateral movement between slices. Traditional perimeter defenses and system-overload protections should continue to be used to complement ZTA for defense against external attacks on availability.
Existing Mitigations	<ul style="list-style-type: none"> > Slice Access Control: 3GPP R16 Network Slice Specific Authentication and Authorization (NSSAA) should be used for secondary authentication by the UE to the network slice. > Isolation: Implementation of slice isolation requires combining network isolation components in the RAN (RAN Slicing), transport (MPLS VPN segments, IPsec), and Core (shared and dedicated network functions). > Encryption and Authentication: Implementation of Core SBA network function encryption and mutual authentication using TLS 1.3 and OAuth2 with PKI across different vendors and slices. > Monitoring and Visibility: Network slice specific monitoring for visibility.

5.12

Use Case 12 – 5G Supply Chain

Use Case ID	UC-12
Use Case Name	5G Supply Chain
Description	The 5GS consists of a broad array of hardware and software solutions from a variety of vendors. 5G is the first cellular generation that was standardized to utilize virtualization and a cloud-native approach for the 5GC and the RAN. Compute platforms must interoperate seamlessly for 5G’s network functions to perform properly and at large scale. The virtual compute stack consists of

network function virtualization infrastructure (NFVi), a cloud management platform that can manage the compute, network, and storage resources, an operating system, and the application itself. This architecture can be used either on-premises or via a public hyperscaler through a hybrid and/or public model. All of these hardware and software components can be sourced from different vendors, and many of them will include free and/or open source software (FOSS).

Risk Types

External and Internal

Risk Summary

The U.S. government has identified telecommunication networks as critical infrastructure for national security. Software developers, software vendors, hardware vendors, and operators that do not align with a mature DevSecOps model could allow a threat actor to maliciously implant code into these products. That nefarious code could include a zero day attack, send compromised data to a command-and-control server, give the threat actor remote access to the network, etc., all of which could be very difficult to detect. In addition, without a mature DevSecOps program, packages with known vulnerabilities may remain in the supply chain. The ICT supply chain consists of retailers, distributors, and suppliers that participate in the sale, delivery, and production of hardware and software [29]. Common attack techniques include hijacking updates, undermining code signing, compromising open source code [29], and attacking known vulnerabilities in code bases.

Impact

A threat actor that successfully compromises the supply chain could disrupt the availability of telecommunication services, potentially impacting millions of customers. Threat actors could compromise the confidentiality and integrity of the messages that are being sent on any one of the network traffic planes, capturing information about network user communication.

Benefits of ZT

Some ZTA architecture techniques, like DevSecOps, micro-segmentation at the software and virtual compute workload level, and network segmentation, can mitigate the risks of supply chain attacks and limit the potential blast radius associated with a successful attack. In addition, all entities in the 5G supply chain should follow the shared responsibility model, outlined in 7.3, when deployed in third-party clouds.

Potential Mitigations

- > Require all vendors to undergo a cyber risk assessment.
- > Require all vendors to identify and fix known vulnerabilities before delivering software.
- > Require all vendors to rapidly deliver fixes for newly discovered zero day vulnerabilities.
- > Limit and/or restrict acquisition of hardware and/or software from any untrusted suppliers.
- > Implementation of micro-segmentation of the software as a part of a zero-trust architecture.
- > Implementation of network segmentation.
- > DevSecOps.
- > Adherence to the NIST Secure Software Development Framework (SSDF) [30].
- > Use of digital signatures for the signing of software binaries from the vendors.
- > Use of a consumable Software Bill of Material (SBOM) for automated vulnerability management scanning.
- > For any use of FOSS in the software supply chain, it is recommended that:
 - > Industry collaborate to foster improved communications and alerting relating to vulnerable software code.
 - > OpenSSF's Security Score Cards are used when selecting open source software [31].



6. HOW ZERO TRUST APPLIES TO 5G TODAY

6.1 Zero Trust in the 5GC

Wireless connectivity has become ubiquitous, and with the huge number of connected devices, the attack surface is broad. Wireless communication networks are made up of multiple interconnected networks, which may be trusted and/or untrusted. Any trusted network that interfaces or communicates with an untrusted network becomes in itself untrusted. Therefore, 5G networks should be treated as untrusted.

ZT ignores implicit trust throughout the 5G network by consistently putting the “never trust, always verify and re-verify” mindset into practice. Implementation requires unified security analytics, enforcement, and visibility across the 5G network. All 5G network assets, from devices to applications, that interact with other 5G assets are considered untrusted in a ZTA and must be authenticated and authorized for access.

Specifically, for its 5G ZT study, 3GPP has been focusing only on the Service Based Interface (SBI), but not other aspects of the 5G network, such as the RAN. The SBI is central to the Service Based Architecture (SBA) that powers the 5GC control plane. Figure 6.1-1 below shows the 3GPP perspective of how ZT applies to the 5GC, in which communication between 5GC network functions is secure regardless of network location and implementation form (PNF, VNF, CNF).

SBI specifies that secure API-based communication can take place between two or more network functions within the 5G SBA. 3GPP recommends that 5G networks use mutual Transport Layer Security (mTLS), PKI, and OAuth to secure communications between 5GC NFs. An NF can utilize an API call over the SBI to invoke a particular service or service operation in a secure fashion.

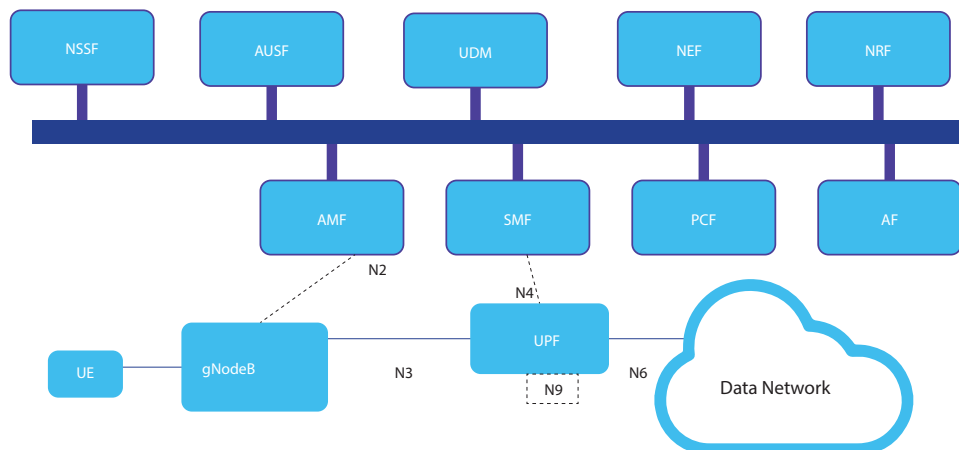


Figure 6.1-1. Service Based Interfaces (SBI) Communication in the 5G Core

6.2 5GS Security Domains

The 3GPP 5G System security architecture is specified to have the following security domains:

- I. **Network Access Security:** Network access security provides a set of security features that permit UE to authenticate and access services (both non-3GPP and 3GPP accesses), and to safeguard against attacks on radio interfaces. Network access security also entails the communication of security context from 5G NR to the UE.
- II. **Network Domain Security:** Network domain security includes security features that allow user plane and signaling data to be securely exchanged between network nodes.
- III. **User Domain Security:** User domain security focuses on security tools that protect a user’s access to mobile equipment.
- IV. **Application Domain Security:** Application domain security provides a set of security features that enable applications in the user domain and in the provider domain to securely exchange messages.
- V. **SBA (Service-Based Architecture) Domain Security:** SBA domain security specifies the method for private communication between NFs within the serving network domain and with other network domains via the HTTP/2-based SBI. The SBA defines flat, peer-to-peer interactions between NFs.
- VI. **Visibility and configurability of security:** The visibility and configurability security domain provide a group of functions that enable a user to determine whether a security feature is active.

Figure 6.2-1 illustrates some of these security domains. The six security domains are discussed in more detail in the subsections below.

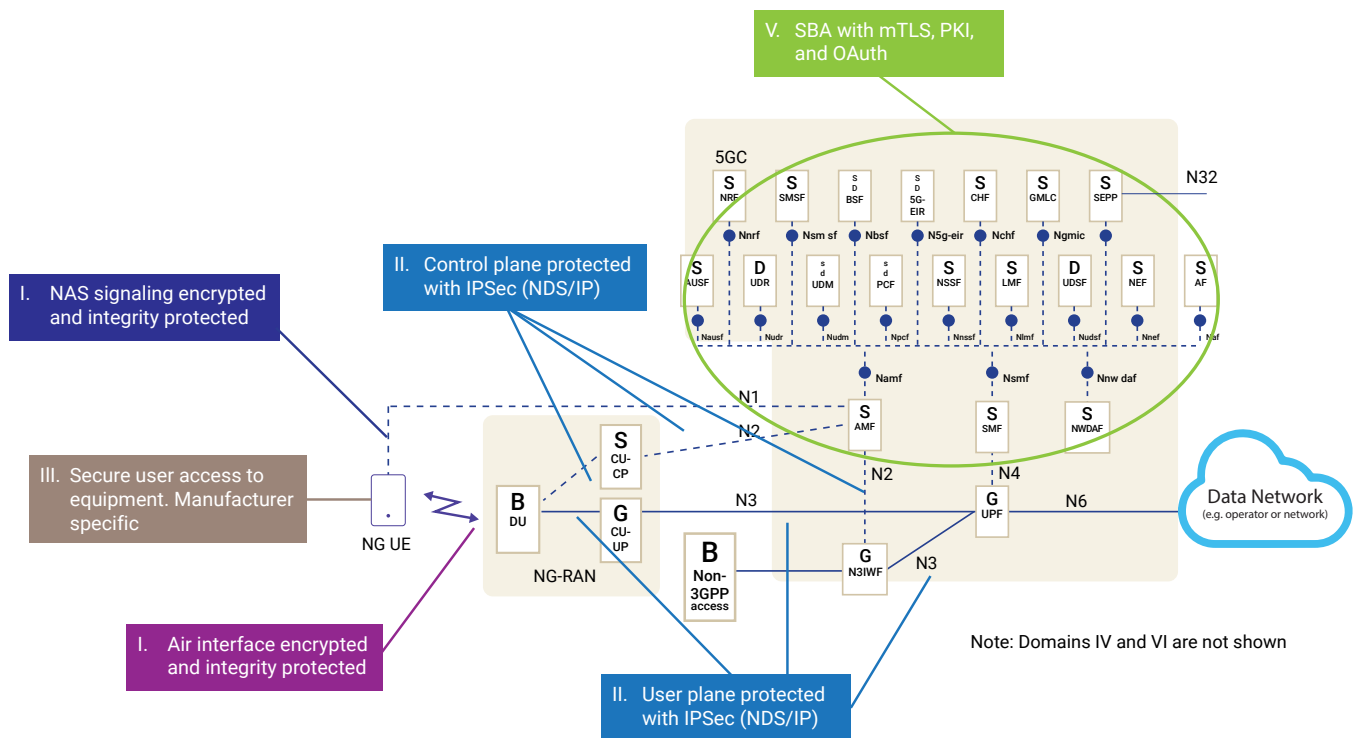


Figure 6.2-1. 5G Security Domains and Secure Communications

6.2.1 5G Network Access Security (I)

Network access security is concerned with ensuring that a 5G UE can authenticate and securely access services via a 5G network. A critical aspect of the 5GS security architecture is the various security identifiers, keys, cryptographic algorithms, and protocols used to provide security and privacy features. Some of these are touched upon below in general terms.

5G Identifiers

A 5G UE typically has a USIM/eSIM, which contains a Subscription Permanent Identifier (SUPI) that identifies the 5G UE subscription information and home network of a 5G subscriber. To protect the 5G subscriber's privacy, the SUPI is never transmitted in the clear but instead concealed in a Subscription Concealed Identifier (SUCI), which the home 5G network de-conceals to reveal the SUPI and hence identify the 5G subscriber subscription information. After a 5G UE has successfully registered with a 5G network, the 5G UE is assigned a temporary identifier, known as a 5G Globally Unique Temporary Identifier (5G GUTI), which is updated frequently by the 5G network to protect the 5G subscriber privacy.

Additionally, the USIM/eSIM and the 5G network both hold the same secret key (K), which is never exposed externally, and the same set of cryptographic algorithms, which are used to mutually authenticate the 5G UE and 5G network.

5G Primary Authentication and Authorization

To obtain services, a 5G UE must first register with a selected 5G network using a number of steps:

- > The 5G UE selects a 5G network and sends a request, containing its SUCI, to register. The 5G network uses the SUCI to reveal the SUPI which identifies the 5G UE subscription information which includes the secret key (K).
- > The 5G network triggers an authentication procedure by running a challenge-response protocol, which when successful results in proving that both parties have knowledge of the same secret key (K) and hence mutually authenticate each other.
- > Secure communication channels are established between the 5G UE and 5G network using security keys derived from K and agreed security algorithms. Specifically, signaling between the 5G UE and 5GC and signaling between the 5G UE and 5G RAN is protected from confidentiality, integrity, and replay attacks.
- > Finally, based on the 5G UE subscription data and 5G network policies, the 5G network will authorize the services the 5G UE is allowed to access, such as allowed data and/or voice services, network slice(s), geographical areas, roaming services, and so on.

The 5GS supports two authentication methods, namely 5G Authentication & Key Agreement (AKA) and the Extensible Authentication Protocol (EAP) AKA authentication methods. In non-public networks, other EAP methods and associated credentials may be supported, such as EAP TLS.

The establishment or modification of 5G data connections between a 5G UE and 5G network to carry user data is subject to authorization by the 5G network based on subscription and policy information.

5G Secondary Authentication and Authorization

In addition to primary authentication and authorization, the 5GS also supports secondary authentication and authorization of 5G UE data connections to external data networks. An example is those of an enterprise, enabling the enterprise to enforce its own security policies on users of its network. Similarly, the 5GS also supports network-slice-specific authentication and authorization.

5GS Policy Framework

The 5GS incorporates a policy framework, which generates policies based on a combination of inputs that include locally configured policy rules, static information (such as 5G UE subscription information), and dynamic information (such as 5G UE geographic location, time of day, and various event triggers.) These policies are distributed to enforcement end points in the 5GS such as the 5G UE, 5G RAN, and 5GC NFs.

6.2.2 5G Network Domain (II)

The 5G network security domain is concerned with ensuring that 5G network nodes, such as those in the 5G RAN and/or 5GC, can exchange signaling and user data securely.

The authentication and security protection (integrity, confidentiality, and replay-attack protection) of IP-based signaling interfaces in the 5GS, such as those between the 5G RAN and 5GC domains, is realized using IPsec/IKE technologies.

To support call routing and/or roaming, 5G networks interconnect via Security Edge Protection Proxies (SEPPs) to protect the inter-network service-based signaling. The interface between SEPPs is mutually authenticated and protected by TLS and in certain cases by PRINS (Protocol for N32 Interconnect Security). Inter-network user data protection, which does not traverse the SEPP, is provided using Inter PLMN User Plane Security (IPUPS).

6.2.3 User Domain Security (III)

User domain security refers to the passwords, fingerprints, facial recognition and other security features that enable secure user access to mobile equipment. These features are specific to mobile equipment manufacturers and are not standardized. Security features to ensure that only an authorized USIM may access mobile equipment are standardized by 3GPP, as are features that ensure only authorized users may access a USIM, such as with a PIN.

6.2.4 Application Domain Security (IV)

3GPP application domain security refers to the set of security features that enable end-to-end security of applications in the user and application domain, including applications running on the USIM, in the ME, or in the provider domain. Examples of these applications include the 3GPP-standardized USIM application [32], GSMA Remote SIM Provisioning (RSP), and GSMA Secure Application for Mobile (SAM). End-user applications defined outside the scope of 3GPP (e.g., Facebook and Twitter) and that run on the mobile equipment OS as opposed to on the USIM employ their own approach

to realize end-to-end security. Many of these applications that run over the top (OTT) use IETF-defined and/or other dedicated mechanisms for end-to-end (E2E) security. UE applications, and the application security domain, are considered out of scope for MNO ZTA initiatives, but in scope for the application owners.

6.2.5 Service-Based Architecture (SBA) (V)

In the 5GC, SBIs between NFs are authenticated using mutually authenticated TLS, and transport protection (integrity, confidentiality, and replay-attack protection) is provided by TLS. Protection may also be provided by other means (e.g., physical protection of the interfaces). Authorization is also supported using the OAuth 2.0 Framework. PKI provides the enablers to support the above technologies.

6.2.6 Visibility and Configurability of Security (VI)

Security visibility refers to the capability to provide an indication to the end user or application whether security mechanisms, such as confidentiality and/or integrity protection, may be enabled on the radio access, allowing the decision whether to proceed with a call. 3GPP has standardized some features in this space.

Security configurability enables the user to control certain security features, for example to configure whether User-USIM authentication is enabled or disabled.

6.3

Alignment of 5G to NIST ZT Tenets

As discussed in a previous section, NIST has defined seven tenets of ZT [4]. This section described the alignment of 5G security standards to those seven tenets.

Tenet 1 requires the system owner to identify resources within the 5GS so that the system owner constructs the ZTA around these resources. In 5G, network assets—including UEs, RAN, transport, Core, applications, and services—are considered assets and data sources that have different system owners. Compute resources running NFs and applications are also considered assets.

Tenet 2 requires that all network communications provide confidentiality, integrity, and source authentication. In 5G, encryption and integrity protection, authentication schemes, and security protocols are leveraged to secure (or authenticate) communication amongst participating entities in the network.

Tenet 3 requires that any service request includes authenticating the requestor prior to granting access to a resource and limiting that access to only what is required to complete the task. In 5G, token-based authorization using OAuth 2.0 is an optional security feature for an NF service consumer to access services provided by an NF producer.

Tenet 4 requires that a requestor is evaluated to be in a secure state prior to being allowed access to a resource. In 5G, the Core evaluates what resources connect to it by

authenticating user devices, applying policies, and managing the mobility of devices before allowing access to mobile operator services or the internet.

Tenet 5 requires all resources and owned associated assets are continually monitored and evaluated to ensure they are in a secure state. In 5G, the monitoring and evaluation of resources and owned assets are typically performed by non-5G entities. Examples include processing security-related logs by a SIEM, systems that process IPFIX and full packet capture data, and endpoint agents such as EDR. These monitoring entities may be outside the scope of the 3GPP-specified 5GS.

Tenet 6 requires that on-going communication sessions between a requestor and a resource be evaluated to ensure both parties are in a secure state. In 5G, NF service consumers can be periodically evaluated prior to consuming services via OAuth 2.0 authorization.

Tenet 7 requires ZT policies and relevant data be periodically reevaluated for their security effectiveness based on past performance and updated as needed. In 5G, there is no inherent network function or service that takes into consideration the collection of dynamic data and current states of assets to enhance the network's security effectiveness.



7. ZTA FOR CLOUD/VIRTUALIZATION/CONTAINERIZATION

7.1 Cloud Infrastructure

Cloud Infrastructure consists of underlying computing, network, and storage components. NFs running on cloud infrastructure can either be virtualized through an NFV framework, with VNFs having a guest OS or packaged under the VM through a hypervisor. NFs may also be deployed in the form of a cloud-native architecture, comprised of containers supported by container orchestration platforms like Kubernetes for better agility, scalability, and flexibility.

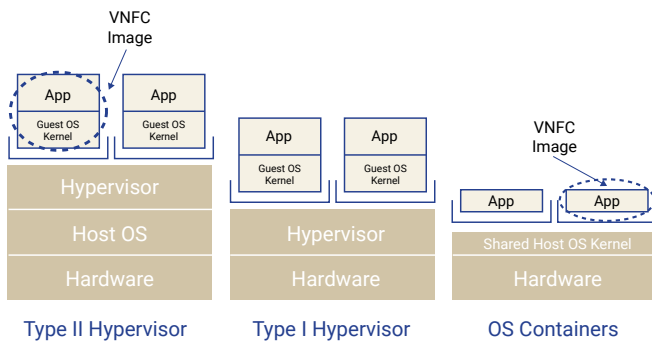


Figure 7.1-1. Comparison between CNF and VNF Virtualized Stack [33]

In the virtualization environment, a VNF is instantiated in a VM, which abstracts the physical hardware, managed by a virtualized manager, which supports orchestrating, controlling, and scheduling the VNFs, as well as the virtual infrastructure. CNFs may also be comprised of control/master nodes, worker nodes, edge nodes, and storage nodes. A Control Manager can also be optionally used for bare metal deployments. Other deployment models include a virtualized deployment on open stack.

7.2 Virtual Machines, Containers, and Kubernetes

ZT security to 5G virtualization and containerization means:

- > Identifying the network infrastructure and applications.
- > Assessing the architecture from threats.
- > The type of access it allows and under what condition,
- > Evaluating the design for direct and compensating security controls.
- > Mapping the attack vectors from the lens of internal and external adversaries.
- > Classifying the vulnerabilities into impact and probability of occurrence.

- > To define security monitoring across the stack and ensure that policy enforcements are synchronized across the different layers.

Figure 7.2-1 shows the lifecycle of the ZT security framework that applies to the 5G virtualized cloud infrastructure.



Figure 7.2-1. ZT Framework for 5G Cloud Infrastructure

Virtual Machines and Network Functions

The approach to mitigate infrastructure threats in NFV is to first classify the stack into risk domains, then identify and classify threats across each domain, and finally to prioritize direct and compensating controls.

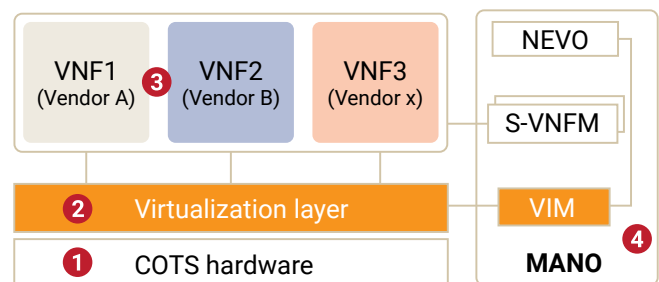


Figure 7.2-2. 5G VNF Attack Domains

Domain	Risk Category	Risk Impact and Mitigation
0	Attack Vector	<ul style="list-style-type: none"> > Confidentiality attacks on VNF: user privacy information (location, CDR, IP address), session data, 5G auth vectors, SIM keys, security context, guest OS > Root kit for infra compute > Hypervisor controller attacks like hyper jacking > DDOS on public APIs for MANO
1	Assessing Infrastructure	<ul style="list-style-type: none"> > Residual data breach due to shared storage > No clear isolation matrix, large attack area > No clear boundary between logical networks > Detection of threats and response > Security logging
2	Application Security	<ul style="list-style-type: none"> > Open source software introduces more vulnerabilities > Software composition analysis > Difficult to monitor the internal traffic of the virtual network
3	Operational Security	<ul style="list-style-type: none"> > Increase the risk of security issue identification > Trustworthiness of VNFs > Data security with trusted processing > Multi-vendor account, permission, and authentication management > Security monitoring with loops > Dedicated PKI
4	Security Policy Induction and Orchestration	<ul style="list-style-type: none"> > Induction of policy points from Active Directory, hypervisor, and Linux kernel > Security policies need to be automatically adjusted during service migration through MANO

Table 7.2-1: Risk Categories and Mitigations for 5G VNF Deployments

Now that we have explored the threat analysis and different risk domains on the VNF stack, it is important to classify this risk in terms of exposure and impact. The mitigation strategy should tie back to the ZT security framework, such that each layer should have its own policy checkpoint. On the other hand, container, network, and management functions present different sets of challenges and need to be examined, as well, in the overall effort of applying a ZT security framework to 5G.

Containers and Kubernetes

The 5G cloud-native architecture is generally comprised of control/master nodes, worker nodes, edge nodes and storage nodes. A control manager can optionally be used for bare metal deployments, and other deployments include virtualization on open stack.

The master node is designed to be bound to the public O&M network, and it is an add-on to a generic Kubernetes master. Keepalive messages are used to support a virtual IP address across the master nodes. A worker node is equivalent to a generic Kubernetes node, which is designed to run applications. An edge node is designed to interface with an external network, and it provides a proxy for data traffic in and out of the Kubernetes cluster. Figure 7.2-3 depicts a typical network design with internal and external accessibility points.

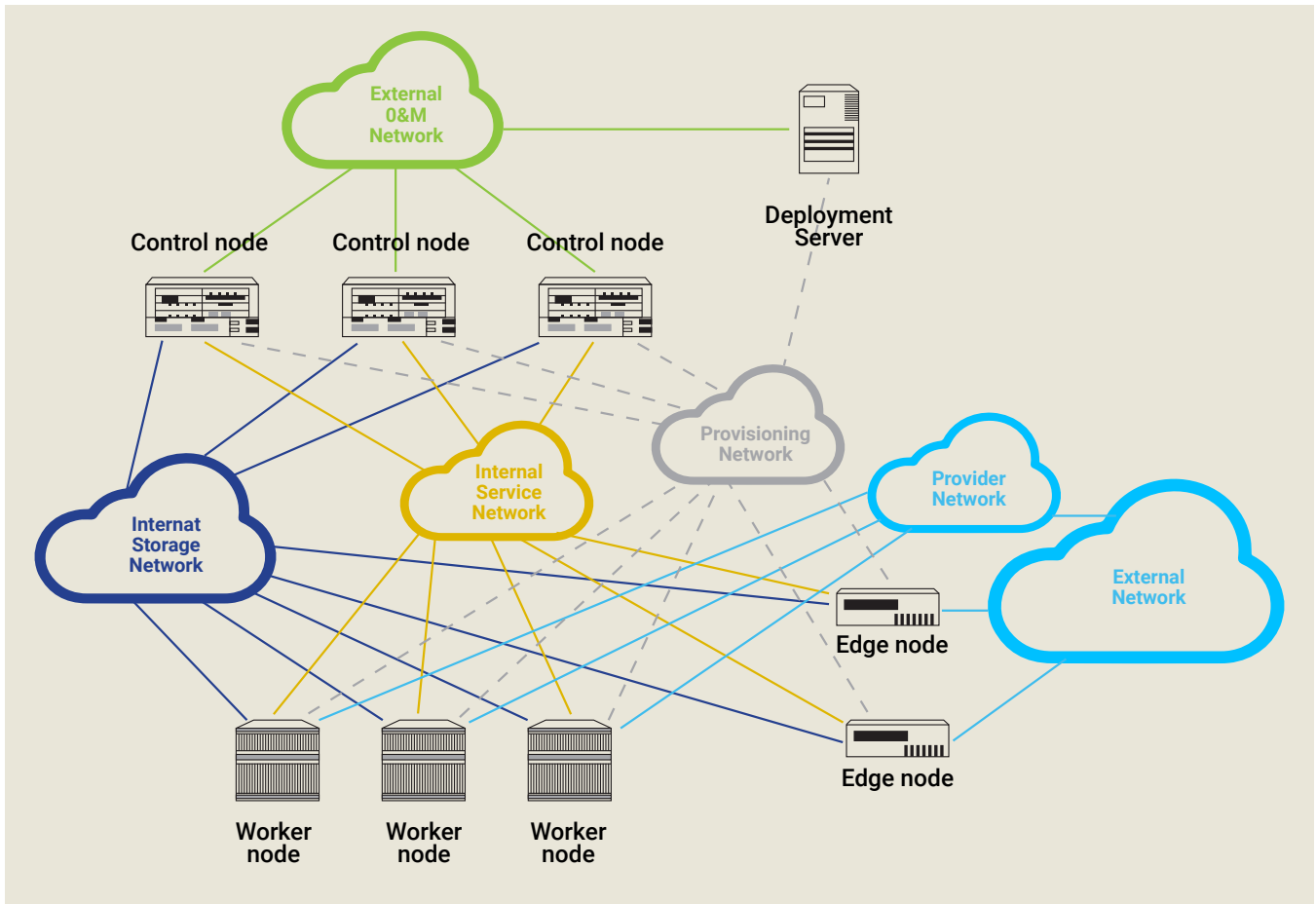


Figure 7.2-3. Cloud-Native Deployment Model

If we examine the attack's vectors in the Kubernetes stack, master nodes are the primary focus of the attackers. That's because it is exposed to the enterprise through an API server, but worker nodes are workload runs, so both layers need to be looked upon carefully. Figure 7.2-4 shows how the attack vectors fall into different layers.

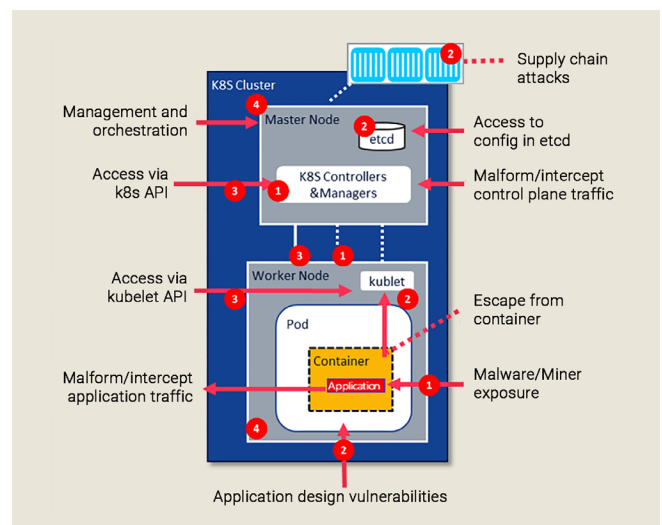


Figure 7.2-4. 5G CNF Attack Domains

Domain	Risk Category	Risk Impact and Mitigation
0	Attack Vector	<ul style="list-style-type: none"> > Intercepting control plane traffic > Accessing the ETCD directory to steal the service account tokens > API server been accessed by authentication/un-authorized calls > Worker nodes accessed by Kubelet API; these are sensitive pods, should be non-routable
1	Assessing Container Stack	<ul style="list-style-type: none"> > Identity and Access Management into 5G CNF's Infra Master node from external network is done via privilege access management (PAM) tool as primary and jump host as a break-glass scenario with no direct access through worker and edge nodes > Enterprise AD to enforce the policy for SSO, LDAP server to authorize key cloak users towards Kubernetes Master > Strict Communication Matrix for trusted networks, which has control node IPs mapping with relevant source IP (administrator and provisioning) and ports > IP Tables with the use of firewall rules for inter/intra-CNF traffic, separate IP tables are needed from control node, edge, and worker nodes
2	Application Security	<ul style="list-style-type: none"> > Open source software introduces more vulnerabilities, to be validated for CVEs > Software Composition Analysis for Registries and OS binaries through authenticated scanning and signatures > Malform/intercept application traffic can be prevented with service account authorization through MTLS, with namespace isolation > ETCD to support encryption at rest with no human/external API access > Static and dynamic application security testing to identify vulnerabilities in code written by an NF provider
3	Operational Security	<ul style="list-style-type: none"> > Isolate pods, hosts, and namespaces through policies defined in security context > API request through cluster to be validated by admission controller > Traffic filtering and zoning for master and worker nodes > Runtime security detection with EDR; logs to be sent to SIEM
4	Security Policy Induction and Orchestration	<ul style="list-style-type: none"> > Induction of policy points from Active Directory for human users, and security context of cluster, SELINUX profiles in host > Security policies to be set for each namespace with labels with action set, deployment server to be used as an orchestration layer

Table 7.2-2. Risk Categories and Mitigations for Kubernetes Stack

Further guidance for cloud security best practices is provided by Open Web Application Security Project (OWASP) [16], Center for Internet Security (CIS) [17], Cloud Security Alliance (CSA) [18], and U.S. DoC NIST [19].

Figure 7.2-5 shows a reference ZT architecture for virtual and containerized NFs. PDP is the control plane, which hosts policy engines such as identity and access management, PKI, governance, risk control checks, and SIEM platform. The policy gets the feed from the threat analysis and mitigation exercise, whereas the data plane consists of the policy enforcement points across different layers. The security enforcements include authentication/authorization checks from subjects, as well as encryption in transit and at rest using certificates and keys. It also contains malware-detection agents in the OS, security, and audit logs streams.

latency thresholds can be achieved.

The 5GC is comprised of an SBA, with control plane and user plane functionality, and with the northbound interface interworking with the OSS and BSS systems for provisioning and billing functions. The 5GC and its network management system can be hosted in a public cloud, whereas the BSS interworking part can be deployed in an on-prem environment.

With infrastructure deployment in the operator data center, the assets are managed locally. Security components like hardware security module, vaults, Active Directory, LDAP servers, and endpoint detection tools are all hosted locally, so that access control and security posture can be controlled. Data privacy is also an important reason why operators host the network architecture locally. Databases

containing subscriber profiles and CDRs are other datasets that are processed and stored in the Core are recommended to be hosted locally because security controls are data privacy issues that outweigh the benefit of hosting in a public cloud.

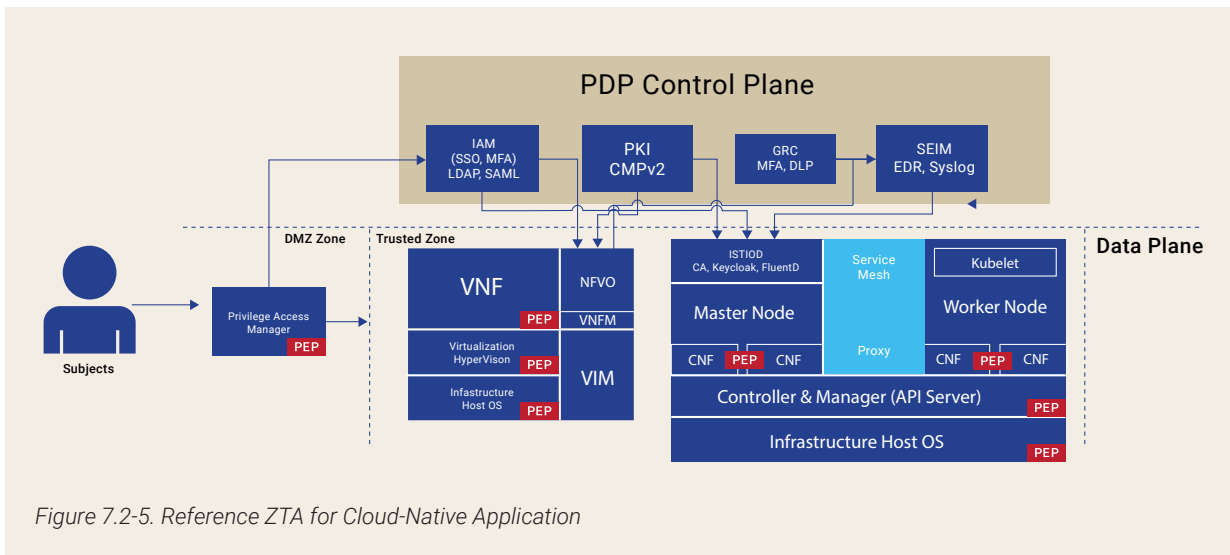


Figure 7.2-5. Reference ZTA for Cloud-Native Application

7.3 Deployment Models (e.g., Public/Private Clouds)

VM and containerized deployments use public, private, and hybrid cloud methods. The models depend on technical, business, data privacy, and legal considerations. In public cloud deployments, the cloud provider supplies the infrastructure (e.g., AWS, Microsoft, Google, etc.), whereas hybrid cloud deployments use a mixture of private on-premise equipment and some portion of the public cloud.

To illustrate this further, using RAN as an example, we see it is comprised of RUs, CUs, and DUs. The RU will be on the cell site, while the DU carries out real-time functions like scheduling, which means it should be near the cell site. The CU (both CU-CP and CU-UP) performs non-real-time functions, so it can be located far from the cell site, except when an URLLC service is required. Then the CU-UP should be near the cell site, as well.

If public cloud service providers like AWS, Microsoft Azure, or Google have enough data centers distributed across the 5G network coverage, it will be possible to move some of the RAN functions to the public cloud, assuming the required

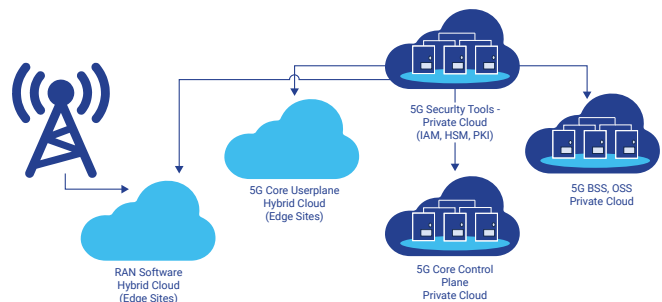


Figure 7.3-1. Example Cloud Deployment Model for a 5G Network

The importance of security considerations for mobile networks deployed on cloud infrastructure is expected to grow as 5G use cases are realized. A key concern when moving NFs to the cloud is the introduction of new risks and vulnerabilities that can be exploited by a malicious external or internal threat actor. MNOs will need to mitigate these risks in the cloud even without having full control of infrastructure. Delegation of security responsibilities to achieve a ZTA can be complicated for 5G critical infrastructure deployed in public and hybrid clouds. While the Cloud Shared Responsibility

Model provides guidance for security responsibility, a lack of a standard security framework for cloud deployments of 5G critical infrastructure puts further responsibilities on the MNO and their software suppliers, as shown in Figure 7.3-2.

























	Infrastructure-as-a-service (IaaS)	Platform-as-a-service (PaaS)
 Human access	 MNO/Software Supplier	 MNO/Software Supplier
 Data	 MNO/Software Supplier	 MNO Software Supplier
 Application	 Software Supplier	 Software Supplier
 Operating system	 MNO	 Cloud Service Provider
 Virtual networks	 MNO	 Cloud Service Provider
 Hypervisors	 Cloud Service Provider	 Cloud Service Provider
 Servers and storage	 Cloud Service Provider	 Cloud Service Provider
 Physical networks	 Cloud Service Provider	 Cloud Service Provider

Figure 7.3-2. Cloud Shared Responsibility Model



8. IMPLEMENTATION AND OPERATIONAL CONSIDERATIONS

8.1 Edge Deployments

3GPP Release 17 introduced a new edge compute architecture that defines new network domains, network functions, interfaces, flows, etc. 3GPP incorporated additional aspects of ZT principles in this new architecture, including the PDP and PEP functionalities, along with authentication, authorization, and the application of policies. The 3GPP architecture supports various deployment options, which allows flexibility in implementing ZT principles. Edge deployments can vary based on factors such as network topology, security needs, and other specific requirements. This adaptability enables alignment with principles of ZT and facilitates the implementation of ZT with the new edge compute architecture. The bulk of the technical specifications are captured in TS 23.548 5G System Enhancements for Edge Computing [34], TS 23.558 Architecture for enabling Edge Applications [35], TS 28.538 Management and orchestration for Edge Computing Management [36], and TS 33.558 Security Aspects of Enhancement of Support for Enabling Edge Applications [37]. This section will touch upon the high points from a ZT perspective.

In reviewing these relevant technical specifications for edge computing for this ZT study, it can be stated that the 3GPP Edge Computing's Edge Configuration Server (ECS) is functioning as a PDP and the EES is functioning as a PEP. Figure 8.1-1 illustrates the high-level edge compute architecture.

For the UE to function and operate in this new edge computing architecture, the UE must consist of an Edge Enabler Client (EEC), Application Client (AC), and Notification Management Client. The UE must be authenticated and authorized by the ECS via the TLS handshake or the General Public Subscription Identifier (GPSI) before the UE can access this architecture. After that, the UE can be registered and validated by the EES.

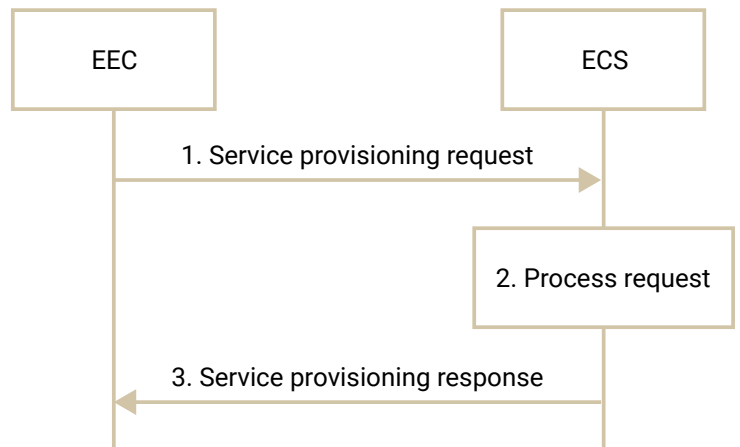


Figure 8.1-2. Edge Enabler Client (EEC) Authentication/Authorization with the Edge Configuration Server (ECS)

The EEC sends a service provisioning request to the ECS, and the ECS authenticates the EEC via the TLS handshake or GPSI. The ECS performs an authorization check to verify whether the EEC has authorization to perform the operation. The ECS will identify the EES based upon the AC profile and UE location. The ECS will apply the appropriate policy, including identification of the Edge Data Network (EDN), EDN service area, and EES endpoints.

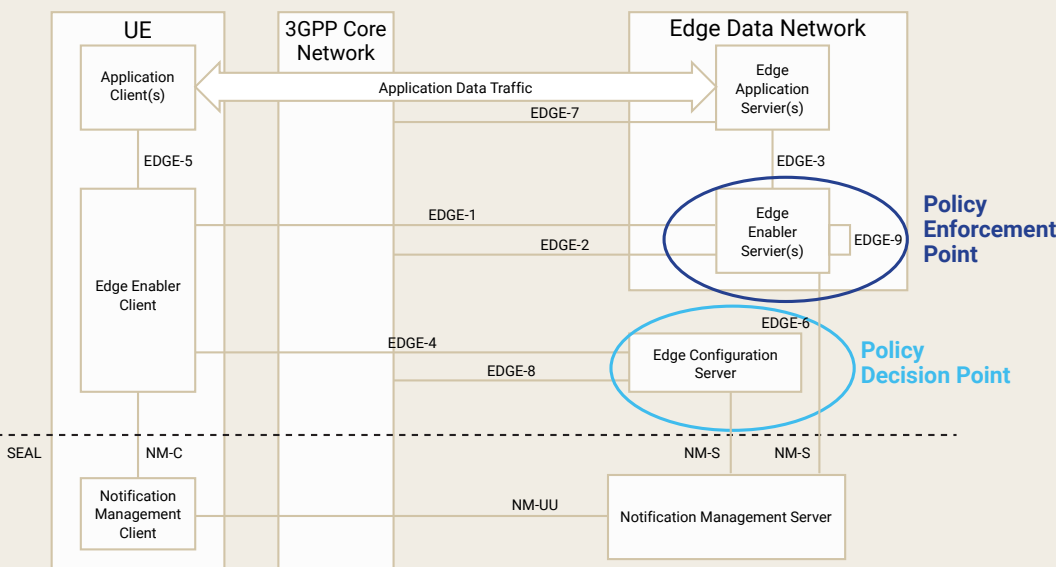


Figure 8.1-1. Architecture for Enabling Edge Applications

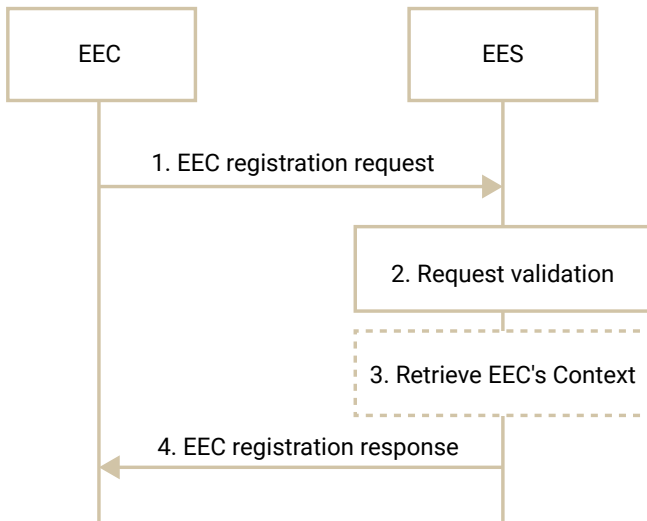


Figure 8.1-3. Edge Enabler Client (EEC) registration to the Edge Enabler Server (EES)

After the EEC is authenticated and authorized by the ECS, the EEC will register with the appropriate EES. The EES authenticates the EEC via the TLS handshake or GPSI. The EES is to validate the registration request and verify the security credentials. If the registration is successful, the EES sends the EEC a registration ID, which allows the AC to discover and communicate with the appropriate EAS.

The EES and Edge Application Servers (EAS) can be deployed and operated within a third-party location (e.g., Application Service Provider's (ASP) network or within an Edge Computing Service Provider's (ECSP) network). Within the 3GPP specifications, they have identified a concept of a "trusted" and "untrusted" edge computing network domain. For trusted domains, the EES and EAS servers can directly access the PCF that is within the MNO's core network. For untrusted domains, the EES and EAS server can only indirectly access the PCF by way of the Network Exposure Function (NEF) or Service Capability Exposure Function (SCEF).

From a ZT perspective, 3GPP has defined numerous security requirements already including details relating to the credentials. Within the architecture documents, the client identity that is defined is the GPSI which could be the user's mobile phone number (aka MSISDN). Within the security specifications [37], 3GPP SA3 states that authentication between the EEC and ECS shall be done during the execution of the TLS handshake protocol. For authorization, SA3 states that the client credentials must use JSON Web Tokens (JWT).

The mobile service provider might host an EES on behalf of an enterprise ECS within the 5G network. The EES can either be the frontend PEP or an intermediate PEP, for instance on a per-slice basis. From a practical implementation perspective, the EES can be under either enterprise control exclusively or under joint control of both the mobile service provider and the enterprise. EES can be shared between multiple ECSs in various access authorization domains. In general, there can be many relationships between a domain's ECS and EASs. For instance, a mobile service provider can share an ECS in concert with a given enterprise domain, or an ECS can reside and be under the control of the enterprise domain and outside the jurisdiction of the mobile service provider domain. Using static identifiers (such as an individual's phone number) is not recommended because it could assist a threat actor in taking over someone's account.

8.2 Closed-Loop Automation

The complexity, scale, and performance requirements of 5G networks require fully automated management, including installation, commissioning, configuration, ongoing operations, software upgrades, and decommissioning. Full automation will enable a network to sense its environment and adapt to changes with little to no human intervention. Automation technologies will include scripting, continuous integration and continuous delivery (CI/CD) practices, policy-based intent driven networks, and AI/ML driving real-time analysis and decision making. For communication service providers (CSPs), automated management provides better

network security by eliminating direct human management of the network, but automated management also introduces new security risks to the network in the use of automation software, AI/ML, and network data.

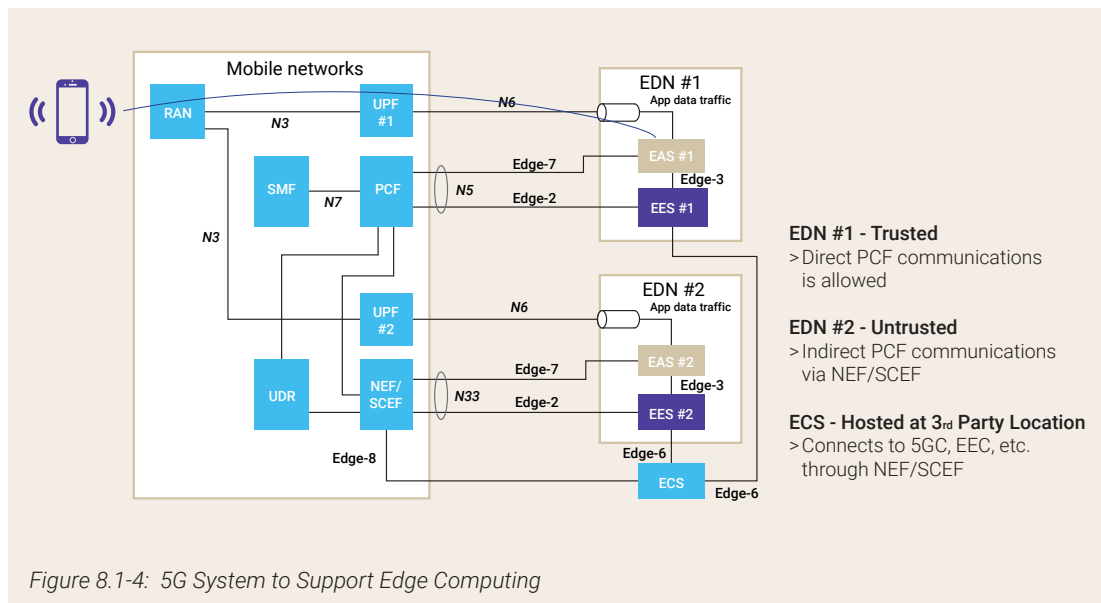


Figure 8.1-4: 5G System to Support Edge Computing

Humans cannot keep up with rapidly evolving operational requirements, such as dynamic slice allocation and deallocation or the vast number of network events that may require action. Closed-loop automation, also called zero-touch automation, is the move from manual to automated operations with AI complementing and enhancing human knowledge [38]. Zero-touch network and service management (ZSM) is ETSI's formalization of closed-loop automation for 5G networks and beyond [39]. Closed-loop automation, and specifically ZSM, have been driven by the transformation of networks into programmable, software-driven, service-based network architectures. These networks require a radical change in network and network services management and orchestration [40].

The ultimate automation target is to enable largely autonomous networks, which will be driven by high-level policies and rules; these networks will be capable of self-configuration, self-monitoring, self-healing, and self-optimization without further human intervention. All this requires a new horizontal and vertical end-to-end architecture framework designed for closed-loop automation and optimized for data-driven ML/AI algorithms [41].

ZSM provides a framework that incorporates practices from CI/CD, intent-based networking, closed-loop automation, and automated tracing. This framework will ensure more resiliency and improved security. The adoption of CI/CD practices will enable faster elimination of software bugs. Automated tracing will use AI/ML techniques to accelerate troubleshooting, root cause analysis, optimization, and detection of cyberattacks. Intent-driven networks will allow CSPs to define the behavior they expect from their network through policies, with the system translating it into real-time network action [38].

ZSM consists of three high level processes: Service on-boarding, service fulfillment, and service assurance. The service on-boarding process adds a new service model to the E2E service management domains service catalogue. The service fulfillment process enables a service instance based on the on-boarded service model, configures the service instance, activates it, and terminates it when it is no longer needed. The service assurance process keep the service free of faults and optimizes service quality [41].

It should be noted that other organizations are also doing work directly related to ZSM, including the 3GPP (SA2 and SA5), TM Forum, GSMA, MEF, ONF, ITU-T (SG13), IETF, IRTF, Broadband Forum (BBF), and ETSI ISG Securing Artificial Intelligence (SAI). Additionally, there are several open source communities implementing ZSM-related platforms such as Open Source MANO (OSM), Open Platform for NFV (OPNFV), OpenStack, the Open Network Automation Platform (ONAP), and OpenSlice.

The ZSM framework has been guided by the architectural principles of modularity, extensibility, scalability, open model-drive interfaces, closed-loop management automation, stateless management functions, resiliency, separation of concerns, service composability, intent-based interfaces, functional abstraction, simplicity, and automation [42]. The architectural principles are complemented by nine categories

of scenarios to be supported by ZSM: E2E network and service management, network as a service, analytics and ML, collaborative/federated service management, security, testing, and tracing [43]. E2E network and service management includes the automation of E2E lifecycle management of network resources and services, including installation, commissioning, configuration, day-2 operations, software upgrades, and decommissioning. The network as a service (NaaS) scenarios describe the service exposure capabilities for managed resources in the ZSM framework. Analytics and ML scenarios outline the analytics and AI/ML capabilities needed by the ZSM framework. Collaborative/federated service management scenarios provide input for supporting ZSM in a multi-provider service. The security scenarios outline the controls needed for temporary decryption during automated troubleshooting. The testing scenarios outline the support required for automated testing in production environments. Automated tracing capabilities are crucial for root cause analysis, infrastructure/service optimization, predicting future behaviors, preventing, and mitigating cyberattacks, delivering data to AI/ML capabilities, explaining AI/ML decisions, and network testing. Finally, integration and operation scenarios identify the ZSM's interaction points with external services, such as BSS/OSS and compute platforms.

ZSM will provide ZTA with additional controls that include verifying human and non-human users; restricting user access via least privilege; restricting data access to authenticated and authorized entities; managing identity lifecycle; and monitoring the network constantly. Together these controls ensure network availability by minimizing the risk of unauthorized changes and ensure the confidentiality and integrity of network, management, and personal data. The monitoring capabilities will rely on AI/ML to support automated attack/incident detection, prevention, and mitigation, thereby enabling rapid and appropriate response to anomalous behaviors. AI/ML will also enable the network to learn from past behaviors, accelerating attack detection and response. The use of tested policies and management scripts substantially reduces the risk that unexpected software is installed in the network or that vulnerable configuration changes occur. The accidental or malicious introduction of vulnerabilities is also substantially reduced by eliminating direct human access to the management plane. CI/CD integration enables the acceleration of new software releases, which in turn reduces the time to deploy bug corrections.

Although a carefully implemented ZSM will enhance the security of the network, it also has risks [44]. A vulnerability in the ZSM management service could be compromised and used by an adversary to attack other management services or the ZSM framework. Broken access control could allow an adversary to exhaust the management resources, causing a DoS of the ZSM framework, tamper with management policies, or exfiltrate data. Where the ZSM framework runs in a multi-tenant environment, a vulnerability in the ZSM framework could be used to circumvent isolation mechanisms, which could cause the loss of sensitive data or of a ZSM service.

Finally, the ZSM framework will operate across diverse management and service domains that will require trust relationships to be established among them. Without these

trust relationships, the full benefits of zero-touch automation will be impossible to achieve. The trust models defined in NIST 800-39 and ETSI NFV are static and do not adapt to the dynamics of a ZSM framework [45]. ETSI ZSM defines a reflective and adaptive trust model to build mutual trust between entities inside a management domain or different domains of ZSM framework based on a common trust entity that can “vouch” for the trust level of each domain in the form of a chain of trust, trust profile, and trust assurance [45]. With this information, each management or service domain can create authentication and access policies to enforce on each domain.

CSPs and vendors should use the ETSI ZSM and other closed-loop automation frameworks as the basis of 5G closed-loop automation. Closed-loop automation should focus on strong implementations of ZT controls, supply chain security, multi-tenancy isolation controls, and establishing trust relationships between different networks and management domains. The most significant area for further study is security threat analysis and the development of countermeasures to protect AI/ML services and functions. The work underway in the ETSI ISG SAI should be studied and adopted as part of ZSM. Finally, as recommended by ETSI, the ZSM platform reference architecture should be enhanced with capability that allow the supervision and audit of the AI/ML decisions, especially those that affect privacy and security.

8.3 **AI/ML**

ML systems are a new area of attack. There are frequent reports of ML systems being tricked or evaded. Cyberattacks increasingly use data poisoning, model theft, or adversarial attacks. Adversarial attacks on ML-enabled closed-loop automation may cause the closed-loop automation framework to degrade network performance, leak proprietary/personal information, or stop working entirely. The inability of the AI/ML to detect spoofed network data will allow an adversary to misrepresent the state of the network, leading to network performance issues. The opaqueness of ML models can prevent a CSP from understanding why a poor automation action was taken. The unpredictability of AI/ML results and the tendency toward biased results can also be leveraged by an adversary. Data poisoning can cause the AI/ML to trigger a malfunction in the network or closed-loop automation framework. An adversary with knowledge of a closed-loop automation ML model can also trigger network or closed loop malfunction. AI/ML in closed-loop automation will have to address the security of data supply chain, model supply chain, model deployed in shared framework, interaction between multiple domains, and trust between AI/ML service producer and consumer.



9. ZT POLICY MANAGEMENT

9.1 Introduction

An attack surface may be defined as the sum of all possible security risk exposures and vulnerabilities (known and unknown) and controls across all hardware, software, and network components. The attack surface of an application is the sum of all paths for data/commands into and out of the application, and the controls that protect these paths, including resource connection and authentication, authorization, activity logging, data validation, and encoding.

Critical to these controls are the access and other policies they enforce. A collection of the policies is a logical system, and as such can use techniques of formal logic to prove an assertion about that system true or false. This is important to a ZTA because the ability of the system to prevent an attack can be proven by reasoning about the policies using automated proof techniques. Policy management is an important factor to achieve ZT that will continue to evolve as policies increase in number and complexity. This section will introduce potential solutions being developed to achieve consistency in policy management to achieve a ZTA. There are a number of current systemic problems in policy management, which can be addressed by current and future work in cognitive systems. Five important areas are:

1. Access control policy makers and implementers use different technical languages for expressing their policies and associated behaviors that have not been addressed by current policy management approaches.
2. Context and situational awareness are not considered.
3. As networks become more decentralized and autonomous, network entities need to make decisions locally to maximize the network performance amid uncertainty of network environment.
4. Policy conflict detection and remediation is overly complex.
5. Systems are unable to perform predictions and converge toward implementing a set of optimal policies to manage behavior.

It is difficult for current policy-based systems to perform predictions and converge toward implementing a set of optimal policies to manage behavior. For example, a service provider may have a business rule to optimize revenue based on different types of users and the applications that they are using. This can become complex because different users often have different requirements for the same application

(e.g., one user wants low-cost voice calls while another user wants high reliability and security and doesn't care about cost).

The following subsections provide an overview the roles of formal logic, explainable AI (XAI), and Explainable Security (XSec) in ZT and examine the NIST Policy Machine, the ANSI/INCITS Next Generation Access Control (NGAC), developments in cognitive networks, and the implications of these initiatives on ZT.

9.2 Formal Logic in Zero Trust

Formal logic and cognitive mechanisms help make policies, a key component of a ZTA, more consistent and **explainable**. This has three profound implications for building more secure systems.

1. Policy-driven behavior can be proven to have a set of desired features, such as being able to protect against a set of specific threats (see sections 8.2.3 and 8.2.4).
2. Proofs about the policy decisions made by 5G and security systems can be used to explain these decisions.
3. Policy-driven behavior can now inspire trust.

Mathematical proofs show that the stated assumptions logically guarantee a conclusion in a rigorous and formal manner. Formal logic is a specific type of mathematical proof that can help prove security features by verifying that the system expressing the features satisfies certain properties or specifications and by detecting any errors, inconsistencies, or vulnerabilities in the system.

Logical proofs also provide human-understandable explanations of the decisions made by software or policy-intensive systems, such as a ZTA or a 3GPP system, because the proofs include the reasons for outcomes, as well as limitations and uncertainties in the proof. They can also help verify the accuracy, robustness, and privacy of a ZTA or 3GPP system, and detect any potential vulnerabilities, biases, and errors. They can also be used to improve the performance and quality of a ZTSA or 3GPP system by helping users and stakeholders to challenge and correct any inconsistencies discovered by proofs.

9.3 Explainable AI in Zero Trust

Understanding the actions of security systems augmented with AI is even more challenging. XAI is a type of AI that can explain the purpose, rationale, and decision-making process of an AI-based system or mechanism in a way that

can be understood by humans. XAI is a promising security technology because it may help security operations assess the potential threats and reduce alert fatigue. XAI can help ZT by providing transparent and interpretable explanations of the AI models used to evaluate the risk associated with each access request. XAI can also help monitor and audit the AI security models for accuracy, fairness, and compliance, and can help mitigate the risks of bias and drift in AI systems.

XSec applies formal logic to security controls. XSec and XAI can be used to reason about ZT controls. XSec is a new paradigm in security research that tries to provide explanations for decisions made by security systems, especially those that involve AI components. Similar to XAI, XSec can help users and stakeholders to understand, trust, and manage the security systems, and to verify, detect, and correct any errors, vulnerabilities, or attacks. The importance of XAI and XSec will increase as governments, SDOs, and others demand accountability for the emerging generation of AI systems. Some examples of XSec that are applicable to both ZT and 3GPP systems include listing the factors that contributed to an access decision and the confidence in flagging an event as an anomaly. Additionally, XSec can be used to construct better security policies and systems.

9.4 NIST Policy Machine Extensions

The Policy Machine (PM), defined in [46], extends the NIST ZT logical architecture, as shown in Figure 9.4-1.

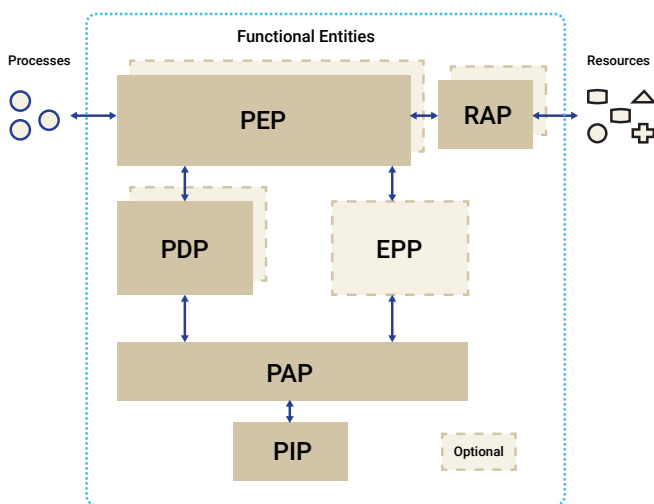


Figure 9.4-1. Policy Machine and NGAC ZT Logical Architecture [46]

The expansion of policy language can enhance 5G ZT by enabling mathematically proven security using a richer policy language that allows the consideration of more factors to make an access decision. The second extension allows the PM access control model to use basic principles of deontic logic in policies, including prohibited (e.g., the suppression of access rights), permitted (e.g., allowed or not forbidden), and obligatory (e.g., a requirement to grant access rights) operations. Deontic logic provides a framework to prove that access rights are handled in a logically consistent manner.

The PM model also differentiates between operational and administrative actions (i.e., those access requests for the creation and maintenance of policy elements). This is analogous to separating the data, control, and management planes in networking. It also follows a fundamental principle of software engineering – the Single Responsibility Principle – which enhances system flexibility by allowing each policy to be replaced, updated, or modified without affecting other policies.

A third extension is that more than one PEP may exist to service access requests from a Subject Actor. Alternatively, a PEP may be dedicated to a particular type of access request. Similarly, more than one PDP may exist to service access requests from a PEP, or a PDP may be dedicated to a particular type of access request. This combination of features provides more manageability over access to particularly sensitive information through the use of dedicated policy-based mechanisms.

A fourth extension is the insertion of multiple entities to better isolate the elements of an access request. The PDP obtains the information needed to verify the access request from the Policy Information Point. A further extension is that once the access is granted, the Resource Access Point (RAP) allows one or more PEPs to gain access to protected resources. The RAP controls access to all Target Actors. Multiple RAPs can exist, but each Target Actor is accessible only through a single RAP.

A final extension is that the PEP and PDP generate events describing the access request, which are sent to the Event Processing Point (EPP) for eventual processing. The state of the Policy Machine may dynamically change as a consequence of different types of access control decisions.

The EPP functions like a transaction processing monitor and avoids contention with active PEPs accessing the same PDP. It also enables access control decisions to be realized as events, which facilitates its integration with event-driven systems.

9.5 Next-Generation Access Controls (NGAC) Extensions to the Policy Machine

The Policy Machine is designed in support of, and in alignment with, a NIST-led American National Standards Institute/International Committee for Information Technology Standards (ANSI/INCITS) standard under the title of Next Generation Access Control (NGAC).

NGAC is a fundamental reworking of traditional access control to better suit the needs of modern, distributed, interconnected systems. Each NGAC command describes specific changes made to the authorization state of an entity. Behavior is controlled using predicates that define the required behavior and may be augmented by pre- and post-conditions and invariants to provide semantic clarity of the required behavior. Interoperability is enhanced by defining a standard set of data elements and relations that can be configured to express access control policies in support of a wide variety of data services and applications. It also includes

a generic set of operations that can be performed on resource data, as well as administrative operations for configuring the data elements and relations that represent policies. These features collectively overcome the following limits of the Policy Machine and other access control schemes:

- > The standard set of data elements, relations, and operations offer support for a microservice-based implementation, which is not possible in the Policy Machine due to its monolithic architecture. This also improves scalability (e.g., by adding or removing microservices as needed).
- > NGAC uses a distributed database, making it easier to handle large amounts of data. In contrast, the Policy Machine uses a centralized database, which can become a performance bottleneck.

NGAC is designed to enable simultaneous instantiation of multiple access control policies. This enables it to guard against insider access by using multiple fine-grained, specific policies that are instantiated simultaneously. This avoids the very difficult problem of trying to write a single “universal” access control policy that covers every variation in the changing context in which it can be executed.

NGAC also offers greater operational efficiency because it computes decisions differently by applying a single combining algorithm over applicable policies that don't conflict. With NGAC, computation of a decision is through an algorithm that is linear. The information necessary in computing an access decision can reside in memory that is updated to reflect each administrative change.

The increased syntactic and semantic specificity of policy languages provided by NGAC provides improved computational efficiency, increased preciseness, and enhanced interoperability between different policies and policy-based systems. The NGAC uses an ANSI access control model that can express arbitrary access control policies modeling access decisions as a graph. This enables a systematic, policy-consistent, and scalable approach to access control. NGAC can help protect against insider attacks by allowing fine-grained control over user permissions, enforcing separation of duty and least privilege principles, and supporting audit and review of user activities.

9.6 Future Areas of Study for Policy Management

Cognition is the process of acquiring and understanding data and information and producing new data, information, and knowledge. In a cognitive system, the knowledge repositories are always changing as new information is acquired and validated, and previous information is updated.

Different technical languages and terminologies make it difficult to equate the syntax and especially the semantics of policies for different constituencies. Conceptually, mappings are needed between the language and terminology of each constituency. In addition, the lack of collaboration between access control policy makers and implementers may lead

to a misunderstanding of security policy semantics. This is because policy makers and implementers use different technical languages for communication, which leads to miscommunication between the two parties. For example, security breaches have resulted from a lack of understanding of the semantics of access control policies and the implications of policy settings that protect information assets. What is needed is a consensual centralized data dictionary for mapping different terminologies of each constituency to each other. This should be augmented by a lexicon, so that ML approaches can be used to ensure that the mappings are correct by aligning the meaning of each set of mapped terms.

Most systems use static policies, which are policies that are pre-defined to solve a particular problem. However, network systems are constantly changing because their context is a combination of current business rules, regulatory requirements, customer SLAs, environmental conditions, current mix of user applications, and other factors. A static policy is built to control the behavior of the system for a particular context and may be at best irrelevant and at worst detrimental for the current context. This may render the behavior from a set of statically defined policies in conflict with the behavior needed for a new context.

A contextually aware policy is a policy that is matched to the current context, where context is defined as the collection of measured and inferred knowledge that describe the environment in which an entity exists or has existed. Situation awareness is the perception of data and behavior that pertain to the relevant circumstances and/or conditions of a system or process, the comprehension of the meaning and significance of these data and behaviors, and how processes, actions, and new situations inferred from these data and processes are likely to evolve in the near future. Both features are also being worked on in ETSI ENI.

As networks become more decentralized and autonomous, network entities need to make decisions locally to maximize the network performance under uncertainty of the overall network environment. This requires a distributed policy management system. This subject is beyond the scope of this white paper because it raises technical issues. For example, if there are multiple PDPs in a system, how are policies managed? Ideally, these PDPs could collaborate, but that brings up trust issues and may widen the attack surface unless special measures are taken. Another example is extending a policy management architecture to a distributed multi-domain system, where different domains are used to isolate different policy functions (e.g., separate monitoring policies from configuration policies). This enables more fine-grained policies to be used to manage the behavior of entities in a domain but increases management and administration costs.

Policy conflict detection and remediation increases in complexity as both the number of active policies and the complexity of the behavior controlled by policies increases. For example, RBAC enforces role assignment, role authorization, and permission authorization by a technique known as role engineering. Advanced systems add elements of ABAC into RBAC by, for example, using attributes to define the different role engineering functions in more detail.



10. ASSESSMENT OF 3GPP SA3 ZERO TRUST SECURITY ACTIVITIES

10.1 Introduction

In June 2022, 3GPP's SA3 work group started a study to align the 5G architecture with NIST's Zero Trust Architecture [47]. This study is still ongoing as of the publication of this report.

Based upon this work group's review of the SA3 study, this ATIS study group has created a table to compare NIST's tenet descriptions to those from the SA3 study. In addition, the table also includes recommendations for 3GPP SA3 to address each tenet. A major identified gap is that the SA3 study considers ZTA for only the 5GC and does not consider ZTA for the UE and RAN domains. The recommendation from this report is that the SA3 study should expand their ZTA efforts to encompass the entire 5GS consisting of the UE, RAN, and core domains. The SA3 study [47] currently has the following preliminary conclusions:

- > NIST ZTA Tenets in-scope for 3GPP 5GC and no further standardization work is needed
 - > Tenets 1, 2, 3, 6
- > NIST ZTA Tenets out-of-scope for 3GPP 5GC due to operational dependencies
- > Tenets 4, 5, 7
 - > NOTE: Potential for further analysis on need for standardization of monitoring

10.2 Gap Analysis

Table 10.2-1 and text below provide recommendations to 3GPP, including recommendations to expand the ZT study to be more encompassing. The legend in front of the table explains the various use of symbols and the recommendation numbering in use.

Legend			
Not included in the 3GPP SA3 Study	✗	ATIS does not Recommend Further Study	-
Included in the 3GPP SA3 Study	✓	ATIS Recommends Additional Study	+
Global ATIS Recommendation	*Rx	Specific ATIS Recommendation	Rx.y

NOTE: 3GPP TR 33.894 addresses 5GC only. 5G RAN and UE were not in-scope.

ZT Tenets	NIST and 3GPP Descriptions	3GPP TR 33.894 Study	ATIS Recommended Future 3GPP Study			ATIS eZT5G Work Group Recommendations
		Core	UE	RAN	Core	
T1	NIST 800.207 – All data sources and computing services are considered resources. 3GPP TR 33.894 – Any Network Function and their services in the 5G Core are considered a resource.	✓	+	+	-	R1.1 = 3GPP should include the NIST ZT architectural components (e.g., PEP, PDP).
T2	NIST 800.207 – All communication is secured regardless of network location. 3GPP TR 33.894 – All the 5GS Core network communications should be done in the most secure manner available - with confidentiality, integrity, and source authentication (as applicable).	✓	+	+	-	R2.1 = Specifically further study (see R0.1) should investigate the wireless data communications between the UE and 5G radio. The 3 recommendations below are just some ex-amples: R2.1.1 = The Access Stratum's (AS) Signaling Radio Bearers (SRB) and Data Radio Bear-ers (DRB) should be evaluated as it relates to T2. R2.1.2 = The Non-Access Stratum (NAS), NG(N2/N3), F1, E1 and Xn communications should also be evaluated. R2.1.3 = As device credentials technologies (e.g. SIM, eSIM, USIM, ISIM) evolve, 3GPP ZT should address evolving related security gaps

Continued on next page - Table 10.2-1. NIST ZT Tenet and 3GPP ZT Study Gap Analysis and Recommendations

ZT Tenets	NIST and 3GPP Descriptions	3GPP TR 33.894 Study	ATIS Recommended Future 3GPP Study			ATIS eZT5G Work Group Recommendations
		Core	UE	RAN	Core	
T3	<p>NIST 800.207 – Access to individual enterprise resources is granted on a per-session basis.</p> <p>3GPP TR 33.894 – This tenet is about access authorization to resources.</p>	✓	+	+	+	<p>R3.1 = 3GPP should study how all access requests that consume network resources, such as the establishment of signaling/data channels, could be authenticated and authorized on a per session basis with a least privilege approach.</p> <p>R3.2 = 3GPP, in particular, should expand ZTA study to further include areas such as access authentication and authorization to Network Slicing and Edge Compute resources.</p> <p>R3.3 = 3GPP should accelerate enhancements to support PKI-based mutual authentication for machine-to-machine communications such as zero touch automation environments and sessions between network functions.</p> <p>R3.4 = 3GPP should study opportunities to introduce multi-factor authentication (MFA) for human user access.</p>
T4	<p>NIST 800.207 – 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.</p> <p>3GPP TR 33.894 – Tenet 4 was entirely omitted in the latest draft publication - Study basically skipped the evaluation of this tenet.</p>	✗	+	+	+	<p>R4.1 = To support dynamic policy, 3GPP should explore applying ZTA to access a 5G network function's (i.e., Producer) application/service using the network function consumer's identity and any additional attributes as it relates to this Tenet.</p> <p>R4.2 = 3GPP should enhance the identity and access management specifications for network functions to provide dynamic identity security assessment of ZTA policy decision.</p> <p>R4.3 = 3GPP should study how to (re)evaluate and revoke authorization permissions granted to network function(s) when deemed necessary, for example when a network function is identified as compromised. The study should also address mechanisms to enable notification of the revocation of authorization permissions in a timely manner to concerned network functions.</p>
T5	<p>NIST 800.207 – The enterprise monitors and measures the integrity and security posture of all owned and associated assets.</p> <p>3GPP TR 33.894 – Study states the NIST description of the tenet which implies their agreement for this study.</p>	✗	+	+	+	<p>R5.1 = 3GPP should develop support for security monitoring, measuring integrity, and security posture of all 5GS assets (UE, RAN and Core) such as behavioral analytics, configuration management, software integrity checks, vulnerability scans, security anomaly detection, etc.</p>
T6	<p>NIST 800.207 – All resource authentication and authorization are dynamic and strictly enforced before access is allowed.</p> <p>3GPP TR 33.894 – In the 5G Core context, this tenet also relates to how the access by service consumers to the services of producers is secured.</p>	✗	+	+	+	<p>R6.1 = 3GPP should study, in addition to the strict enforcement of authentication and authorization of access to network resources, mechanisms to realize periodic re-evaluation of authentications and authorizations based upon for example dynamic policies and/or other relevant information.</p> <p>R6.1.1 = This study could be the reauthentication and reauthorization of the UE's access to network resources for Network Slicing, Edge Compute, etc.</p> <p>R6.1.2 = This study should include how an MNO remotely accesses the UE to perform an over-the-air (OTA) update on the SIM.</p>
T7	<p>NIST 800.207 – 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.</p> <p>3GPP TR 33.894 – The tenet describes that "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..."</p>	✗	+	+	+	<p>R7.1 = 3GPP should develop the capability to collect security related metrics, events, and data from 3GPP functional elements. This would be similar intent to the performance based NWDAF.</p> <p>R7.1.1 = The new security capability would focus on the security aspects and posture of the individual components and functions of the RAN and Core. This could be introduced, for example, as an extension to TS 23.288.</p> <p>R7.2 = 3GPP should develop security configuration validation and checking specifications for the 5GS.</p>

Table 10.2-1. NIST ZT Tenet and 3GPP ZT Study Gap Analysis and Recommendations

10.3 Consolidated Recommendations for 3GPP

5G is the most secure generation of standardized mobile technology available commercially today. 5G is specified with many features of a ZTA but it does not embody a full implementation of a ZTA. Multiple use cases for ZT in 5G are described in this report, and these should be used to guide further enhancement of ZTA in 5G standards. The recommendations listed below provide areas for enhancement of ZTA in a 5GS, as introduced in Table 10.2-1. These recommendations are a collection of the various recommendations and security control groups that have been discussed thus far in this report within scope of 3GPP.

10.3.1 General Recommendations

- R0.1** = 3GPP should extend ZT beyond the core network (CN) and SBA to cover UE and RAN components of the 5GS.
- R0.2** = 3GPP should coordinate with appropriate SDOs to integrate non-3GPP ZT solutions with 3GPP's ZT solution.
- R0.3** = 3GPP should extend the core network's ZTA to encompass all core network functions and deployment models.
- R0.4** = 3GPP should support ZT policy tools and functions to enable an operator to develop and enforce ZT policies (e.g., RBAC, ABAC, etc.) more effectively.
- R0.5** = 3GPP's ZTA should be extended to allow use of IT enterprise security tools and functions as part of ZT policy decisions
- R0.6** = As 3GPP develops a policy decision framework to support a ZTA, it should allow the flexibility for the operator to use its choice of different policy decision approaches and architectures (e.g., RBAC, ABAC, NGAC).
- R0.7** = 3GPP should study effective testing, and potentially define SCAS, for 3GPP ZT enhancements to 5G.

10.3.2 Tenet 1 – All Data Sources and Computing Services are Considered Resources

- R1.1** = 3GPP should be built upon the NIST ZTA components (e.g., PEP, PDP).

10.3.3 Tenet 2 – All Communications is Secured Regardless of Network Location

- R2.1** = Specifically, further study (see R0.1) should investigate the wireless data communications between the UE and 5G radio. The three recommendations below

are just some examples:

- R2.1.1** = The Access Stratum's (AS) Signaling Radio Bearers (SRBs) and Data Radio Bearers (DRBs) should be evaluated as it relates to T2.

- R2.1.2** = The Non-Access Stratum (NAS), NG(N2/N3), F1, E1, and Xn communications should also be evaluated.

- R2.1.3** = As device credentials technologies (e.g., SIM, eSIM, USIM, ISIM) evolve, 3GPP ZT should address evolving related security gaps.

10.3.4 Tenet 3 – Access to Individual Resources is Granted on a Per-Session Basis

- R3.1** = 3GPP should study how all access requests that consume network resources, such as the establishment of signaling/data channels, could be authenticated and authorized on a per-session basis with a least privilege approach.

- R3.2** = 3GPP, in particular, should expand ZTA study to further include areas such as access authentication and authorization to network slicing and edge compute resources.

- R3.3** = 3GPP should accelerate enhancements to support PKI-based mutual authentication for machine-to-machine communications such as ZT automation environments and sessions between network functions.

- R3.4** = 3GPP should study opportunities to introduce multi-factor authentication (MFA) for human user access.

10.3.5 Tenet 4 – Access to Resources is Determined by Dynamic Policy

- R4.1** = To support dynamic policy, 3GPP should explore applying ZTA to access a 5G network function's (i.e., Producer) application/service using the NF consumer's identity and any additional attributes as it relates to this tenet.

- R4.2** = 3GPP should enhance the identity and access management specifications for network functions to provide dynamic identity security assessment of ZTA policy decision.

- R4.3** = 3GPP should study how to (re)evaluate and revoke authorization permissions granted to NF(s) when deemed necessary, for example when a NF is identified as compromised. The study should also address mechanisms to enable notification of the revocation of authorization permissions in a timely manner to concerned network functions.

10.3.6 Tenet 5 – Monitoring/Measuring the Integrity and Security Posture of all Assets

- R5.1** = 3GPP should develop support for security

monitoring, measuring integrity, and security posture of all 5GS assets (UE, RAN, and core), such as behavioral analytics, configuration management, software integrity checks, vulnerability scans, security anomaly detection, etc.

10.3.7 Tenet 6 – All Resource Authentications/Authorizations are Dynamic and Strictly Enforced Before Access is Allowed

R6.1 = 3GPP should study, in addition to the strict enforcement of authentication and authorization of access to network resources, mechanisms to realize periodic re-evaluation of authentications and authorizations based upon for example dynamic policies and/or other relevant information.

R6.1.1 = This study could be the reauthentication and reauthorization of the UE's access to network resources for network slicing, edge compute, etc.

R6.1.2 = This study should include how an MNO remotely accesses the UE to perform an over-the-air (OTA) update on the SIM.

10.3.8 Tenet 7 – Improving the Security Posture of all Assets, Network Infrastructure, and Communications

R7.1 = 3GPP should develop the capability to collect security related metrics, events, and data from 3GPP functional elements. This would be similar intent to the performance based NWDAF.

R7.1.1 = The new security capability would focus on the security aspects and posture of the individual components and functions of the RAN and Core. This could be introduced, for example, as an extension to TS 23.288.

R7.2 = 3GPP should develop security configuration validation and checking specifications for the 5GS.



11. CONCLUSIONS AND NEXT STEPS

This ATIS report was completed by a collective group of industry experts from equipment providers, communication service providers, transport providers, industry think tanks, and government agencies. After months of research, analysis, and discussions, the work group created a list of actionable recommendations that the various industry bodies should consume to publish specifications, recommendations, and/or guidance that are more aligned with the principles of ZT. This paper identified 12 use cases for ZT in 5G and 12 ZTA security control groups that can be implemented to provide protection from external and internal threats for each of the use cases.

ZTA is a plan based upon the concept of ZT. It is imperative that the 5G networks, as critical infrastructure, strive towards the goal of ZTA for the data, control, and management planes in the 5GS, including RAN and core. As noted in this report, each of the NIST's seven tenets for ZT can be applied to a 5GS ZTA. Any tenet that may be considered by 3GPP to be outside of its scope should be addressed in another relevant industry body, such as ATIS. NIST's ZT logical components include a PDP and PEP that may be implemented as standalone NFs or logical functions within an asset, such as a NF, that serves as a micro-perimeter.

Multiple U.S. federal agencies are addressing ZT within the various individual agency environments and for critical infrastructure, including cellular communications. The relevant agencies for 5G and ZT are the White House's ONCD, the Department of Commerce's NIST, the Department of Homeland Security's CISA, and the National Security Agency's ESF. ESF's "Security Guidelines for 5G Cloud Infrastructures" provides a playbook for adapting NIST ZTA to 5G that should be addressed by the relevant 5G industry bodies.

It is recommended that the ATIS NextG Alliance extend the security gap analysis documented in this paper to ZTA requirements for 6G. Areas for further study for ZTA in 6G mobile networks include Continuous Monitoring, Anomalous Behavior Detection, Policy Management, TDR/EDR, Threat Intelligence, and SIEM/SOAR integration. Cloud security best practices are also evolving to support the security needs of 5G and other critical infrastructure and should be included in any future enhancements to 6G ZTA.

In summary, the ATIS 5G ZT study group has the following recommendations for next steps:

- > Identify the appropriate industry body to address any NIST ZT tenets applicable to 5G networks and beyond that are outside the scope of 3GPP.
- > Bring ESF's "Security Guidelines for 5G Cloud Infrastructures" guidelines adapting NIST ZTA into 5G and 6G supporting standards.
- > NextG Alliance extend ZTA from 5G described in this paper to 6G, including Continuous Monitoring, Anomalous Behavior Detection, Policy Management, TDR/EDR, and Threat Intelligence.
- > Evolve cloud security best practices for 5G critical infrastructures.

These important activities will support significant gains in 5G and future 6G network security to support a ZTA as the network threat landscape continues to evolve.



REFERENCES

- [1] S. Rose, O. Borchert, S. Mitchell, S. Connelly, "Zero Trust Architecture", NIST Special Publication 800-207, August 2020 <https://csrc.nist.gov/publications/detail/sp/800-207/final>
- [2] Security Guidance for 5G Cloud Infrastructures, volume I, NSA ESF and DHS CISA, October 2021, https://www.cisa.gov/sites/default/files/publications/Security_Guidance_For_5G_Cloud_Infrastructures_Part_I_508_Compliant.pdf
- [3] <https://www.whitehouse.gov/wp-content/uploads/2023/03/National-Cybersecurity-Strategy-2023.pdf>
- [4] S. Rose, O. Borchert, S. Mitchell, S. Connelly, "Zero Trust Architecture", NIST Special Publication 800-207, August 2020 <https://csrc.nist.gov/publications/detail/sp/800-207/final>
- [5] https://media.defense.gov/2021/Nov/18/2002895143/-1/-1/0/SECURITY_GUIDANCE_FOR_5G_CLOUD_INFRASTRUCTURES_PART_II_20211118.PDF
- [6] https://media.defense.gov/2021/Dec/01/2002901540/-1/-1/0/SECURITY_GUIDANCE_FOR_5G_CLOUD_INFRASTRUCTURES_PART_III_508%20COMPLIANT.PDF
- [7] https://www.cisa.gov/sites/default/files/publications/Security_Guidance_For_5G_Cloud_Infrastructures_Part_IV_508_Compliant.pdf
- [8] <https://csrc.nist.gov/Projects/devsecops>
- [9] <https://csrc.nist.gov/publications/detail/sp/800-204c/final>
- [10] <https://csrc.nist.gov/publications/detail/sp/800-218/final>
- [11] [Zero Trust Maturity Model, v2.0. US DHS CISA, April 2023, Zero Trust Maturity Model Version 2.0 \(cisa.gov\)](#)
- [12] [Executive Order on Improving the Nations Cybersecurity, EO 14028, US White House, May 12, 2021, Executive Order on Improving the Nation's Cybersecurity | The White House](#)
- [13] [Zero Trust Strategy, US DoD, Oct 2022, DoD Zero Trust Strategy \(defense.gov\).](#)
- [14] "Final Report – Recommendations to Identifying Optional Security Features that can Diminish the Effectiveness of 5G Security", FCC CSRIC VII WG3, March 2021.
- [15] <https://www.fcc.gov/file/20606/download>
- [16] Top 10 Proactive Controls, OWASP
- [17] Top 10 Proactive Controls, CIS
- [18] Top Threats to Cloud Computing: The Pandemic Eleven, Cloud Security Alliance, 2022.
- [19] Security Strategies for Microservices-based Application Systems, NIST SP 800-204, US DoC NIST.
- [20] [Kubernetes Hardening Guidance, NSA ESF and US DHS CISA, August 2022, CTR_KUBERNETES_HARDENING_GUIDANCE_1.2_20220829.PDF \(defense.gov\)](#)
- [21] Open RAN, NSA ESF and US DHS CISA, Sept 2022.
- [22] Report on Open RAN, CSRIC VIII WG2, US FCC, Dec 2022.
- [23] [Artificial Intelligence Risk Management Framework, NIST AI RMF 1.0, US DoC NIST, January 2023. Artificial Intelligence Risk Management Framework \(AI RMF 1.0\) \(nist.gov\)](#)
- [24] [Securing Artificial Intelligence \(SAI\): Mitigation Strategy Report, ETSI GR SAI 005, v1.1.1, March 2021, GR SAI 005 - V1.1.1 - Securing Artificial Intelligence \(SAI\): Mitigation Strategy Report \(etsi.org\)](#)
- [25] NIST AI Risk Management Framework <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf>



REFERENCES

- [26] NSA ESF Report on 5G Network Slicing, Dec 2022.
- [27] E2E Network Slicing Architecture, GSMA, June 3 2021.
- [28] Potential Threats to 5G Network Slicing, NSA ESF and CISA, Dec 13 2022
- [29] NIST CISA Defending Against Software Supply Chain Attacks (April 2021) - https://www.cisa.gov/sites/default/files/publications/defending_against_software_supply_chain_attacks_508_1.pdf
- [30] NIST Computer Security Resource Center, Secure Software Development Framework (SSDF) SP 800-218 (Feb 2022) - <https://csrc.nist.gov/Projects/ssdf/publications>
- [31] OpenSSF Security Scorecards - Security health metrics for Open Source, <https://github.com/ossf/scorecard>
- [32] TS 31.102 Universal Mobile Telecommunications System (UMTS); LTE; 5G; Characteristics of the Universal Subscriber
- [33] Identity Module (USIM) application, https://www.etsi.org/deliver/etsi_ts/131100_131199/131102/16.04.00_60/ts_131102v160400p.pdf
- [34] Network Functions Virtualisation (NFV); Virtualisation Technologies; Report on the application of Different Virtualisation Technologies in the NFV Framework, ETSI GS NFV-EVE 004, v1.1.1, March 2016, https://www.etsi.org/deliver/etsi_gs/NFV-EVE/001_099/004/01.01.01_60/gs_nfv-eve004v010101p.pdf
- [35] TS 23.548 5G System Enhancements for Edge Computing
- [36] TS 23.558 Architecture for enabling Edge Applications
- [37] TS 28.538 Management and orchestration for Edge Computing Management
- [38] TS 33.558 Security Aspects of Enhancement of Support for Enabling Edge Applications
- [39] "Operations of the future: Reaching for the North Star of zero-touch operations." <https://www.ericsson.com/4a80d9/assets/local/managed-services/documents/05102021-operations-of-the-future-report.pdf>
- [40] "Zero touch network & Service Management (ZSM)." <https://www.etsi.org/technologies/zero-touch-network-service-management>
- [41] "Industry Specification Group (ISG) Zero Touch Network and Service Management (ZSM)." <https://www.etsi.org/committee/zsm>
- [42] Liyanage, Madhusanka, et al. "A survey on Zero touch network and Service Management (ZSM) for 5G and beyond networks." Journal of Network and Computer Applications 203(11). March 1, 2022. https://www.researchgate.net/publication/358976899_A_survey_on_Zero_touch_network_and_Service_Management_ZSM_for_5G_and_beyond_networks
- [43] "Zero-touch network and Service Management (ZSM); Reference Architecture." https://www.etsi.org/deliver/etsi_gs/ZSM/001_099/002/01.01.01_60/gs_ZSM002v010101p.pdf
- [44] "Zero-touch network and Service Management (ZSM); Requirements based on documented scenarios."
- [45] https://www.etsi.org/deliver/etsi_gs/ZSM/001_099/001/01.01.01_60/gs_ZSM001v010101p.pdf
- [46] "Zero-touch network and Service Management (ZSM); General Security Aspects." https://www.etsi.org/deliver/etsi_gr/ZSM/001_099/010/01.01.01_60/gr_ZSM010v010101p.pdf
- [47] MEF, "Zero Trust Framework for MEF Services", October 2022. Available at <https://www.mef.net/resources/mef-118-zero-trust-framework-for-mef-services/> [PMNIST] D. Ferraiolo, S. Gavrila, W. Jansen, "Policy Machine: Features, Architecture, and Specification", NISTIR 7987 revision 1, October 2015
- [48] 3GPP SA3 TR 33.894 Study on Zero-Trust Security Principles in Mobile Networks - <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=4086>

COPYRIGHT
AND
DISCLAIMER

ATIS-I-0000095 Published June 2023

Copyright © 2023 by Alliance for Telecommunications Industry Solutions

All rights reserved.

Alliance for Telecommunications Industry Solutions
1200 G Street, NW, Suite 500
Washington, DC 20005

No part of this publication may be reproduced in any form, in an electronic retrieval system or otherwise, without the prior written permission of the publisher. For information, contact ATIS at (202) 628-6380. ATIS is online at <http://www.atis.org>.

The information provided in this document is directed solely to professionals who have the appropriate degree of experience to understand and interpret its contents in accordance with generally accepted engineering or other professional standards and applicable regulations. No recommendation as to products or vendors is made or should be implied.

NO REPRESENTATION OR WARRANTY IS MADE THAT THE INFORMATION IS TECHNICALLY ACCURATE OR SUFFICIENT OR CONFORMS TO ANY STATUTE, GOVERNMENTAL RULE OR REGULATION, AND FURTHER, NO REPRESENTATION OR WARRANTY IS MADE OF MERCHANTABILITY OR FITNESS FOR ANY PARTICULAR PURPOSE OR AGAINST INFRINGEMENT OF INTELLECTUAL PROPERTY RIGHTS. ATIS SHALL NOT BE LIABLE, BEYOND THE AMOUNT OF ANY SUM RECEIVED IN PAYMENT BY ATIS FOR THIS DOCUMENT, AND IN NO EVENT SHALL ATIS BE LIABLE FOR LOST PROFITS OR OTHER INCIDENTAL OR CONSEQUENTIAL DAMAGES. ATIS EXPRESSLY ADVISES THAT ANY AND ALL USE OF OR RELIANCE UPON THE INFORMATION PROVIDED IN THIS DOCUMENT IS AT THE RISK OF THE USER.



www.atis.org

For information, contact ATIS at (202) 628-6380.