Improving Resilience of Future Mobile Network Generations Implementing Zero Trust Paradigm

Kamyar Abedi[†] Giang T. Nguyen[*‡] Thorsten Strufe[†‡]
[†]Karlsruhe Institute of Technology
[*]Haptic Communication Systems, TU Dresden
[‡]Centre for Tactile Internet with Human-in-the-Loop, TU Dresden

***Abstract*—Using virtualized network management functions, the Service Based Architecture will replace the Reference Point Architecture for managing future mobile networks. It naturally lacks a clear security perimeter and has an increased attack surface, so defending the control plane against attacks requires a novel protection paradigm. Both National Institute of Standards and Technology and 3rd Generation Partnership Project suggest moving from perimeter security to a Zero Trust Architecture (ZTA), authenticating all request initiators and controlling access to all resources for each request. However, it insofar remains somewhat unclear to which extent the suggested management protocols do indeed meet the ZTA. We are exploring the standardized communication management protocols in this paper. Our analysis indicates that with careful implementation, the existing network functions and protocols can indeed achieve comprehensive authentication and access control so that the ZTA can be met.**

I. Introduction

Future 5G and 6G mobile communication systems will leverage Service Based Architecture (SBA). SBA implements network management as a Control Plane (CP) and handles all user traffic of a User Plane (UP), enabling end users to connect reliably and securely to the network and provide access to its services. SBA provides vital functions in mobile networks, such as Session Management, Mobility Management, Authentication, Authorization, and Policy Administration.

As the CP manages all networking assets, it must be considered the most prominent target for attackers who aim to compromise network and communications security. Protecting both CP and signaling has already been difficult in the past. Numerous incidents[1,2] show how adversaries managed to exploit vulnerabilities in the CP, for instance, at the Signalling System 7. The attacks occurred even when the attack surface was small. The employed Perimeter Security Model (PSM) [1] only allowed chosen entities to access the majority of interfaces.

Evolving to a virtualized CP, deployed probably even to shared cloud infrastructures, renders PSM a futile endeavor. Network Functions (NF) become mobile software services, and there is no clear boundary between the CP of the network core and outside networks. The previous trust assumptions about components vanish because adversaries may quickly gain control over processes executed in the virtualization infrastructure and potentially contact arbitrary services, potentially even impersonating seemingly legitimate entities.

Contrary to the Perimeter Security Model, the Zero Trust Architecture (ZTA) [2] assumes that no part of the network can be trusted, and adversaries can be within any network perimeter. Additionally, ZTA assumes all communication over the network to be potentially compromised. Requirements in turn are that the origin of any communication is authenticated before entering any interaction, and that requested access to data or functions is verified for effective permissions and access control is enforced. The approach ensures that insider attackers who have compromised security within a network perimeter can be repelled in their attempts to escalate their privileges or gaining access to sensitive resources. Adopting the ZTA has frequently been suggested as a natural fit for the SBA of future mobile communication. However, there has yet to be an explicit analysis of whether communication management in 5G and 6G networks implements the ZTA.

Several research studies have explored the ZTA and the implementation challenges within 5G/6G networks, presenting several solutions to mitigate these challenges. [3], [4], [5]. Ramezanpour [6] introduced the concept of an Intelligent Zero Trust Architecture (i-ZTA) as a security framework that can enforce ZT principles leveraging Machine Learning (ML) and Artificial Intelligence (AI) algorithms. Authors in [7] analyze all ZTA principles and how those principles and features help to develop a security protection architecture in edge cloud and IoT smart grid. However, the papers do not investigate ZTA applications in 5G/6G networks and their alignment with standards.

In this paper, we investigate the common communication scenarios in 5G SBA, focusing on authentication and access control compared to the ZTA. We show that the suggested SBA protocols reflect the necessary security features and allow for correctly implementing authentication and access control, thus

meeting ZT paradigms.

II. Background

5G Mobile communication systems usually separate traffic into distinct planes, the so-called Control and User Plane Separation (CUPS). The UP transfers user-generated data, i.e., payload, across the network. The CP manages and controls the operation of the network, including setting up, configuration, and routing. The separation allows for flexible deployment models, whether distributed or centralized, and permits the independent scaling of control plane and user plane functions without affecting the functionality of existing network nodes subject to this division. Additionally, alongside the CP and UP, there is a Management Plane (MP) that is exclusively designated for transporting administrative traffic. This clear separation of functions optimizes network performance and ensures efficient traffic management in modern telecommunications systems [8].

A. 5G Service Based Architecture (SBA)

The Service Based Architecture (SBA) for the 5G core network, depicted in Fig. 1, relies on modular NFs deployed as services. These services are designed to be independent of specific hardware, offering flexibility in deployment across centralized data centers, edge computing nodes, or virtualized infrastructure. The interaction between network services follows a Provider/Consumer relationship, with a Service Producer (SP) providing services and a Service Consumer (SC) exploiting them through reactive Request/Response interactions. To support the development and deployment of new services, all services register upon deployment in the core network, exposing their functions through a Service-Based Interface (SBI). The SBA, encompasses essential functionalities for secure session establishment and data forwarding, ensuring data connectivity for mobile devices.

Within the User Plane (UP), the User Plane Function (UPF) processes and forwards user data under the control of the Session Management Function (SMF). The UPF acts as a stable IP anchor point for devices, connecting to external IP networks and facilitating data reachability as devices move within the network. UPF handles tasks such as generating traffic usage reports, packet inspection, and enforcing network and user policies, including traffic gating, redirection, and data rate limitations. When a device is idle, the UPF buffers incoming traffic and triggers a network page to bring the device back into a connected state. The remaining functions are part of the CP.

The SMF manages end-user (device) sessions, including establishing, modifying, allocating, and releasing IP addresses. The SMF is responsible for interacting with the data plane, creating, updating, and removing Protocol Data Unit (PDU) sessions, and managing session context with theUPF.

The AMF controls device registration, authentication, and mobility management and supports encrypted signaling connections. The Unified Data Management (UDM) is the front end for user subscription data, generating device authentication

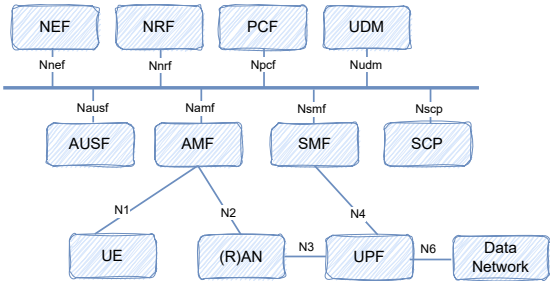

Fig. 1. 5G Service Based Architecture (cf. [9]). Note their organization around a bus rather than specific links per interface.

credentials and authorizing user access based on subscription profile. Authentication Server Function (AUSF) authenticates devices, consuming authentication credentials from UDM. AUSF also generates cryptographic material for secure updates of, e.g., roaming information.

The 5G core also introduces new network functions that did not exist in earlier generations. Network Repository Function (NRF) is the central registry holding comprehensive information for every Network Function, thus serving to discover active NFs. NRF is also the authorization server, exploiting the OAuth2.0 protocol. Service Communication Proxy (SCP) is a new network function managing and routing service-related messages within the 5G core. 5G SBA exploits SCP to expose services to authorized consumers, such as application developers or third-party service providers. SCP resolves NF discovery requests by communicating with the NRF and facilitates indirect communication between two NFs

B. Perimeter Security Model (PSM)

The PSM is the traditional model to protect a network's infrastructure by establishing a secure boundary (or perimeter) around it. This model prevents attacks, threats, and unauthorized access from reaching the internal network, exploiting several components and practices such as firewalls, intrusion detection systems, and access control.

Consequently, the model assumes that all elements within the network are inherently considered trustworthy entities. However, this has proven to be a strong assumption because, in practice, attacks breached the traditional perimeter security measures, such as in SolarWinds Supply Chain Attack (2020) and Log4j Vulnerability (2021). Therefore, the PSM is inadequate to achieve the security objectives against insider attacks [10]. In such cases, the adversaries may move laterally within the network, benefiting from an extended attack surface and exploiting vulnerabilities no longer hidden behind the perimeter. Consequently, they may access additional systems, data, or critical resources. This highlights the need for a more comprehensive and adaptable security model to address the potential risks associated with breaches of the perimeter security model.

C. Zero Trust Architecture (ZTA)

Since the traditional PSM model is ineffective against internal threats, a fundamental shift is required. The concept of ZT, dating back to 2002 as introduced by Sood et al. [2], suggests solutions. Aiming at enhancing intrusion tolerance in computer systems, the authors presented the Self-Cleansing Intrusion Tolerance, combining the principles of ZT and system self-cleansing to protect against unauthorized intrusions. Later, Forrester Research picked up the term ZT [11] for the enterprise domain. They reinforce to never to make any assumptions about the trustworthiness of peers.

The ZTA assumes that all entities are inherently untrusted regardless of their location. Therefore, the model demands entities to verify every request, which requires authentication and access control functionality.

1) Authentication: Authentication is the corresponding process of verifying the identity of a user, system, or entity attempting to access a resource. Authentication ensures that the entity requesting access has the claimed identity [12].

Zero Trust (ZT) operates by validating all requests to access resources, regardless of whether the request is coming from within or outside the network, and ensuring only authorized and approved subjects are granted access to the resources by establishing a secure access network through authentication. The ZT improves security by blocking unauthorized access and preventing lateral movement by attackers already inside the network.

2) Access Control: Upon receiving a request at its interface, every service must control access to its assets, i.e., functionality and data. Access control is the mediation process between the requests of the subjects of a system to perform specific actions on specific objects. The main task of access control now is to decide, based on a defined policy, whether or not a specific access can be permitted and to enforce this decision [12]. The access control policy contains explicit permissions, which can be granted or revoked in the authorization process. Fig. 2 illustrates the access control mechanism, which decides permission ascertainment for every access request. Permissions are directly or indirectly tied to the identities of the requesting entities, requiring verification of whether the entity initiates the request with a claimed identity. Then, the access control calls on whether to grant or deny access; considering the security policies established in the system, access is granted with the least privileges needed to complete the task.

D. National Institute of Standards and Technology (NIST) 800-207 Operationalization

Regarding access control, the NIST 800-207 standard distinguishes between the place of permission verification and its enforcement, where Policy Decision Point (PDP) and Policy Engine Point (PEP) are crucial components. Figure 2 illustrates the logical components and their relationship.

The PDP constitutes the core of the policy framework, acting as the central authority that evaluates resource access requests from subjects based on authentication and authorizations configured in the policy. The Policy Engine (PE) is ultimately responsible for granting access to a resource for a given subject. The Policy Administrator (PA) manages communication between the subject and resources.

PEP plays a critical role in protecting resources by assuming each process resides in its trusted domain and distrusts anything originating from outside of this domain. Hence, it authenticates and controls access to all requests from other processes.

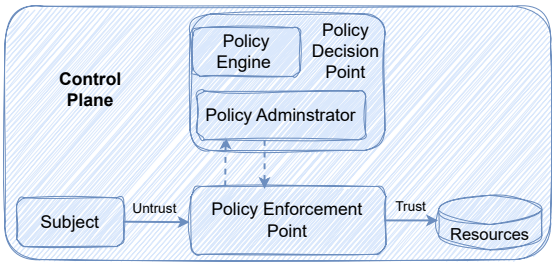

Fig. 2. Access Control Process And Corresponding Components (after [13]).

III. Mapping Zero Trust Architecture to the Service Based Architecture

SBA Domain Security defines mechanisms for secure communication between NFs, employing encryption, authentication, and data integrity mechanisms to prevent eavesdropping, tampering, or interception of sensitive data and control messages. SBA Domain Security also ensures that only verified and authorized network functions and external applications can interact with the network through APIs. Mechanisms such as certificate-based authentication, token-based authorization, and network segmentation can help establish trust and restrict access to trustworthy entities. In the following, we want to investigate the implementation of ZTA within the SBA.

A. Extension to the SBA

To implement ZTA, networks must incorporate authentication and authorization, according to NIST, 800-207, authentication must occur before granting access to the resources by the authorization server. It is required for every request, regardless of the location of the requesting entity, to prove its identity. Therefore, all access requests must be identified and authenticated leveraging Mutual Transport Layer Security (TLS). Then, the access control mechanism, performed by the token-based authorization server running OAuth 2.0, checks whether the request is authorized and then grants permission by issuing a token to the requester based on security policy. In SBA Network Repository Function (NRF) acts like an authorization server, according to 3GPP TS33.501 [9].

B. Mapping ZTA principles to scenarios and 5G SBA core

Scenario 1 (Direct communication): This scenario occurs when an NF consumer requests services directly from an NF

provider. The communication between the two NFs is secured with TLS. Afterward, the NF consumer and provider require token-based authorization for service access. The authorization requires an authorization server,NRF, implementing OAuth 2.0. When a network consumer needs service from a network function provider, the NF consumer requests authorization from the NRF. Compared to NIST Special Publication 800-207, NRF acts as a Policy Decision Point (PDP). The NRF decides whether to authorize the request based on security policies. If the request is authorized, the NRF issues a token with the scope of access to the resource. The NF consumer presents the token to the NF provider to show it has access to the service. The NF provider verifies the token using the NRF's public key since the NRF's private key has signed the token given to the NF consumer.

Scenario 2 (Indirect communication): In this scenario, the NF consumer and NF provider communicate via a Service Communication Proxy (SCP), which receives the NF consumer's requests on the NF provider's behalf, acquires the provider's responses, and replies to the consumer. Instead of merely forwarding requests to the provider, SCP can actively modify the requests. This scenario raises distinguished technical challenges. One of those is that end-to-end mTLS, as with direct communication, is no longer feasible. This requires different techniques to implement ZTA.

There is a solution in the form of token-based authorization, similar to what was used in direct communication. The SCP presents a valid access token issued to the NF consumer. It allows the SCP to prove to the provider that the consumer authorizes the SCP to act on the consumer's behalf. In some deployment scenarios, the SCP can request these access tokens on behalf of the consumer. The authorization must rely on NRF. Only the SCP authorized by the consumer can request these tokens. To achieve this, the consumer sends a self-signed assertion to the SCP. Then, the SCP uses this assertion to prove to the NRF that the consumer authorizes it. During communication between the NF service consumer and producer, the SCP can use the access token to demonstrate that it has the consumer's authorization to represent it.

Scenario 3 (Application Programming Interface (API) Gateway): API exposure in vertical communication empowers developers and third-party apps to seamlessly access and efficiently leverage services and functionalities offered by the 5G network. It opens services to third-party applications through a dedicated NF, Network Exposure Function (NEF) [14] as an API gateway, i.e., the intermediary between the 5G network infrastructure and external applications. It exposes APIs to enable them to leverage the capabilities of the 5G network.

For mapping ZTA into this scenario, NEF authenticates and authorizes third-party requests to the 5G core network. The third-party application sends a request to the NEF. The NEF authenticates the application request using mTLS. The NEF authorizes the application to access the requested capabilities and act as an authorization server based on the network operator's authorization policy. If the application is authorized, the NEF forwards the request to the appropriate NFs. The NFs process the request and return a response to the NEF. The NEF forwards the response to the third-party application.

Scenario 4 (5G roaming): 5G roaming is interconnecting Home Public Land Mobile Nework (HPLMN) and Visiting Public Land Mobile Network (VPLMN) via the Security Edge Protection proxy (SEPP). SEPP is a key network element that provides security for signaling exchange between roaming PLMNs in the 5G SBA core at the control plane level. SEPP performs several security functions, including authentication, authorization of signaling messages, encryption, integrity protection of signaling messages, filtering and policing of signaling messages, and detection and prevention of signaling attacks. SEPP is deployed as a transparent proxy at the core of PLMNs, ensuring secure signaling messages without modifying them.

When the HPLMN wants to communicate with the VPLMN, home's SEPP sends a request to the counterpart SEPP of the VPLMN. The visited network SEPP verifies the authenticity of the incoming request. It checks whether HPLMNs SEPP is a trusted entity. This authentication is often based on mTLS. Post authentication, the visited SEPP verifies if the home SEPP or the HPLMN has the necessary permissions to access the services or data it requests. The SEPP utilizes a token-based authorization system where the requesting SEPP would first obtain an authorization token that implies its right to access specific services. The visited SEPP will validate this token to grant access. The authorization process may involve checking against predefined rules, policies, or agreements between the operators.

In this scenario, we are describing the role of SEPP where two 5G SBA cores connect to provide roaming based on [15], where authentication and authorization mechanisms between HPLMN and VPLMN core need to be in place to meet the requirements of ZTA.

IV. Conclusion

The core of 6G mobile communication systems is expected to evolve to the SBA using virtualized network management functions for the CP. Defending CP against attacks, in this case, requires a novel protection paradigm. Even though NIST and 3GPP suggest moving from the Perimeter Security Model to Zero Trust Architecture, authenticating all request initiators, and controlling access to all resources for each request, it remains unclear to what extent the suggested management protocols meet the zero trust paradigm.

In this paper, we have analyzed the common communication scenarios in the 6G core compared to the requirements of the ZTA. Our analysis indicates that with careful implementation, the existing network functions and protocols can achieve comprehensive authentication and access control to meet the ZTA. This positive finding suggests that the 6G core specifications are already well-aligned with the principles of ZTA.

REFERENCES

[1] C. Riggs, *Network perimeter security: building defense in-depth*. CRC Press, 2003.

[2] A. Sood *et al.*, "Zero trust intrusion containment for telemedicine," George Mason University, USA, Tech. Rep. ADA415878, 2002.

[3] S. Sarkar *et al.*, "Security of zero trust networks in cloud computing: A comparative review," *Sustainability*, vol. 14, no. 18, p. 11213, 2022.

[4] B. Mao, *et al.*, "Security and privacy on 6g network edge: A survey," *IEEE Communications Surveys & Tutorials*, 2023.

[5] H. A. Kholidy *et al.*, "Toward zero trust security in 5g open architecture network slices," in *IEEE Military Communications Conference*, 2022, pp. 577–582.

[6] K. Ramezanpour and J. Jagannath, "Intelligent zero trust architecture for 5g/6g networks: Principles, challenges, and the role of machine learning in the context of o-ran," *Computer Networks*, 2022.

[7] M. A. Alipour *et al.*, "Enabling a zero trust architecture in a 5g-enabled smart grid," *arXiv preprint arXiv:2210.01739*, 2022.

[8] 3GPP, "Universal mobile telecommunications system (umts); lte; architecture enhancements for control and user plane separation of epc nodes (3gpp ts 23.214 version 16.1.0 release 16)," 2020. [Online]. Available: https://www.etsi.org/deliver/etsi_ts/123200_123299/123214/16.01.00_60/ts_123214v160100p.pdf

[9] ——, "Security architecture and procedures for 5g system, (3gpp ts 33.501 version 16.3.0 release 16)," 2020. [Online]. Available: https://www.etsi.org/deliver/etsi_ts/133500_133599/133501/16.03.00_60/ts_133501v160300p.pdf

[10] R. Rais *et al.*, *Zero Trust Networks, 2nd Edition*. O'Reilly Media, Inc., 2023.

[11] J. Kindervag *et al.*, "No more chewy centers: Introducing the zero trust model of information security," Forrester Research, 2010.

[12] G. Schaefer and M. Rossberg, *Security in Fixed and Wireless Networks*. Wiley, 2016.

[13] V. Stafford, "Zero trust architecture," *NIST special publication*, vol. 800, p. 207, 2020.

[14] 3GPP, "5g; 5g system; network exposure function southbound services; stage 3 (3gpp ts 29.591 version 16.6.0 release 16)," 2021. [Online]. Available: https://www.etsi.org/deliver/etsi_ts/129500_129599/129591/16.06.00_60/ts_129591v160600p.pdf

[15] ——, "5g; 5g system; public land mobile network (plmn) interconnection; stage 3 (3gpp ts 29.573 version 16.3.0 release 16)," 2020. [Online]. Available: https://www.etsi.org/deliver/etsi_ts/129500_129599/129573/16.03.00_60/ts_129573v160300p.pdf