TRAFICOM

Finnish Transport and Communications Agency
National Cyber Security Centre

# 5G Security Architecture

4/2022

# Sisällys

Introduction ..... 4
About This Guideline ..... 5
5G Technology Concepts ..... 6
Service-Based Architecture ..... 9
Security Aspects ..... 11
Mapping to the EU Toolbox Risks ..... 13
Cloud Concepts and Security ..... 14
Cloud Architecture ..... 15
Cloud Security Management ..... 18
Cloud Workload Concepts ..... 20
Cloud Workload Protection ..... 26
MEC Security ..... 29
Network Function Security Concepts ..... 30
NEF Considerations ..... 31
SCP Considerations ..... 32
SEPP Considerations ..... 33
Integrations ..... 35
Data Transmission Security Considerations ..... 36
Base Station Physical Threats ..... 37

Date of publication
28.2.2022

**Title of publication**
5G Security Architecture

**Author(s)**
National Cyber Security Centre Finland

**Commissioned by, date**
National Cyber Security Centre Finland under The Transport and Communications Agency Traficom
February 28, 2022

**Publication series and number**
**Traficom Research Reports**
4/2022

ISBN 978-952-311-780-8

**Keywords**
5G, Cyber Security, Security Architecture

**Abstract**

The current 5G standardisation focuses on the security of the protocols and application interfaces between 5G functions. However, there is only little publicly available information on securely owning and operating a 5G network, including the best possible way of protecting the underlying ICT platforms that run the 5G network.

5G network architectures and use cases come in various forms. Like many applications, mobile networks have also moved towards a software-based architecture, allowing a high scale of decentralisation. It means they can be protected with the same building blocks as many other microservice-oriented environments.

The Traficom 5G Security Architecture provides high-level guidelines and checklists to support building and operating a 5G network. It aims to help network owners in designing, maintaining and auditing a 5G network that is based on efficient and effective security mechanisms.

The Traficom 5G Security Architecture complements the current 5G standards and mainly focuses on a cloud- based 5G network implementation. By following the principles defined in the Traficom 5G Security Architecture and utilising the existing tools available in the modern cloud security technologies, the network owners can efficiently maintain the security and trust in the services their networks provide.

| **Contact person** | **Language** | **Confidence status** | **Pages, total** |
|---|---|---|---|
| Teemu Juujärvi | English | Public | 37 |

**Distributed by**
Transport and Communications Agency,
Cyber Security Centre Finland

**Published by**
Transport and Communications Agency,
Cyber Security Centre
Finland

# Introduction

One of the most pressing questions about the security of and trust in our digital infrastructure in the 2020s is the implementation of the new 5G standalone communications infrastructures. This technology is expected to become the ubiquitous data transmissions and services backbone for not only much of our interpersonal communications and e-commerce, but also for connected devices, machines, sensors, logistics and even vehicles. It is paramount that these use cases be built on foundations that can keep our personal data private, businesses secure, and devices safe.

This document aims to present the important concepts that network operators and other practitioners involved in system architecture and networking design, technical implementation, and monitoring should be aware of when it comes to the security of the 5G infrastructures. By recognising these concepts and applying this knowledge in their work, system architects, cloud and system engineers, developers and security operations teams will be better equipped to design, implement and maintain secure infrastructure operations.

## SCOPE AND TARGET AUDIENCE

The guideline focuses on 5G architecture security concepts and especially the cloud architecture and software aspects. As will be discussed later in this document, the standardised network functions are essentially modeled as applications with well-defined APIs. The document focuses on those security aspects that are new when it comes to 5G infrastructure.

The intended audience for this guideline are operators and other stakeholders familiar with the previous generation's telecom infrastructures and networking that are now planning to implement 5G standalone infrastructures.

The guideline does not contain original research or propose novel findings. Instead, this guideline is based on currently known practical security challenges and solutions in 5G, cloud computing, and software. Many critical decisions about security happen during architecting, when making technology choices and when planning integrations.

This document does not aim to cover all the security and privacy aspects of 3GPP specifications. Instead, the guidelines focus on bridging the gap between specifications and security theory with system and security operations practices.

At the time of writing this document, the 5G specifications and solutions offered by vendors are being developed at a rapid pace. We will periodically revise and amend this guideline as new innovations and developments become known.

## RISK-BASED APPROACH

In this guideline we are not suggesting a security baseline or a collection of controls as the best practice. Instead, we want to inform the reader of the different functionality aspects they should be aware of when considering security risks and appropriate procurement requirements, technology options and operational procedures to support their security objectives.

No two 5G implementations are going to be the same. The cyber threat environment and associated risks will vary based on geographical aspects, supported use cases, risk profile of the users, network architecture and technology choices, and many more aspects. We encourage all readers to perform their own assessments and plan for cyber security. This guideline can, at best, offer input for that process.

**You will find hint boxes like this one throughout the guideline.**
The advice contained is intended for operators and architects planning 5G standalone implementations as input for their risk-assessment and threat modeling work.

The hints are formulated as questions that we suggest the practitioners consider when setting security requirements for one's own network implementation.

# About This Guideline

The first version of this guideline was created in late 2020 and early 2021. The guideline was commissioned by The Finnish National Cyber Security Centre NCSC-FI under The Finnish Transport and Communications Agency Traficom with the intention of creating an open and continuously developing guideline about 5G best practices for those about to implement the first standalone networks and services. The document is not part of any legislation or a mandatory security standard. It is a collection of best practices and advice that is voluntary for anyone to apply.

To solicit the content and views for the guideline, a series of interviews was organised with operators and vendors invited by Traficom. They and other stakeholders contributed also by commenting on a draft deliverable before the release of the first version in June 2021.

The work to create this document was facilitated by Marko Buuri and Tuomo Makkonen from the cyber security professional services company Fraktal, who also wrote the first version of the document.

# 5G Technology Concepts

5G is the 5th generation of mobile communications technologies that will power digital services throughout the 2020s. 5G network implementations have already started, and the first mobile devices have become available.

In 5G, there are two types of implementations: standalone and non-standalone. In non-standalone implementations, previous-generation components are used in the core, whereas standalone implementations use only 5G components. Non-standalone implementations are seen as transitional, so this document will mostly focus on 5G standalone. At the time of writing this guideline, 5G standalone implementations are not commonplace in production. The reader is reminded that these topics and the guideline overall are forward-looking; as such, best practices expected to be commonplace in production at the time of writing will not be discussed.

**Consider these questions when planning a 5G infrastructure**

- How to gain assurance that the slicing technology is able to keep the logical service experiences isolated from unintentional service quality effects or intentional disturbance from one or more of the slices?
- Can private services with use cases that are sensitive from a confidentiality or service-level point of view be offered using the same technology stack as public services?

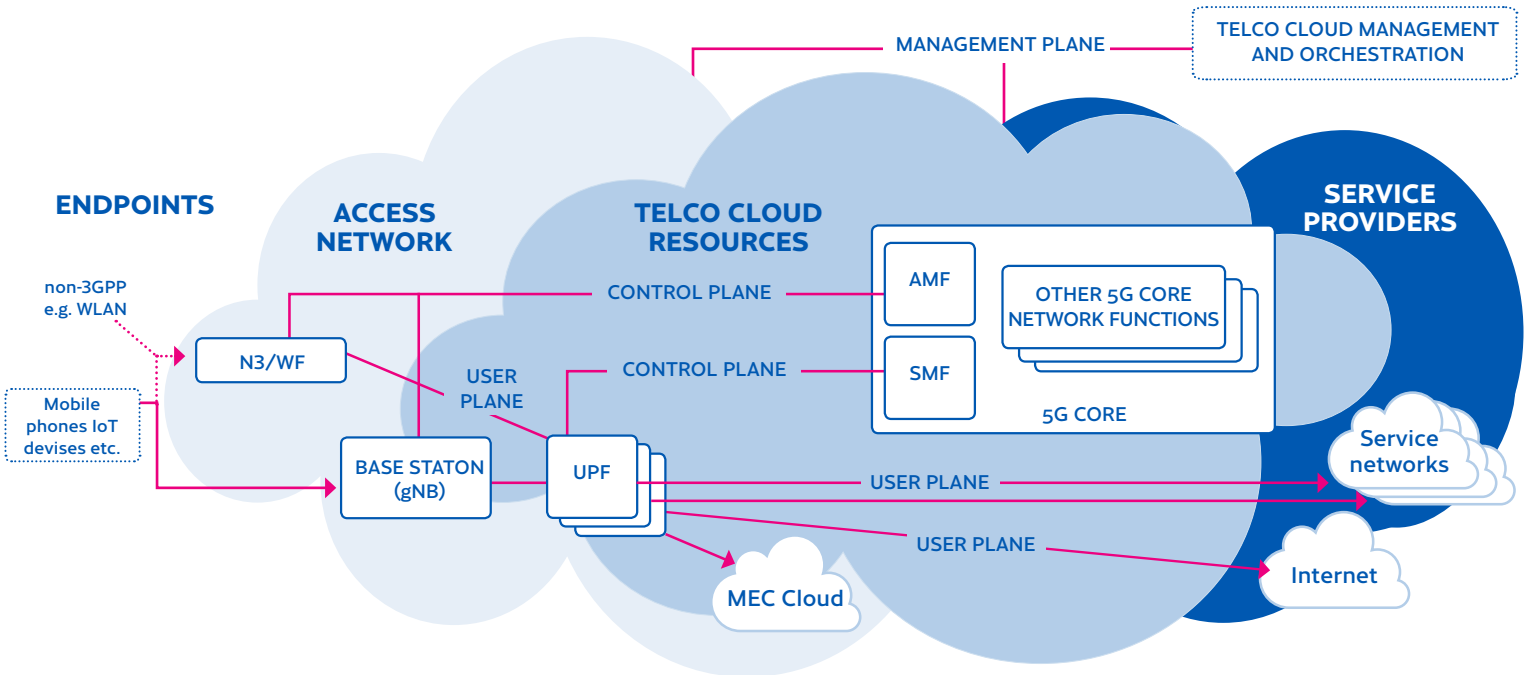



*Figure 1: Overview of a standalone 5G architecture.*

## TECHNICAL ARCHITECTURE

From an architecture and security point of view, 5G brings many changes. The following is a high-level diagram of the end-to-end architecture.

Key system elements in the architecture diagram are:

- **Endpoints** or **user equipment** are the phones, IoT devices and other equipment connected wirelessly to the 5G service. This generation brings new protocols that provide improved benefits in communications security and privacy in wireless channels compared to previous generations. In this guideline we will avoid focusing on the security aspects of the wireless channel.
- **Access network** is the part of the architecture where base stations and the access networking equipment and connectivity reside. We will discuss radio access network security later.
- **Telco cloud** are the data centre resources set up to provide the software-defined cloud computing, networking, and storage services needed to install 5G core network functions and other software required to run and manage mobile network operations. We will discuss the telco cloud and its security aspects later.
- In the telco cloud **5G core network functions** such as AMF and SMF have been provisioned as cloud workloads. 5G network functions are software deployed using cloud automation and orchestration mechanisms. We will discuss cloud workload concepts and network function concepts later.
- Multi-access Edge Computing or **MEC cloud** is an architecture add-on expected to be part of many 5G implementations in the future. Its purpose is to provide a possibility for service and application providers to bring their cloud workloads as close as possible to the mobile users. Doing so would minimise latencies and offer the best access speeds possible. MEC is currently developed in ETSI and is expected to be part of the 3GPP Release 17 specifications. We will discuss MEC concepts later in the guideline.

In 5G there is a new and clearer separation of different communications planes:

- **Control plane** refers to the signaling plane that implements the functionality needed to verify end points, allow them access to the network and activate the requisite service access.
- **User plane** refers to the plane on which service and application traffic is transmitted. This could be e.g. traffic to and from any website on the internet.
- **Management plane** refers to signaling used to deploy, configure, monitor and otherwise manage the infrastructure components, such as devices, cloud workloads and 5G network functions. The management plane is isolated from the other communication planes and critical from a security point of view.

These communication planes must be kept logically isolated. Any possibility for endpoints to access the control plane might have critical security implications, as that position might be used to leverage attacks against the base stations and core network.

## NON-STANDALONE AND STANDALONE IMPLEMENTATIONS

The non-standalone (NSA) 5G refers to implementations that have been built by taking advantage of the previous-generation network core. In this case the core network is called Evolved Packet Core (EPC). The network is amended with faster and more reliable new 5G radio, called 5G NR. The speeds and services that can be offered with this technology exceed 4G LTE offerings.

Recently, 5G services based on the non-standalone approach have been made globally available via many operators. While these networks offer greater speeds and efficiency in data communications for mobile users, far greater benefits and use cases await in the near future. Those use cases will be enabled when new 5G networks are built with standalone 5G network cores. The standalone (SA) 5G refers to implementations with the new 5G Core (5GC). This core option will provide the full benefits and range of 5G services such as ultra-low latency mode and network slicing. Network functions in the 5G Core are specified with telco cloud platforms in mind, i.e. the technologies, their functionality and the services architecture are built around cloud-native thinking.

At the time of writing, 5G services based on standalone cores are not being commonly produced. This is the target mobile architecture for the 2020s and beyond, however, and this guideline is exclusively focused on building security knowledge of the security aspects of standalone core implementation and operations.

## 5G SERVICE MODES

5G networks are intended to support a vast range of service use cases from fixed broadband to low-energy IoT metering. These use cases require a different type of signaling at the radio channel and functionality from the core services. Standalone 5G implementations can support three types of service modes:

- Enhanced Mobile Broadband (eMBB) for high -bandwidth use cases,
- Massive Machine-Type Communications (mMTC) for low-power, low-bandwidth, high-density IoT use cases, and
- Ultra-Reliable and Low-Latency Communications (URLLC) for low-latency use cases.

Examples of mMTC application areas include building automation, logistics, and smart grid. Examples of URLLC application areas include factory automation, robotics, and autonomous driving.

Supporting these service modes may have security implications for the network. These considerations are currently outside the scope of this guideline.

## NETWORK SLICING

Network slicing is a new concept in 5G that enables the delivery of logically isolated end-to-end services using shared infrastructure capabilities, such as physical radio access and cloud platform resources. In order for 5G network slicing to work, the underlying cloud computing, storage, and network resources must be able to keep the logical virtualised service layers apart. This is one of the key concerns touched on throughout this guideline.

It should be noted that full isolation can only be expected to be achieved when operating within a single operator domain.

# Service-Based Architecture

5G core system architecture is based on concepts of independent services. In that respect it follows the cloud-native architecture principle where systems are built on loosely coupled services and integrated using well-defined APIs.

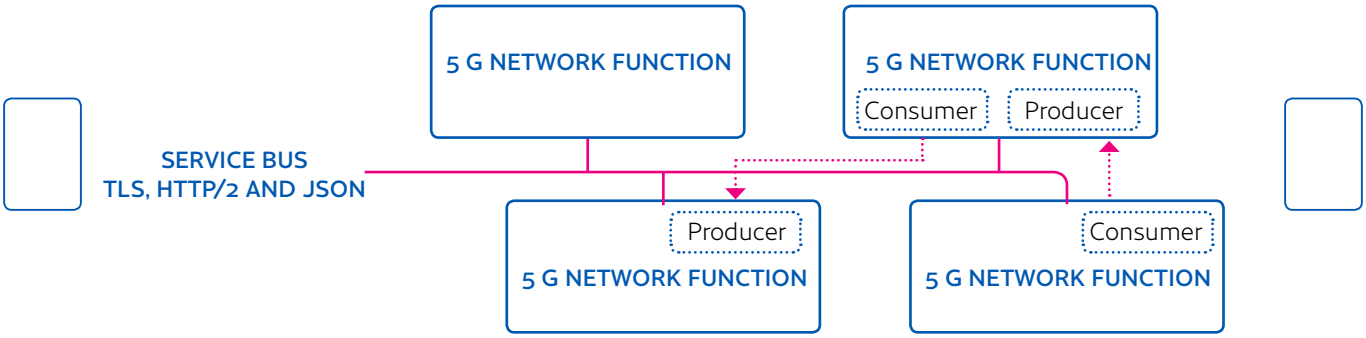



*Figure 2: 5G service-based architecture overview.*

In 5G, core network functions communicate over a service bus which can be e.g. Ethernet-based in a telco cloud implementation.

Network functions provide the essential network functionality for 5G and their functional and protocol specifications are developed by 3GPP. 5G follows the service-based architecture principle where network functions produce and consume services that can be accessed by other network functions. This is also referred to as the producer-consumer architecture pattern. Network functions should be understood as applications that host service APIs and access service APIs provided by other network functions. All this happens with the use of HTTP/2 and JSON serialisation, which are both common internet technologies.

```
# This is an example to illustrate JSON notation

[
    {
        "id": "9816654325",
        "subscriber": "97234987234",
        "service_endpoint": "https://telcocloud.local/v1/user"
    },
    {
        "id": "9816654331",
        "subscriber": "23468797923",
        "service_endpoint": "https://telcocloud.local/v1/user"
    },
]
```

## SERVICE DISCOVERY AND ACCESS AUTHORIZATION

As network functions are dynamic and occasionally short-lived software processes, every 5G core needs a discovery service to enable functions to communicate with each other. In other words, services need a well-defined service to learn about which URLs and IP addresses to communicate to within the telco cloud. A network repository function, or NRF, maintains a repository of functions in the environment. As new network functions are initialised, they will register their information, such as network address, authorisation rules and available services, to the NRF. To discover if there are other essential services that can be found, the functions can use the service discovery functionality provided by the NRF.

An NRF provides security-critical functionality as it contains an OAuth 2.0 authorisation server functionality. The purpose of that server is to issue access tokens for service consumers when they match the allowed rules submitted to the NRF by the service producers. In other words, a service consumer will have to present a valid authorisation token each time it wants to consume a service from a producer.

## LIST OF 5G NETWORK FUNCTIONS

The following 5G core network functions defined for 3GPP for standalone networks are discussed in this guideline. Each function provides a collection of services and consumes services from other network functions.

An interested reader should refer to 3GPP documentation for a full list of 5G services not covered in this guideline.

| Network function | Description |
|---|---|
| **AMF** | Access and Mobility Management Function provides registration, authentication and authorisation services for connected user equipment (UE) such as phones and other mobile devices. |
| **NEF** | Network Exposure Function provides controlled interfaces for untrusted third-party application functions (AFs) to access 5G network events and information. |
| **SCP** | Service Communication Proxy is an optional function that can be used to provide routing, load balancing, visibility and security functionality for service messaging between network functions. |
| **SMF** | Session Management Function provides services for managing the session lifecycle for connected UEs, including e.g. IP allocation services and DHCP functionality. |
| **UDR** | Unified Data Repository is a database for storing subscription data. |
| **UPF** | User Plane Function provides packet routing services for UEs, including e.g. QoS prioritisation and packet inspection functionality. UPF is a gateway to a data network (DN) such as the internet or another service network accessible by the 5G endpoint. |

# Security Aspects

5G specifications introduce many improvements to the security and privacy of the end users. These include e.g. new requirements for encryption and protection of subscriber identities. A 5G network will always encrypt permanent subscriber identifiers when transmitted over a radio interface. In roaming situations the visited network will only have access to a permanent identifier after a successful authentication. This mitigates false base station or "IMSI catcher" attacks used to track mobile device locations.

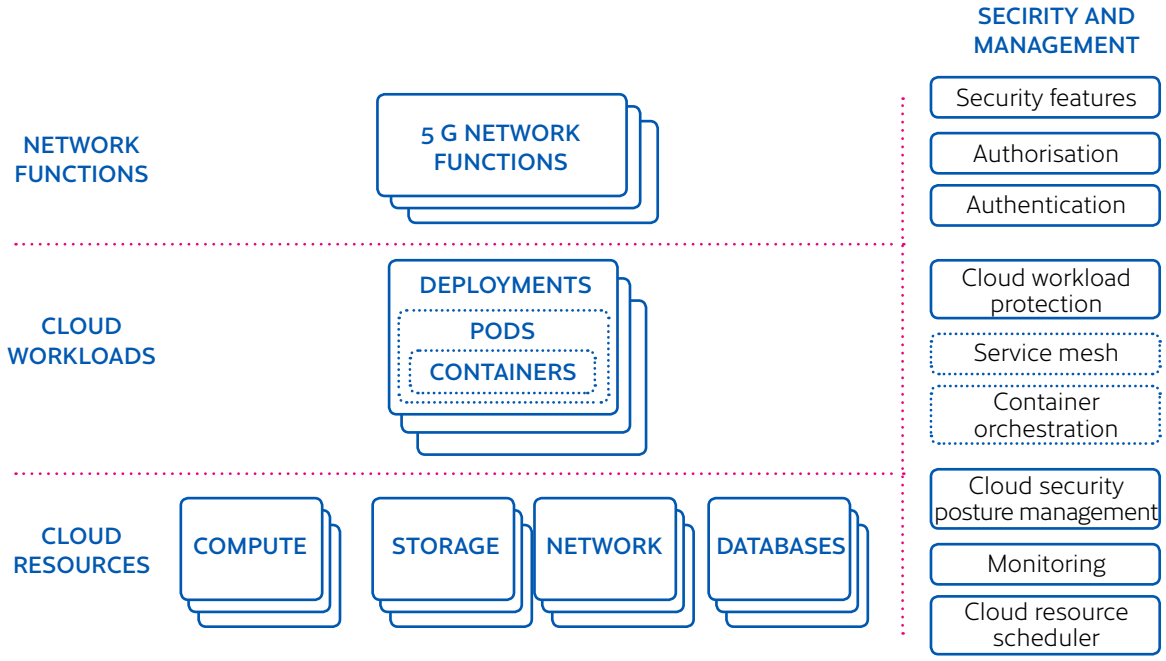



*Figure 3: Architecture levels and examples of their respective security and management functions*

This guideline discusses the 5G security architecture from the following viewpoints.

## SECURITY OF THE TELCO CLOUD

As 5G network functions are implemented as software workloads, operating a stable and secure telco cloud environment is a critical foundation for any 5G operation. This means that the cloud management and orchestration automation should come with features that support managing the security posture of the cloud resources.

## SECURITY OF THE CLOUD WORKLOADS

The first section of this guideline discusses telco cloud security. As 5G network functions are implemented as software workloads, operating a stable and secure telco cloud environment is a critical foundation for any 5G operation. This means that the cloud management and orchestration automation should come with features that support managing the security posture of the cloud resources.

In this version of the guideline we focus solely on private cloud aspects. Multi-tenancy considerations that typically come from public clouds are not currently within the scope of this guideline.

In the Telco Cloud Security section, we will discuss telco cloud concepts and security posture management of the telco cloud environments.

## SECURITY OF THE 5G NETWORK FUNCTIONS

Network functions are the logical building blocks of the 5G networks. The 3GPP specifies key security functionalities that all the network functions share, e.g. authenticating and encrypting communications. Some of the network functions have specific security-supporting responsibilities.

In the Network Function Security section, we will discuss network function key concepts as well as NEF, SCP and SEPP network functions. Additionally, we will discuss the possibility of integrating trusted applications with network functions using ways that are outside the scope of 3GPP specifications.

## SECURITY OF THE RADIO ACCESS AND TRANSPORT NETWORK

The radio access network is the part of the mobile network architecture that contains the base stations and the networking to connect them with the core functions and data services. 5G brings lots of improvements to the security and privacy aspects of the wireless interface and endpoints.

In the RAN and Transport Security chapter, we will discuss transport data protection as well as the physical security aspects of base stations.

# Mapping the EU Toolbox Risks

The European Commission has created a 5G toolbox to promote the secure adoption of 5G technologies with the European Union. A key publication is the document on 5G risk mitigating measures (CG Publication 01/2020).

In this document, we provide guidance to operators and other stakeholders on how to address some of the risks identified in the toolbox. The chapters in this guideline discuss the technical measures and supporting actions as follows:

| # | Title | Applicable chapters |
|---|---|---|
| TM01 | Ensuring the application of baseline security requirements | Telco cloud security<br>Workload security<br>Network function security<br>RAN and transport security |
| TM02 | Ensuring and evaluating the implementation of security measures in existing 5G standards | Network function security<br>RAN and transport security |
| TM03 | Ensuring strict access controls | Telco cloud security<br>Workload security<br>Network function security<br>RAN and transport security |
| TM04 | Increasing the security of virtualised network functions | Telco cloud security<br>Workload security |
| TM05 | Ensuring secure 5G network management, operation and monitoring | Telco cloud security<br>Workload security |
| TM06 | Reinforcing physical security | RAN and transport security |
| TM07 | Reinforcing software integrity, update, and patch management | Telco cloud security<br>Workload security<br>Network function security |
| TM08 | Raising the security standards in suppliers' processes through procurement conditions | This entire guideline contains appropriate recommendations. |
| TM11 | Reinforcing resilience and continuity plans | Telco cloud security<br>Workload security<br>Network function security<br>RAN and transport security |
| SA01 | Reviewing or developing guidelines and best practices on network security | Telco cloud security<br>Workload security<br>Network function security<br>RAN and transport security |

Nevertheless, it must be noted that following this guideline will not guarantee that any risks are adequately mitigated. Operators should conduct thorough security risk assessments and consider their operational specifics, vendors, selected technologies and supported 5G use cases while doing so. Every 5G environment and deployment will be unique, which means that the risk assessment and mitigation plans should likewise be unique.

# Cloud Concepts and Security

5G network functions are software-based services and will be deployed as workloads in data centre environments. We refer to these environments as telco cloud when they are used to deploy and manage cloud-native 5G workloads and have been built to match the capacity, resilience, and visibility demands of operators.

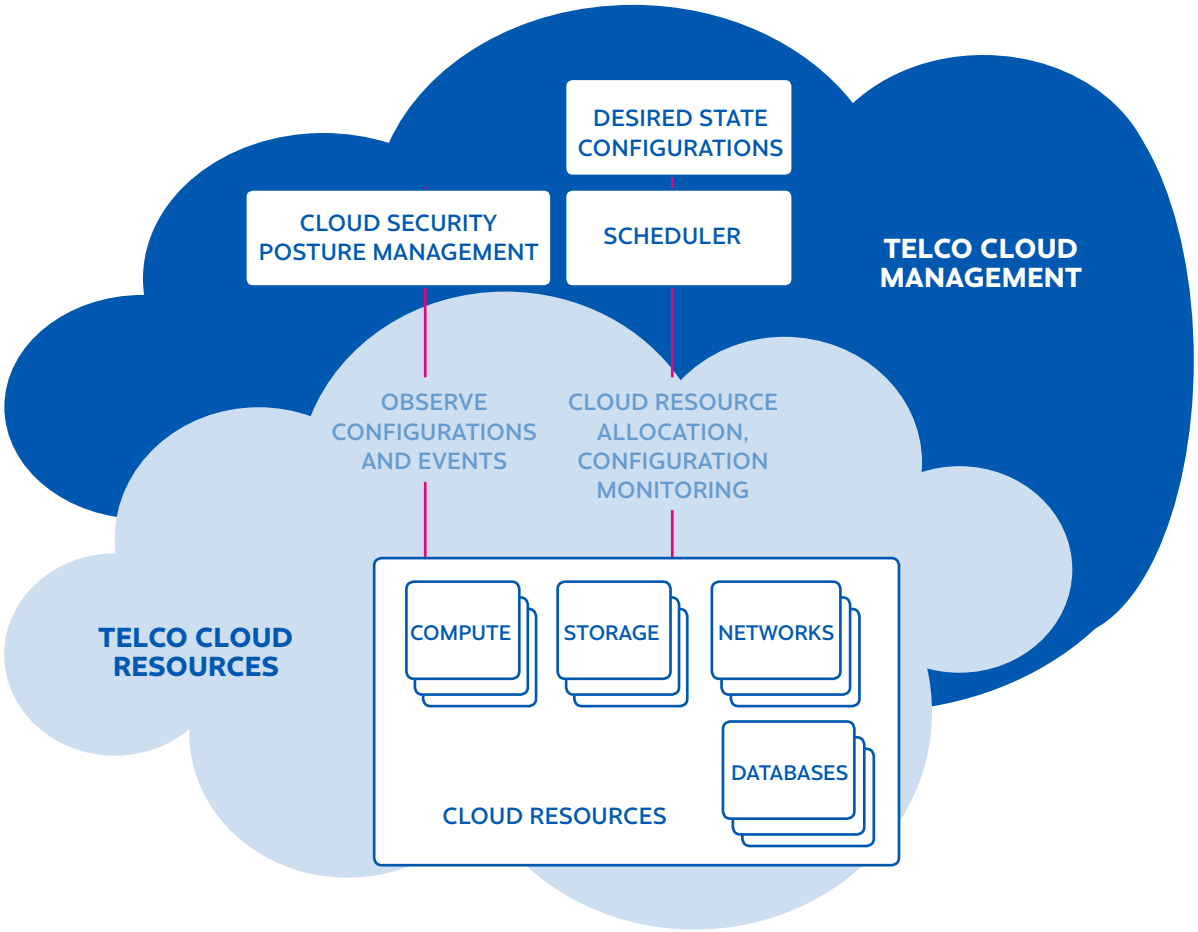



*Figure 4: High-level architecture for cloud resource scheduling and security management.*

5G features rely heavily on modern cloud capabilities, namely virtualisation and scalability of computing, storage, and network resources. The network functions leverage these cloud building blocks to implement the standardised 5G functionality as well as operator-specific functionality.

Traditionally, telco clouds have been understood as being private clouds. That means that the data centres and their capacity has been built and operated by the operator for their own purposes. The cloud-native approach in 5G has led to the current situation, where there are now public cloud offerings for network functions. This means that service providers offer managed 5G network function capabilities hosted from their own cloud platforms. A hybrid cloud approach would take advantage of both ways of producing the services.

Cloud resources are the foundational data centre service building blocks that enable cloud workloads to run on those platforms. Common examples of cloud resources:

- Computing resources include e.g. the possibility to run virtual machines, run containers without maintaining virtual machines, and a possibility to run code without specifying servers (serverless)
- Networking resources enable e.g. the creation of virtual networks and routing
- Block storage resources provide persistent unstructured file storage to compute workloads that might be ephemeral
- Database resources provide persistent structured data storage for computing workloads.

In the cloud, these logical resources are deployed, maintained and monitored by management automation and orchestration systems.

# Cloud Architecture

What makes cloud computing different from a traditional data centre approach is that hardware is treated as bulk capacity on which cloud-based services are deployed using orchestration automation. Whereas previously servers were bought and installed with an intention of running certain applications on them for the duration of their lifetime, in cloud computing the applications can reside on and use the capacity of any available hardware. Often, applications are moved around between servers by using a form of automation called orchestration to optimise their performance and resource consumption.

## CLOUD-NATIVE CONCEPTS

In a cloud-native approach, applications are built so that they are packaged in collections of **containers** that can be easily moved around and deployed rapidly on a compatible host with a container runtime. These applications (cloud workloads) are ephemeral by nature. The number of workloads running on any given time is not based on static software installations. Instead, cloud orchestration automation will decide on the optimal combination of services needed at a given time based on a set of rules.

Cloud-native architectures are **service-based and centred around APIs**. One container typically contains a service that collaborates with other services on other containers using APIs that the services expose. These services are referred to as **microservices**. Together, microservices offer the full functionality of the application. The service interfaces are well-defined and work in a way that frees the consumer from having to know anything about how the functionality has been implemented by the producer.

Additionally, cloud-native architectures are **loosely coupled**, so that when a service is in demand it can be automatically scaled up by moving it on a more powerful hardware or scaled out by increasing the number of copies running in parallel. Because microservices are decoupled from physical resources such as disks, or any specific knowledge about the computing platform they happen to be running on, they can be moved around without any interruption to the application they are a part of. Service catalogue and discovery automation will allow services to find each other in the cloud environment and direct the requests to the most suitable service producers.

## ZERO TRUST ARCHITECTURE

The zero trust architecture (ZTA) model has recently gained attraction as a solution to access control risk management in cloud architectures where resources and workloads are ephemeral and loosely coupled. Though zero trust is not part of 5G specifications, it is discussed here because some vendors are referring to the model in their own materials as something that can be achieved by using their offering.

While the ZT concept has been discussed for over a decade, there have not been many formal definitions for it. In 2020 NIST publicised their paper on the topic, and this chapter is based on their definitions. In ZTA, users access resources using well-defined APIs that

- use only secure protocols that provide messaging integrity and confidentiality,
- enable authentication of all access events, and
- perform dynamic risk assessment before authorising access.

What separates ZT from previous network security thinking is that the risks for all access events are dynamically evaluated at the time of access. In order to implement this functionality, two logical components are needed:

- A policy decision point (PDP) that contains a policy engine to make access decisions and a policy administrator to enforce those decisions.
- A policy enforcement point (PEP) that contains the functionality to enable, monitor and terminate access between the service consumer and producer. There are many technical ways to implement a PEP, such as agents, network equipment and API functionality. PEPs are managed by policy administrators.

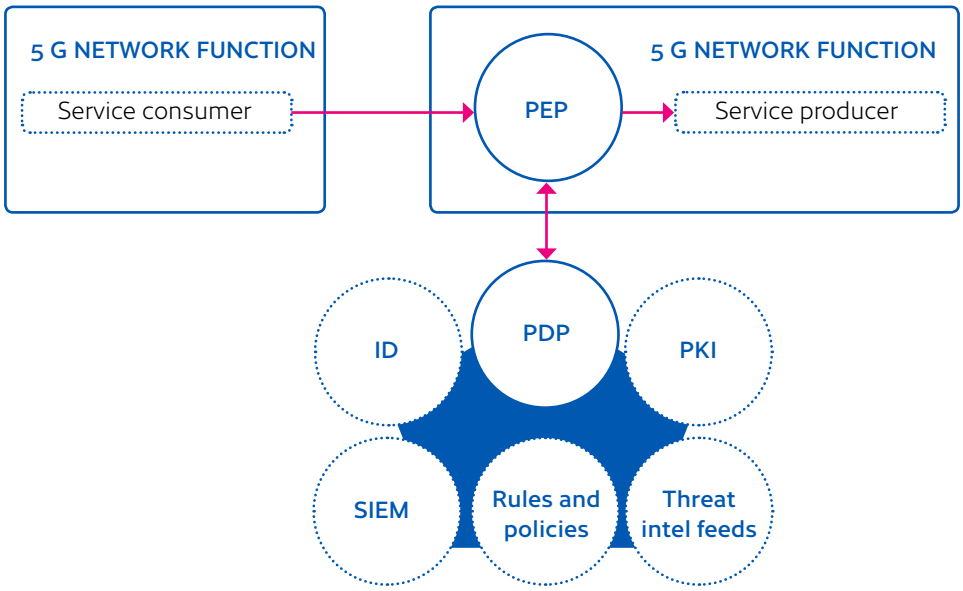



*Figure 5: A conceptual example of ZT components applied to 5G network functions.*

In traditional IT environments, a PEP component could be seen as something that is part of applications and APIs, or possibly as part of network equipment such as a firewall. Similarly, in a 5G context there are several ways to implement PEP functionality:

- Integrate PEP functionality directly to each 5G network function. Such an access component could be e.g. a sidecar proxy service as discussed more in this guideline in the context of cloud workloads and service mesh.
- Integrate PEP functionality to an SCP network function. SCP is a proxy between network functions, and it can include access authorisation and risk assessment logic.
- Additionally, SEPP network function could include a similar dynamic risk-based approach to security functionality.

There are several ways that PDP can and should source threat information to make relevant risk decisions. Those include:

- Access to identities and valid keys or tokens of resources
- Threat intelligence feeds that include attributes that indicate malicious usage
- Access rules and policies of acceptable or unacceptable usage patterns
- Security events and trend data from a monitoring system such as an SIEM.

## CLOUD MANAGEMENT

The processes of managing the cloud resources and orchestrating the collections of services are referred to together as Management and Orchestration (MANO). MANO processes are outside the scope of 3GPP specifications. Cloud management and orchestration processes take care of the following:

- Ensuring that the intended cloud resources are operational with the correct configurations.
- Monitoring performance of the physical hardware, virtualisation, and cloud resources.
- Rightsizing the use of resources to optimise spending.
- Identity and Access Management of cloud resources.

NFV MANO is an architectural framework defined by ETSI ISG for managing virtualised network functions. It aims to define specifications for the components and their functions in telco cloud architectures that host software-based cloud workloads.

The architecture comes with 3 main functional elements:

- NFV Orchestrator (NFVO) for managing the lifecycle of cloud workloads that provide networking services such as 5G network functions.
- VNF Manager (VNFM) for managing the lifecycle of software-defined, virtualised network resources in the telco cloud.
- Virtual Infrastructure Manager (VIM) for managing the telco cloud (hardware) resources, e.g. computing, storage, network, and database resources.

The ETSI working group responsible for NFV MANO undertakes work in 2-year phases. At the time of writing, the work on security aspects is currently active.

**Consider these questions when planning a 5G infrastructure:**

- How can we express the desired security state for the cloud environment so that the management automation is able to assess it?
- How should our tooling on cloud resource management support visibility into security events and ensure a proper security posture against attacks?
- Will there be multiple layers of security management tooling, or can we integrate 5G network function, cloud workload and cloud resource layers under one approach?

# Cloud Security Management

A foundational layer of 5G security is how the telco cloud resources are protected. These are the data centre resources that provide software-defined computing, networking, storage, and database services. The cloud resources must be constantly monitored and assessed for cyber security weaknesses. Most successful attacks on cloud services are the result of errors made in configurations and a lack of visibility of those mistakes.

Any manual or otherwise static approach in providing such visibility into security is incapable of keeping up with the changes. Special tooling is needed to manage overall security of the cloud resource configurations. The industry term Cloud Security Posture Management (CSPM) is defined by Gartner as "a continuous process of cloud security improvement and adaptation to reduce the likelihood of a successful attack."

Many telco cloud environments used to host 5G core services will implement CSPM capabilities to make sure their virtualised cloud hosts and other software-defined infrastructure services are running secure configurations and not exposed to attacks.

A CSPM solution automatically identifies and analyses cloud resources and their configurations and searches the data for possible security weaknesses or risks. This could happen by combining static rules and policies with more sophisticated risk engines, threat feeds and intrusion detection automation. A CSPM solution will typically source this information from the cloud management control plane and interfacing with the data centre equipment using their management APIs.

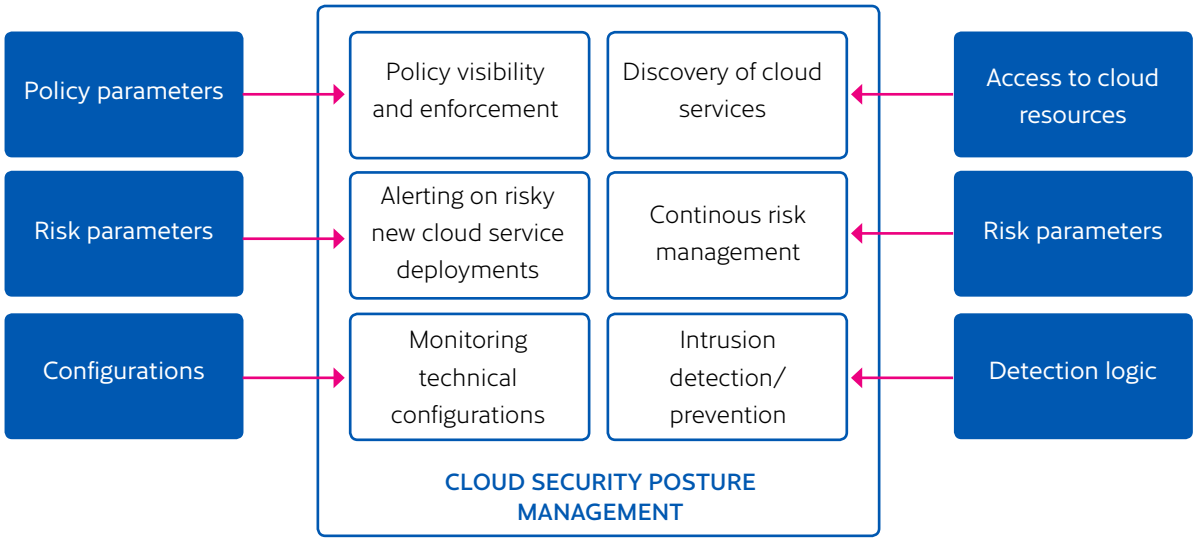



*Figure 6: CSPM solution is used to observe and maintain security configurations of cloud platform resources and monitor for security events.*

Many telco cloud environments used to host 5G core services will implement CSPM capabilities in some form or another. What features those capabilities will offer varies. Those could include e.g. the capability to define and enforce security policies across the telco cloud, and to produce reports and alerts on policy violations.

Examples of risky configurations:

- Deviations from security baselines set for the telco cloud environment
- Privileged access permissions set not according to security policies
- Unusual network configurations
- Telco cloud environment exposing inbound network connectivity from unexpected locations, such as the internet
- Occurrence of unauthenticated APIs
- Database or storage services exposed
- Occurrence of workloads not signed by approved vendors.

By its nature, Cloud Security Posture Management is an agentless approach that focuses on the resource configurations and APIs.

**Consider these questions when planning a 5G infrastructure:**

- What are the key security policies that need to be enforced in the telco cloud?
- What kind of security automation should there be to support policy enforcement and visibility with regard to the configurations being applied?
- Will there be cloud automation supporting the discovery of potentially unexpected workloads and services from the telco cloud environment?
- What would constitute risky deployments and configurations?

# Cloud Workload Concepts

Cloud workload refers to those applications and services that are being run on cloud infrastructure, for example software-based 5G network functions.

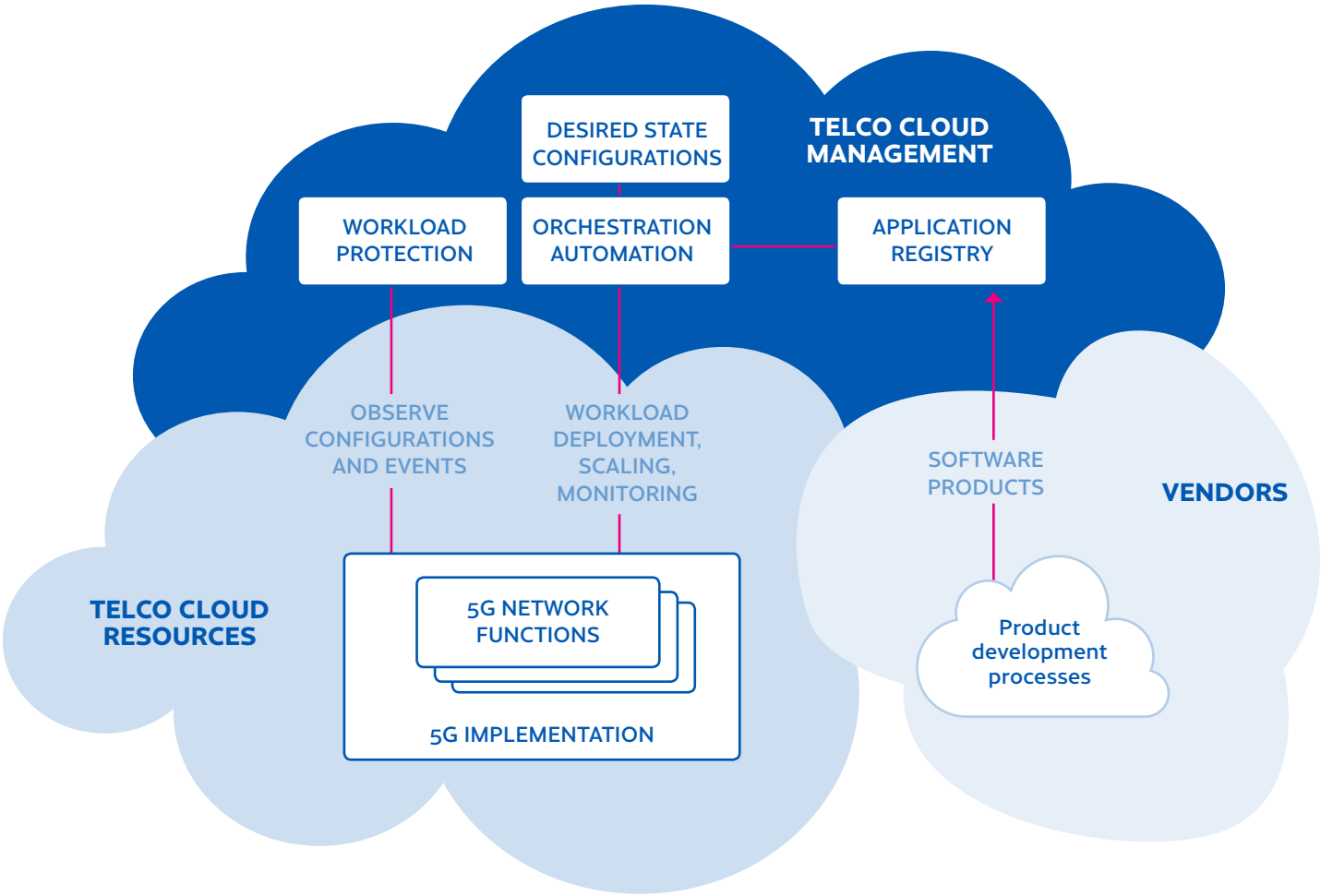



*Figure 7: High-level architecture for application delivery, deployment, and security.*

Solution vendors will continuously develop the features of their software solutions and release new versions. In modern DevOps architecture, the releases will be downloaded to the operator's application registry. The registry is a database of applications from which the operator will deploy them to test and production environments. In practice, the contents of the registry are specific to the operator's cloud implementation and could be VM images, containers and even native application binaries.

As 5G network functions are essentially applications, their configurations and vulnerability management are essential in the overall security posture of a 5G environment. Configurations, software components and behaviours that indicate a security weakness should be detected and responded to. Successful attacks on applications are often due to excessive exposure of settings or vulnerable services.

There are solutions for addressing these cloud workload security concerns. Such solutions are called cloud workload protection (CWP) platforms, as will be discussed later in this guideline.

**Consider these questions when planning a 5G infrastructure:**

- As microservices internal to network functions can be spread across multiple physical hosts, should the network communications between such components be authenticated and encrypted?
- What kind of isolation between microservices implementing different 5G network functions is needed? Is it sufficient to implement them in a single cluster with sufficient visibility, or is cluster-level separation necessary?
- What level of visibility and security controls do we need for the individual software components that reside inside our 5G network functions?

## CLOUD WORKLOAD TYPES

A workload in cloud environments is an overloaded term. In the broadest sense, it can be used to refer to any service or application instance that has been programmatically deployed on a cloud platform.

Software-based 5G network functions will be referred to as cloud workloads when deployed to a telco cloud. There are many ways to do this, from more traditional VM-based installations to container deployments and micro-service orchestration. These options come with trade-offs in performance, scalability and reliability, as well as operational differences in how the workloads are set up, monitored and secured.

As 5G network functions are essentially applications, their configurations and vulnerability management are essential in the overall security posture of a 5G environment. Insecure configurations, vulnerable software components and behaviour that indicate a security issue should be detected and responded to. Successful attacks on applications are often due to security misconfigurations or vulnerable services being exposed to attackers.

5G network function services can be implemented as applications on either virtual machines or containers. These options come with trade-offs in performance, scalability and reliability, as well as operational differences in how the workloads are set up, monitored and secured. Each approach comes with different advantages, and a core network implementation can be a hybrid.

5G specifications do not require a specific deployment model to be implemented, and vendors are likely to come up with varying approaches to implementing the workloads and their orchestration.

### VM-based Deployments

In VM-based deployment, the service is installed on an operating system such as Linux running in a virtualised environment. Each service will be installed on their own VM or, in case of multi-VM deployment, split across more than one VM.

Use of VMs provides an excellent level of isolation between the deployed workloads, although it is possible that hypervisor or processor-level vulnerabilities are discovered over time. Such weaknesses could be used to breach isolation boundaries for the purpose of manipulating or stealing data from other workloads on the same host hardware.

On the negative side, VMs are slower to deploy than containers and generally more wasteful about the capacity of the underlying host hardware, with multiple copies of operating systems running in parallel.

A VM-based deployment does not mean that container technology is not used at all. Rather, it implies that dedicated VMs are deployed for each application or a microservice. There are alternative deployment strategies:

- The application code can be a part of the VM image and deployed as a single package.
- The VM image contains a container runtime, and the workload orchestration deploys a containerised code once a VM is set up.

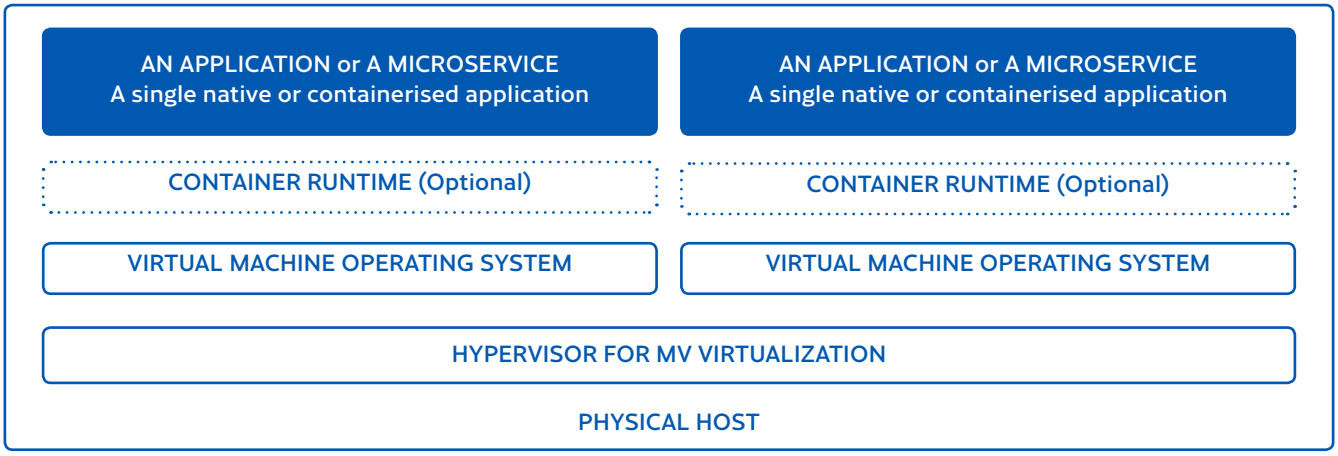



*Figure 8: In VM-based deployment there is one application or a microservice per virtual machine.*

## Container-based Deployments

Container-based deployments mean that the service or part of it is packaged in a software container that is then deployed and executed using a container engine such as Docker on a host operating system such as Linux. A key difference is that a single Linux system can be used to execute several container-based workloads concurrently. The container technology provides isolation between the workloads and the host system. The downside is that the isolation between containers does not equal hardware-level or VM-level isolation. The technology is best suited for situations where the containers on the same physical host share a security boundary.

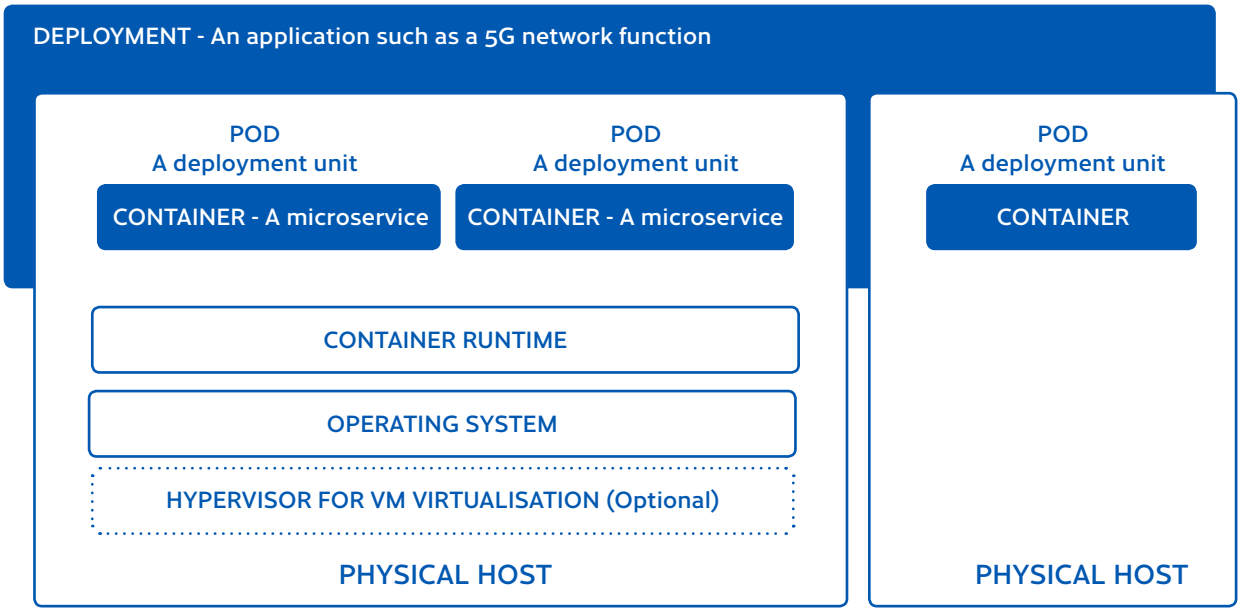



*Figure 9: In a container-based deployment an application's functionality is split across loosely coupled services.*

### Comparison of Features of VMs and Containers

The following table summarises the comparison between virtual machines and containers as deployment units.

| Featur | VM deployment | Container deploymen |
|---|---|---|
| **Deployment unit** | Image | Container |
| **Deployment host** | Hypervisor virtualisation layer running on hardware. | Container engine running on an operating system that could be a virtualised one. |
| **Deployment unit size** | Larger than a comparable container, as an image includes an operating system. | Smaller than a comparable image, as a container contains only the code for a single application or a microservice |
| **Starting of workloads** | Slower than containers, although optimised lightweight VMs may reach boot times close to containers. | Faster than equivalent VMs. |
| **Efficiency** | Multiple copies of the operating system running on a physical host. Typically wasteful of resources. | One operating system per host. Allows better optimisation of resources. |
| **Reliability** | A skilled workforce, tradition of good practices and stable tooling are available. Dependencies are easily managed as VM images contain everything in themselves. | Container-based solutions have reached maturity; however, container engine, kernel and orchestration tooling create dependencies that need to be managed. |
| **Workload isolation** | Hardware-level virtualisation. Provides best isolation for workloads. | OS-level isolation that is best suited for single-tenant use cases where all the workloads on a host share a security boundary. |

## WORKLOAD ORCHESTRATION

Workload orchestration refers to rules and processes for deploying, scheduling, scaling, maintaining and updating cloud workloads such as containers.

Container-based application typically follows the microservices architecture model. That means that the application is based on a collection of containers, with each of them providing a well-defined service. For the full application to work, all the containers need to be in place and operational.

Container orchestration tools such as Kubernetes automate the deployment, management, scaling and networking of container-based applications. Container orchestration comes with its own security concerns.

### Kubernetes

Kubernetes (k8s), an open-source project originally developed by Google, is currently the prominent technology for orchestrating containerised microservice workloads. Kubernetes supports multiple container runtimes; at the time of writing, the most common container runtime in public clouds is Docker.

Kubernetes **clusters** contain a collection of physical or virtual servers called **nodes**, one of which hosts the control plane components such as the workload orchestration automation and health monitoring. Each cluster then has several worker nodes. In Kubernetes a deployment unit is called **a pod**. Pods are considered ephemeral in the sense that they are disposable and non-unique instances of **containers**.

An application such as a 5G network function is modeled in Kubernetes as **a deployment**. A deployment is a desired state description of the operational aspects of the application deployment, such as how many pods should be deployed, which versions of containers should be running, how pods should be scaled in case there is increased load, and how to roll back a deployment in case an update goes wrong. The control plane automation in Kubernetes then works to ensure that this statement is fulfilled.

A key feature in Kubernetes is the continuous ability to monitor the health of the pods and containers. In case there are issues, the automation will deploy new pods on available hosts, enabling high availability of the microservices in case of hardware failures.

## Service Mesh

As containerised applications grow more complex in terms of the number of pods and containers, managing the internal communications between the pods becomes an issue. From a system operations point of view, it is important to have visibility, security through encryption and network traffic controls, and reliability in service-to-service communications within deployments.

Service mesh solutions provide a reusable way to address these issues without having to develop these features inside applications. Service mesh solutions

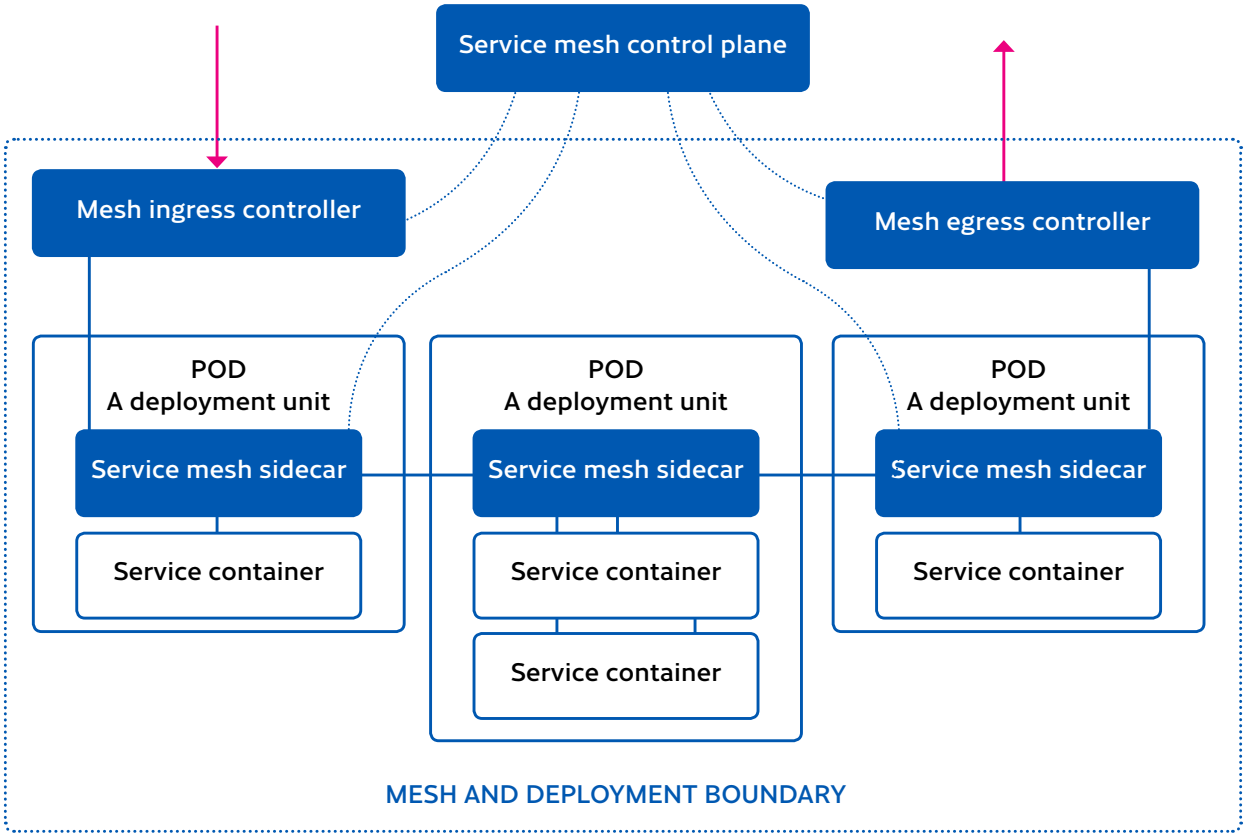



*Figure 10: A service mesh architecture implementing the sidecar pattern.*

**Consider these questions when planning a 5G infrastructure:**

- What are our requirements for detecting threats and malicious activity in the core environment workloads? For security monitoring, visibility to logs and network traffic inside the cluster is critical.
- What are the security-critical configurations for network functions (workloads), and how do we monitor that they are effectively in place?
- To what extent do we need to discover vulnerabilities in network functions (workloads) such as known vulnerable open-source components?
- How is the workload orchestration (e.g. Kubernetes) set up, and how is the access to the control plane protected?
- Do our vendors follow the NESAS scheme, and what other information is available on their secure development practices?

often follow **a sidecar pattern**, which means that a container will be attached with each pod to handle authentication, networking and logging functionality as needed. A sidecar pattern could be used to implement Zero Trust principles.

Containers within a pod share resources and can communicate with each other. Mesh sidecar handles communications between the pods in the same deployment. Typically ingress and egress network controllers are included to provide load balancing and firewall features.

# Cloud Workload Protection

Regardless of the types of workloads being used and how they are orchestrated, a security solution is needed to provide assurance that the software and services deployed in production are running with security configurations and do not contain threats.

Cloud Workload Protection (CWP) solutions are for monitoring, managing and protecting the security of cloud workloads. Use of such solutions would allow an operator to verify that the cloud workloads such as 5G network functions are running securely.

Many telco cloud environments used to host 5G core services will implement CWP capabilities to make sure they are not running 5G software or other cloud workloads that contain security vulnerabilities or malicious code.

A CWP solution automatically identifies and analyses cloud workloads and their configurations, and searches the data for possible security weaknesses or risks. This can be done by combining static rules and policies with more sophisticated risk engines, threat feeds, anti-malware engines and intrusion detection automation. A CWP solution will source this information by directly integrating with workloads using agents or scanning workloads externally. These approaches are compared later in this guideline.

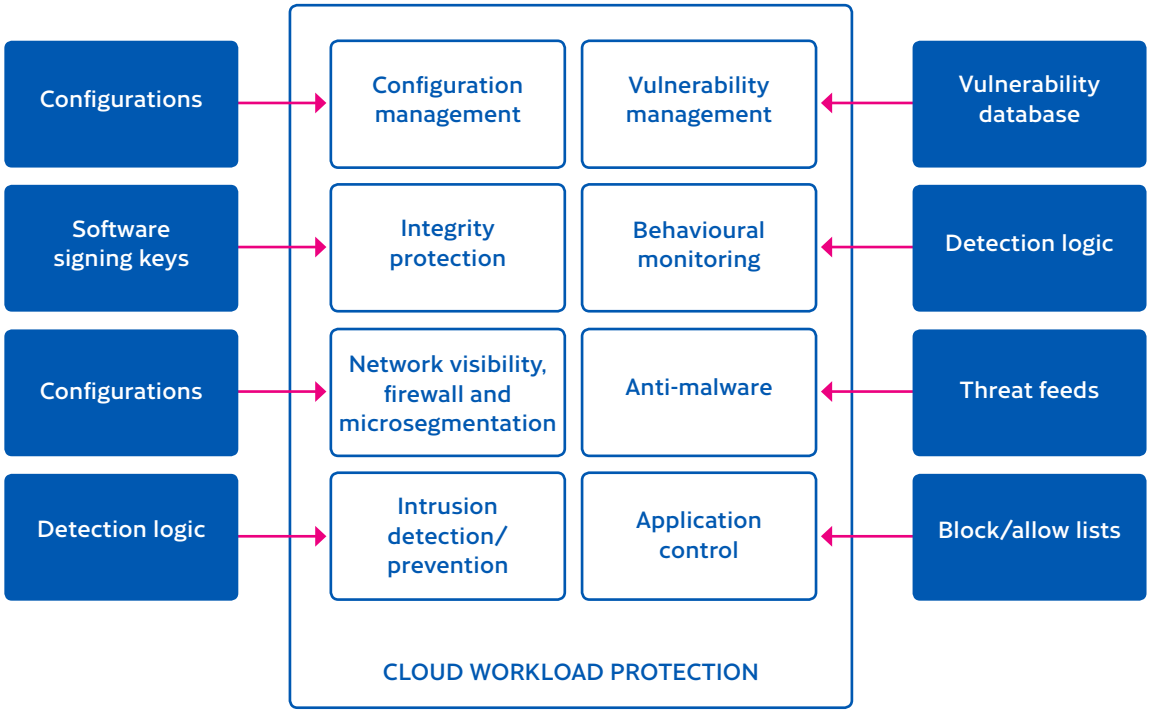



*Figure 11: CWP solutions are used to secure cloud applications and monitor for security events.*

Many 5G telco cloud environments will implement cloud workload protection capabilities in some form. Features of such solutions could include e.g.

- detecting insecure configurations of workloads,
- scanning for vulnerable software components in use,
- detecting known malware, and
- monitoring for anomalous behaviour patterns.

Not all workload protection solutions will offer all the aforementioned features. The feature set most suitable in any environment should be based on **a risk-assessment and consideration** by the operator.

Special consideration should be placed by operators on **the coverage of cloud protection solutions**. A 5G network might consist of software from multiple

different vendors and the 3GPP specifications will ensure they can interface securely. Still, an operator should have visibility into the security of all the workloads regardless of their vendor, software architecture, programming languages, container formats, microservice orchestration and service mesh, open source library versions and the multitude of other nuances that will exist in real implementations.

## TECHNOLOGY APPROACHES FOR CWP

There are several ways of implementing CWP functionality. There are two main approaches: those that rely on installed agents, and agentless solutions.

### Agent-based Approach

In an agent-based approach, the monitored workloads will have an agent component installed to perform the security functionality. This could include e.g. monitoring and exporting logs, processes and API events. The exact functionality of an agent will be dependent on the technology stack used, the workload application architecture, the feature requirements of the agent and other implementation-specific aspects. At the operating system level there is a long tradition of running management, monitoring and security agents on the hosts. An agent-based CWP solution could be much like an endpoint threat detection solution, capable of inspecting an operating system.

In container-based deployments, an agent can be e.g. a separate container included in deployed Kubernetes Pods. This method is called a sidecar pattern. Such a container has the same access to the network and storage as the other containers of that pod and can provide security monitoring functionality. This approach is similar to service mesh, and CWP functionality can very well be part of the same sidecar container.

A concern with agent-based solutions is that an agent, which is embedded in all workloads, is a tempting vector for an attacker to breach. An agent increases the amount of potentially vulnerable code active in the system, and a remotely accessible security vulnerability in an agent would be a critical risk for the 5G environment.

Additionally, a buggy agent implementation might hinder the operations of the 5G network functions by unexpectedly consuming computing resources or blocking service operations. Minding these potential issues adds complexity to the software lifecycle management.

### Agentless Approach

An agentless security solution provides protection for cloud workloads without security processes running in the workloads themselves. There are several approaches to such solutions:

- Scanning of application registry for known malware or unassigned applications
- Scanning of storage and database services for known malware
- Scanning of network services of workloads to discover unexpected services and detect vulnerabilities
- Analysis on configuration database and access policies to detect possibly risky attributes
- Automated log analysis to detect security events
- Network traffic analysis to detect attack patterns, e.g. when integrated to an SCP.

This type of solution integrates with the cloud workloads over remotely accessible interfaces. These could include access to storage resources, for example. Additionally, cloud workloads might contain special APIs to enable health and security monitoring without agent processes.

### Application Registry

An application registry (also called image registry, container registry) is the database of software workloads in a cloud environment. Operators are likely to have multiple registries for production and test software versions. The content of the registries is pulled from the vendors, and the purpose of the registry is to make them available for the workload orchestration of the operator. The orchestration configurations will decide if any of the available applications are deployed anywhere in the telco cloud. An application registry doesn't have to be just a content- agnostic storage location and it can contain a **security-supporting functionality**. At a minimum, the registry should be able to verify that the stored applications, such as containers of 5G network functions or VM images, are signed by authorised parties and that their integrity remains intact. Changes made to the contents of the registry should be protected with good access controls and traceability. Additionally, the registry might be able to scan the contents to detect known malicious code as part of the stored applications or for outdated open source libraries.

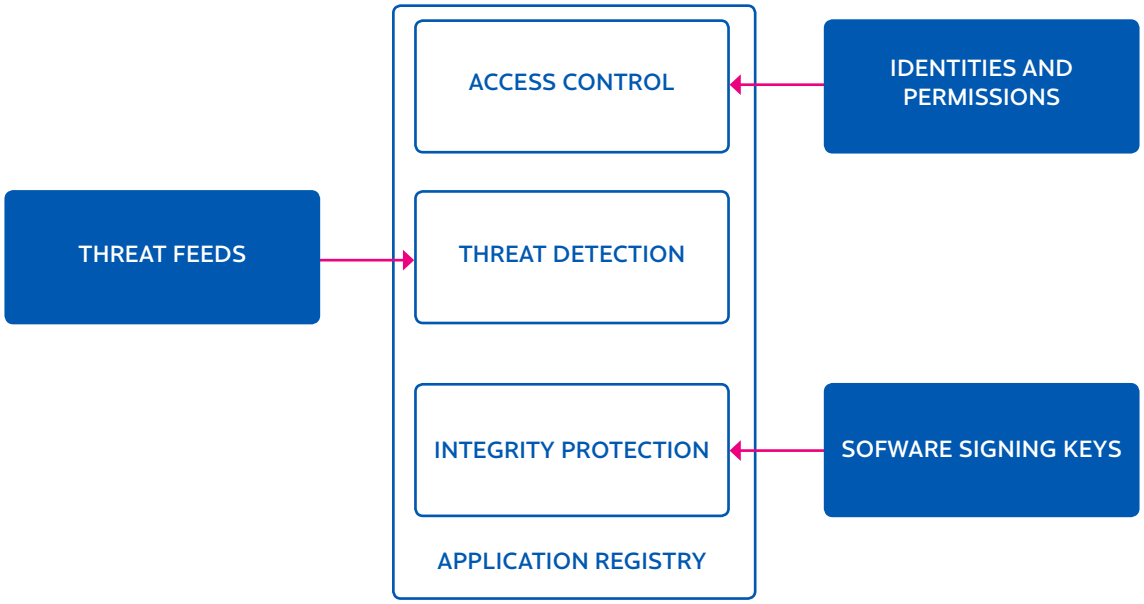



*Figure 12: Application registry can include security supporting functionality.*

## NESAS

GSMA has specified a NESAS assurance scheme for network equipment vendors to implement. Operators should look into how the prospective vendors follow the scheme and what other information is available on their secure development practices.

It should be noted that the objective of NESAS is to verify that the 5G products follow the 3GPP specifications, including security functionality, and it provides good assurances on this. There are, however, many areas of security outside the scope of 3GPP security specifications, such as operational security aspects or cloud and workload technology implementations. Therefore, assurance provided by NESAS alone is not proof of a secure 5G network implementation when in production. Operators should conduct risk assessments at both the operational and solution level, extending beyond the scope of NESAS assurances.

**Consider these questions when planning MEC capabilities:**

- What level of security assurance will be required from edge workload vendors?
- Are there ways to technically ensure security of the workloads?
- How well is a malicious or breached edge workload isolated from critical network functions?
- How should the edge computing platform be monitored so that an attack can be detected?

# MEC Security

Multi-access edge computing (MEC) is a cloud computing architecture concept developed by ETSI's MEC industry specification group (ISG). 3GPP is currently working to amend future specification releases with edge computing.

MEC describes a mechanism for deploying edge service workloads close to users to optimise content delivery and provide ultra-low latency for applicable use cases. In current specification drafts MEC applications are modeled as 5G Application Functions (AF). An AF can set a policy to the network that will influence User Plane Function (UPF) selection for a given user session and that way direct user traffic to an edge workload instead of e.g. a service that resides on the internet.

An operator might authorise edge workloads to access events and information about the operation of the 5G network and connected endpoints. ETSI MEC ISG has e.g. proposed an API with which an edge workload can query locations of endpoints.

From a cyber security point of view, there are concerns related to MEC workloads:

- Depending on the architecture approach taken by an operator, MEC workloads might be deployed on hosts close to critical components of the mobile network. If any of those third-party edge workloads were malicious, they could be well-positioned to attack other workloads and services in the telco cloud. Such malicious functionality could be deliberately placed in the workloads' code beforehand, or attackers could discover and make use of vulnerabilities found in production workloads remotely.
- An operator might choose to authorise APIs for edge workloads to get real-time access to 5G network events and information. Misuse of this information might be harmful in several ways, as it might contain business sensitive information and personal data about the networks' users.

Specifications about integrating MEC with 5G systems are under active development by ETSI and 3GPP. Implementations of MEC systems to support real-life use cases are not expected in the first wave of 5G standalone implementations.

We will monitor the developments in the MEC space and update this segment accordingly.

# Network Function Security Concepts

3GPP specifications define the baseline security requirements for communications, authentication and authorisation for signaling between the 5G network functions. These definitions exist for the purpose of interoperability between the functions which could be sourced from multiple vendors.

As discussed in cloud workload concepts, 5G network functions are software, and each function could come with its own software architecture, programming language, software libraries and container formats. They still need to interface securely, and that is why the specifications in 5G rely on open standards already common on the internet and software application use cases.

## SECURE COMMUNICATIONS BETWEEN 5G NETWORK FUNCTIONS

The specifications define the security baseline for communications between network functions. The elements are

- mutual authentication using TLS 1.2/1.3, and
- transport layer security using TLS 1.2/1.3.

The network functions are connected using a networking connection that provides TCP/IP connectivity between the core network functions. That way the functions can operate in accordance with the intended producer-consumer service model, accessing each other's services as needed.

It should be noted that the specifications only discuss integrity protection and encryption of messaging between network functions, hop-by-hop. Each function has access to unencrypted communications, which is necessary for the function to perform its intended operations. Furthermore, if a network function is implemented as a collection of microservices, the protection of the communications between those internal microservices is not within the scope of the specifications. This is important to understand, as in the telco cloud environment those microservices might be deployed across multiple physical hosts and data centres.

## ACCESS CONTROL TO 5G NETWORK FUNCTIONS

The specifications define that network functions use authorisation tokens based on OAuth 2.0.

In 5G standalone implementations the NRF function assumes the role of authorisation server. As network functions are executed, they will register themselves with the NRF and be added to the repository of services for discovery. At this time, the function will submit the authorisation rule. When another network function seeks to consume services, the NRF will authenticate the service using TLS and then authorise an access token considering the authorisation rules.

An SCP proxy can be used for indirect communications between network function service producers and consumers. In that case, the SCP acts on behalf of the consumer towards the service producer. The SCP is able to provide the consumer's OAuth access token while authenticating itself as SCP using TLS.

## MANAGING SECRETS

TLS and OAuth 2.0 functionality rely on cryptography functions and the way that the secrets needed by those functions are protected and managed throughout their lifecycle.

A public key infrastructure (PKI) is needed to enable the security functionality. How this system works in practice is outside the scope of these specifications and left for the vendors and operators to address. In practice, a form of automation should be in place for creating and issuing public/private keys and managing certificates that couple service identities with their public keys.

**Consider these questions when planning a 5G infrastructure:**

- How does the automation for managing secrets work to protect keys throughout their lifecycle?
- What level of monitoring and auditing capabilities are needed to verify secure operations?
- Is the automation compatible with the whole technology stack, especially in a multi-vendor environment?
- How can security operations respond to expected misuse of tokens or secrets?

# NEF Considerations

There are situations when 5G network function services and events need to be exposed to external applications in a controlled way. In 3GPP specifications, such APIs from the network towards external parties are called northbound interfaces.

In a 5G network, there is a network function called the Network Exposure Function (NEF) that is defined as acting as a gateway to support these use cases. As such, NEF works as an intelligent, service-aware "border gateway" that will enable external Application Functions (AF) to communicate with the 5GC network function in a secure manner over a RESTful API.

The NEF has the following security responsibilities:

- Determining that the AF is authorised to call the relevant NF using the OAuth2.0 authorisation framework.
- Integrity protection, replay protection and confidentiality protection for communication between the NEF and the AF.
- Mutual authentication between the NEF and the AF.
- Subscriber identity (SUPI) protection so that it will not be sent outside the mobile network domain.
- Hiding the details of the 5G core from the AF.

There are many ways such responsibilities may be implemented. For instance, there might be ways to extract information or otherwise utilise the APIs in ways that could not have been foreseen beforehand. At that point, vendors might take different approaches with improving NEF security functionality to e.g. detect known bad AFs or malicious access attempts.

## CAPIF

Independent from 5G, the 3GPP has specified a Common API Framework (CAPIF) for northbound APIs from core networks. CAPIF contains standardised mechanisms for publishing and managing exposed APIs.

CAPIF may be implemented in NEF, in which case the security features from that specification will follow as well. In CAPIF, TLS is used for protecting communications. For AF authentication PKI, TLS-PSK and OAuth 2.0 methods are supported.

**Consider these questions when planning a 5G infrastructure:**

- What are the use cases for external applications to access core network functions?
- How is the access scope for applications controlled?
- How are the identities and secrets such as keys for external applications managed?
- In what ways could the NEF be attacked by Application Functions, and what kind of functionality is needed from management and orchestration systems to respond to such events?

# SCP Considerations

Service communications proxy (SCP) is a network function that can be used in a 5GC for proxy service communication between other network functions. SCP is an optional part of any 5GC.

SCP does not replace the NRF. SCP may become a single point of entry for a collection of network functions once they have been registered with the NRF. SCPs are therefore well-suited as gateways to data centres, segregated networks, or clusters of microservices.

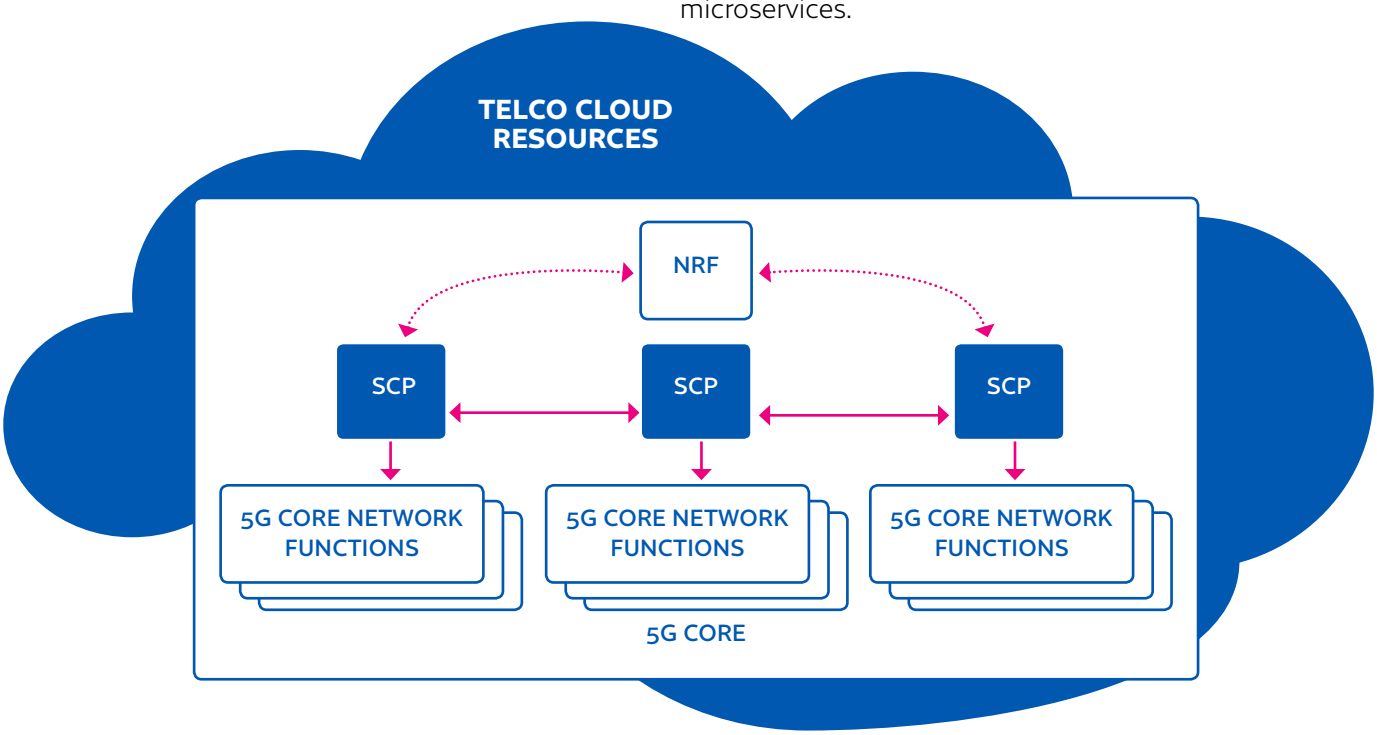



*Figure 13: An example of a distributed setup enabled by SCPs. Functions could be spread across datacenters.*

There are several reasons why using SCPs brings benefits:

- SCPs can be used to implement load balancing functionality in setups where there are several NFs providing the same services.
- SCPs can be used for network topology and routing optimisations. For example, an SBA might extend over multiple routed networks in a telco cloud.
- In such situations, very broad firewall rules would need to be used between the networks to allow service communications between NFs. SCPs can be used as gateways between networks to simplify the setup.
- SCPs can be used to monitor the health and performance of NFs. An SCP will be able to see how fast an NF will respond to service requests and the status codes.
- SCPs can be used to improve visibility regarding how the 5GC operates. In addition to performance metrics, they could implement logging of service requests.
- SCPs can include security functionality, e.g. enable detection of potentially malicious activity in the service requests. Furthermore, SCPs could incorporate a WAF-like firewall and traffic inspection functionality by inspecting service messaging for known patterns of attacks.

While use of SCPs is optional in 5GC, operations might find their possibilities useful in managing the overall security posture.

**Consider these questions when planning a 5G infrastructure:**

- How can we monitor the health of network functions?
- What are the risks for network functions if an attacker can access the SBA bus?
- Are the visibility and logging features provided by network functions sufficient, or should there be a proxy for added visibility?
- Should there be traffic inspection features for protecting network functions?

# SEPP Considerations

Security Edge Protection Proxy (SEPP) is a 5G network function defined by 3GPP. SEPP is a required 5G architecture component in situations where the network supports roaming use cases. Otherwise SEPP is not needed.

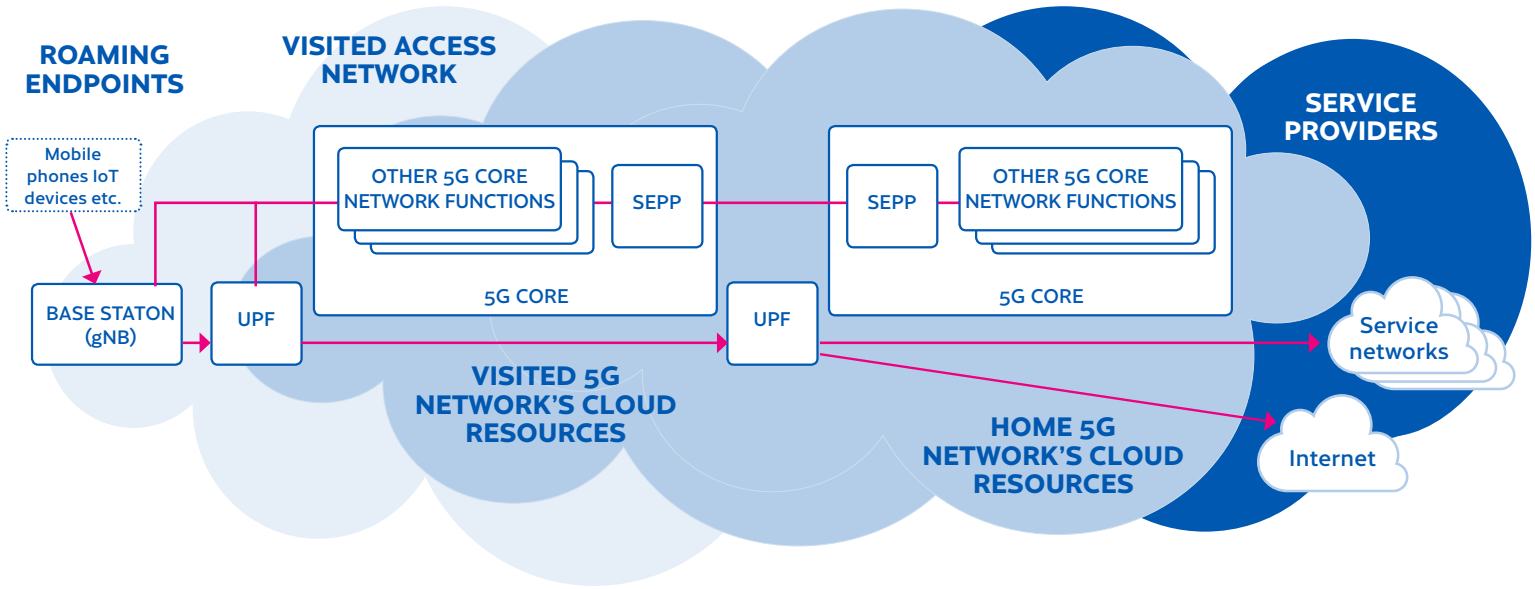



*Figure 14: Overview of roaming architecture when the option of routing through home UPF is used.*

SEPP is a non-transparent proxy allowing secure communications between service-providing and service-consuming NFs in separate 5G networks (PMLN). The separate networks in this case are the 5G home network and the 5G network visited by roaming. Security-related functionalities provided in SEPPs include:

- Topology hiding; hiding the internal architecture and topology of a PLMN from external PLMNs.
- Filtering and policing of information passed through the roaming interface.

SEPPs communicate over an interface called N32 and enforce security policies. 3GPP specifies that Transport Layer Security (TLS) shall be used between SEPPs if no IPX providers are in the path.

## IPX SERVICES

IPX services enable roaming so that mobile network operators do not need to maintain one-to-one connections with all the roaming partners. IPX maintains peering services for operators.

IPX providers may offer value-added services (VAS) on top of the interconnection service. To enable VAS, the provider may need to modify the data passing through

**Consider these questions when planning a 5G infrastructure:**

- What risks could attacks originating via SEPP lead to?
- What kind of message filtering options should we have to mitigate security risks?
- What kind of security policies should we be able to set to limit external exposure of our network?
- Are rule-based policies sufficient, or should we expect SEPP to contain some attack detection capabilities?
- What SEPP deployment model should we use?
- What is our risk appetite regarding IPX roaming traffic modification?

the roaming interface in a controlled way by using an IPX HTTP proxy. The modification policy can be exchanged between SEPPs to define what elements of the message an IPX can modify.

3GPP specifies that PRINS (Protocol for N32 Interconnect Security) shall be used if IPX providers are in the path. To avoid the complexities with maintaining a SEPP with PRINS interface to IPX providers and direct TLS connections to other SEPPs, operators may opt to use a hosted SEPP that is a SEPP service provided by the IPX operators.

**Consider these questions when planning a 5G infrastructure:**

- What dependencies will the network functions have to supporting systems?
- How should these integrations be implemented so that the risks of attacks are sufficiently mitigated?

## SECURITY CONSIDERATIONS

Due to the nature of the SEPP function, when present it is necessarily a critical security component for any 5G implementation. A vulnerability in SEPP could expose 5G core services to potential attackers. Additionally, a failure of SEPP to inspect and filter messaging would lead to exposure of sensitive data from within the 5GC.

It is expected that there will be varying approaches by vendors when it comes to the question of which kind of security functionality to include in their SEPP implementations when it comes to possible content inspection and filtering capabilities. Support of TLS and PRINS is mandated from all implementations. For instance, as attacks might evolve and get more complex, future SEPP implementations could include the use of external threat feeds or behavioural analytics engines to detect and prevent attacks against the gateway and the 5G network.

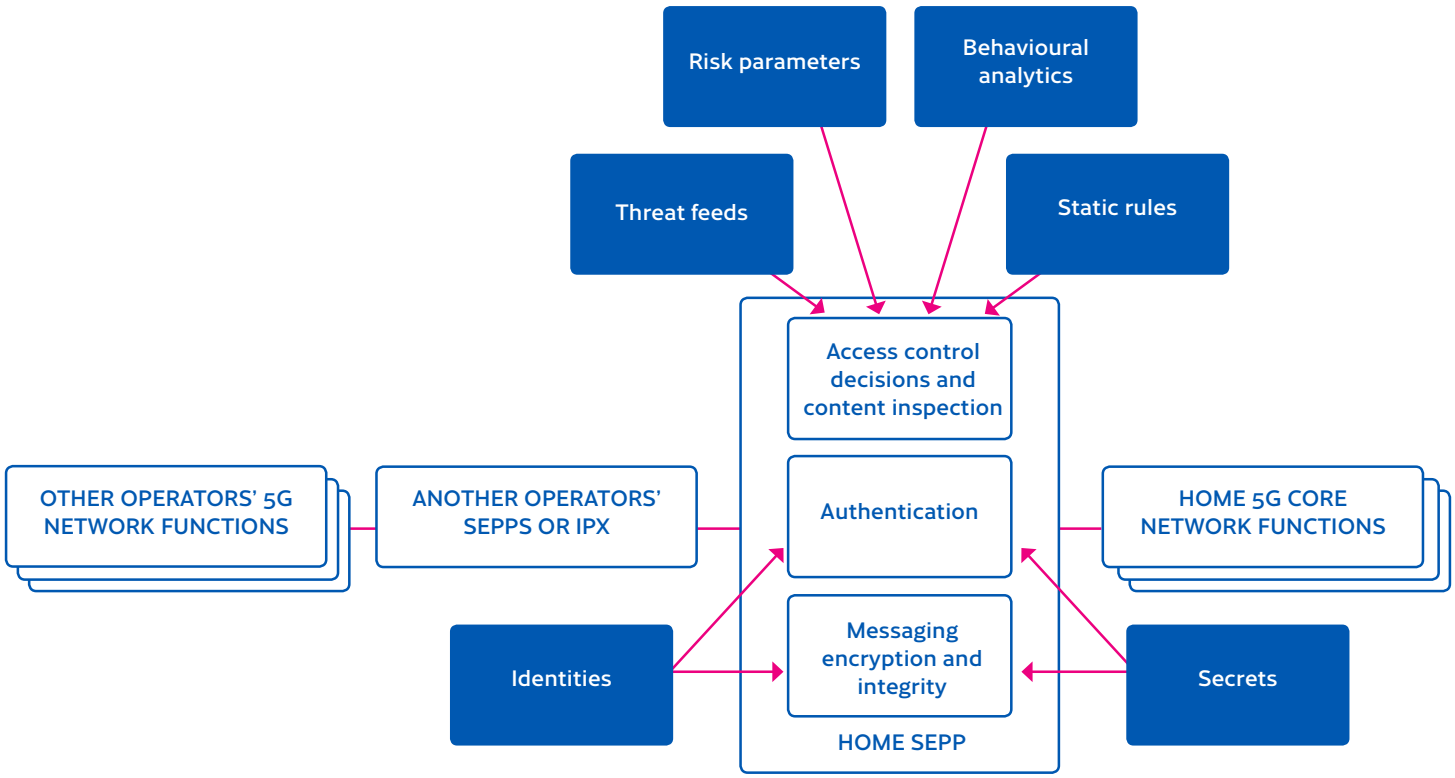



*Figure 15: Security functionality of SEPP is based on configurations, secrets, and possibly dynamic risk engines.*

# Integrations

The 5G system architecture as specified by 3GPP is focused on interoperability and interfaces between the standardised system components. An operator might choose to integrate trusted solutions with the 5G network functions, bypassing any 3GPP-defined APIs and gateways such as CAPIF and NEF.

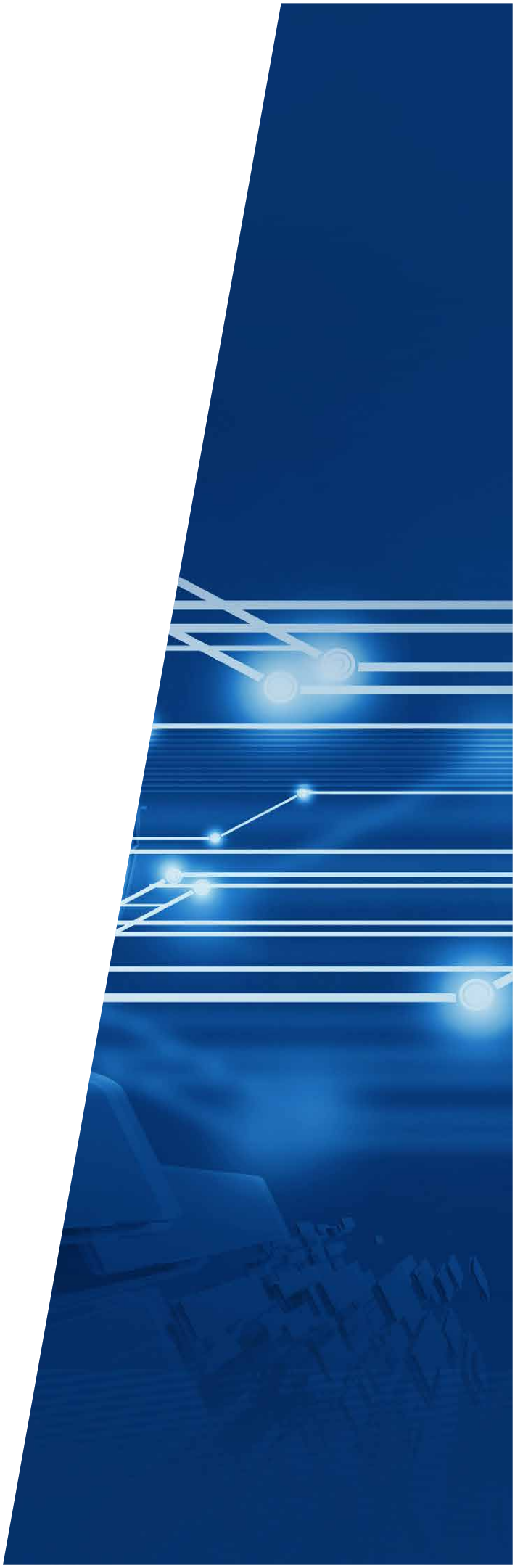
Often these integrations are not discussed in the 5G architecture context, as they are considered outside the scope from the standards' point of view. However, these architecture decisions might have serious security implications in real-world implementations.

## EXAMPLE: UNIFIED DATA REPOSITORY

Unified Data Repository (UDR) is a database service that UDM, PCF and NEF can use. The stored data can include e.g. subscription, authentication, application and policy data. In a 5G environment, there can be multiple UDRs, with each one having different sets of data or serving different network functions, for example.

UDRs are typically integrated with the operator's CRM or OSS/BSS systems to enable subscriber management and billing processes. How these integrations are achieved is often implemented in a way that makes them outside the scope of specifications. Common integration methods include SOAP and FTP.

From a security point of view, the integrations from UDRs to other systems create a potential attack vector avenue for an attacker to gain access to the 5GC service environment. An operator should consider how these integrations are designed and implemented so that they cannot pose a risk to the mobile network services.



# Data Transmission Security Considerations

5G, similar to previous mobile network generations, enables encryption of radio access and transmission communications.

Protection options for transmission from the base station varies between planes:

- For the control plane, IPSec with IKEv2 certificate-based authentication and DTLS are the required protocols to be implemented by vendors for protecting communications between the base stations and the 5G core.
- For the user plane, IPSec with IKEv2 certificate-based authentication shall be implemented by vendors. The encryption could be enabled from the base station to the UPF or terminated some place earlier using a security gateway (SEG) component.

## BASE STATION CU/DU SPLIT

Radio access network (RAN) is the part of the mobile network system that implements radio access. In 5G standalone networks, RAN consists of gNodeB (gNBs) base stations. The gNB s implement the new 5G radio protocols referred to as 5G NR.

The gNB can be a single unit with all the base station functionalities in it. 5G architecture also introduces a functional split into Central Units (CU) and Distributed Units (DU). A CU contains implementation of higher levels of protocol stack and can serve multiple DUs that include the radio functionality.

How the split is made in practice comes with security considerations, especially if DUs and CUs are installed at locations with different physical security protections. The 5G specifications support using IPSec between the CU/DU interfaces. Operators should be mindful about the threat of physical tampering and consider using these optional protections when needed.

## NON-3GPP ACCESS

It is expected that the so-called non-3GPP access methods are going to play an important part in extending 5G service coverage to locations such as office buildings. The wireless technology in this case could be e.g. IEEE 802.11 WLAN.

This access method is considered untrusted, as the wireless access technology is possibly outside the control of the 5G service provider and not part of 3GPP's technology specifications. The connecting endpoint must naturally support this access to 5G services. It will use the WLAN to access the operator's gateway device N3IWF.

The gateway authenticates the endpoint using EAP-AKA`/5G-AKA and after successful authentication establishes separate IPSec connections for control and user plane traffic with the endpoint. N3IWF terminates these IPSec control and user plane IPSec connections and routes the communications to AMF and UPF respectively.

## HOW MOBILE BASE STATION NAMING HAS EVOLVED

| Generation | Base station name |
|---|---|
| 1 | Base Station (BS) |
| 2 | Base Transceiver Station (BTS) |
| 3 | Node B (NB) |
| 4 | evolved Node B (eNodeB, eNB) |
| 5 | ng-eNB, an 4G/LTE base station that connects to 5G core<br><br>gNodeB, gNB, a 5G NR base station that connects to 5G core |

# Base Station Physical Threats

Base stations of 5G implementations will be exposed to physical threats as in any previous generation. In practice, it will not be fully possible to prevent access to and tampering with the installations. Such malicious access should, however, not lead to compromise of user communications, the 5G core functions, or the telco cloud.

## Threats to User Communications

5G specifications include integrity protection and encryption mechanisms for data transmissions. For the user plane, they cover the wireless data communications between user endpoints and gNodeB, and also the user plane communication between gNodeB and UPF. The specifications require equipment vendors to implement these functionalities and support the IPSec protocol for integrity, confidentiality and replay protection of user communications. However, making use of them for user plane protection remains at the discretion of the operators.

From a user's point of view, it is impossible to determine whether an operator is protecting the traffic of the user plane. It is therefore sensible to make use of application-level security protocols as is already commonplace in public networks today for messaging and internet access.

Operators should be mindful about the potential risk of physical tampering of cabling at base station locations and for backhaul connections; if making use of the optional encryption for the user plane would be appropriate, the services offered should be considered.

## Threats to 5G Network Functions

An attacker might tamper with gNodeB devices in an attempt to breach the 5G core using the control interfaces towards AMF and the core functions. Alternatively, an attacker might try to break into UPF to get access to SMF and the core functions. In order to prevent a breach towards the core network functions

- the interfaces of AMF, UPF and SMF should be especially hardened so that an attacker having access to the control or user plane would not be able to compromise the network functions,
- the secrets used for authenticating the base station should be protected in a way that they cannot be extracted from the device.

AMF and SMF could also contain technology to detect attempted attacks originating from the base stations and the user plane.

**Consider these questions when planning a 5G infrastructure:**

- Should we be able to detect attempted tampering of base stations?
- Could an attacker physically extract authentication keys from a base station?
- How are the 5GC network functions protected against attacks from the base station control interfaces?
- Are we going to be able to detect attacks toward the control interfaces?

**Finnish Transport and Communications Agency Traficom**
National Cyber Security Centre Finland
P.O. Box 320, FI-00059 Traficom Switchboard
+358 29 534 5000
traficom.fi/en/kyberturvallisuuskeskus.fi/en

TRAFICOM
Finnish Transport and Communications Agency
National Cyber Security Centre