PNNL-32015

5G Zero Trust Architecture

A Testable and Phased Approach

September 2021

Ryan J Poltermann
Johnathan V Cree
Aarne A Nixon
Patrick J O'Connell

Prepared for the U.S. Department of Energy
under Contract DE-AC05-76RL01830

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor Battelle Memorial Institute, nor any of their employees, makes **any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights**. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or Battelle Memorial Institute. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from the
Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062;
ph: (865) 576-8401
fax: (865) 576-5728
email: reports@adonis.osti.gov

Available to the public from the National Technical Information Service
5301 Shawnee Rd., Alexandria, VA 22312
ph: (800) 553-NTIS (6847)
email: orders@ntis.gov <https://www.ntis.gov/about>
Online ordering: http://www.ntis.gov

PNNL-32015

5G Zero Trust Architecture

A Testable and Phased Approach

September 2021

Ryan J Poltermann
Johnathan V Cree
Aarne A Nixon
Patrick J O’Connell

Prepared for
the U.S. Department of Energy
under Contract DE-AC05-76RL01830

Pacific Northwest National Laboratory
Richland, Washington 99354

Contents

Figures

Acronyms and Abbreviations

AUSF	Authentication Server Function
BGCF	Breakout Gateway Control Function
CU	Centralized Unit
DU	Distributed Unit
I-CSCF	Interrogating - Call Session Control Function
IMS	Internet Protocol Multimedia Subsystem
MC	Mission Critical
ME	Mobile Equipment
MEC	Multi-Access Edge Computing
MMS	Multimedia Messaging Service
N3IWF	Non-3GPP Interworking Function
O-RAN	Open Radio Access Network
OSI	Open System Interconnection
P-CSCF	Proxy - Call Session Control Function
QoS	Quality of Service
RAN	Radio Access Network
RF	Radio Frequency
RU	Radio Unit
S-CSCF	Serving – Call Session Control Function
SA	Stand Alone
SIEM	Security Information and Event Management
SMS	Short Message Service
UE	User Equipment
UPF	User Plane Function
URLLC	Ultra Reliable Low Latency Communications
V2X	Vehicle to Anything
VM	Virtual Machine
W-AGF	Wireline Access Gateway Function

1.0 Introduction

With the emphasis being placed on Zero Trust Architecture in technology and by Executive Order (*Executive Order on Improving the Nation's Cybersecurity*[1]), an investigation was performed as to the feasibility of implementing this approach into 5G. Reference documents include NIST Special Publication 800-207[2] and Department of Defense (DOD) Zero Trust Reference Architecture[3], and these provided guidance as to the overall approach.

Progress is being made within 5G to address the lack of Zero Trust Architecture; a June 2021 DoD press release indicated that "*The prototype 5G network is built on the next generation of Open Radio Network standards and designed to comply with DOD specifications for zero-trust architecture for native security and secure connectivity with other networks.*"

1.1 Zero Trust Architecture Components

Whereas traditional defense-in-depth provides layers of protection, Zero Trust Architecture ensures that communication cannot freely occur within each layer. Zero Trust Architecture extends beyond this, ensuring that each device is verified and that the type of communication being requested is allowed.

NIST SP 800-207 outlines various methods for implementing Zero Trust Architecture within Section 3.1, and our approach includes multiple methods. While NIST SP 800-207 is considered the primary document used, the Department of Defense is more approachable. The Department of Defense has an excellent diagram showing the components within Zero Trust.

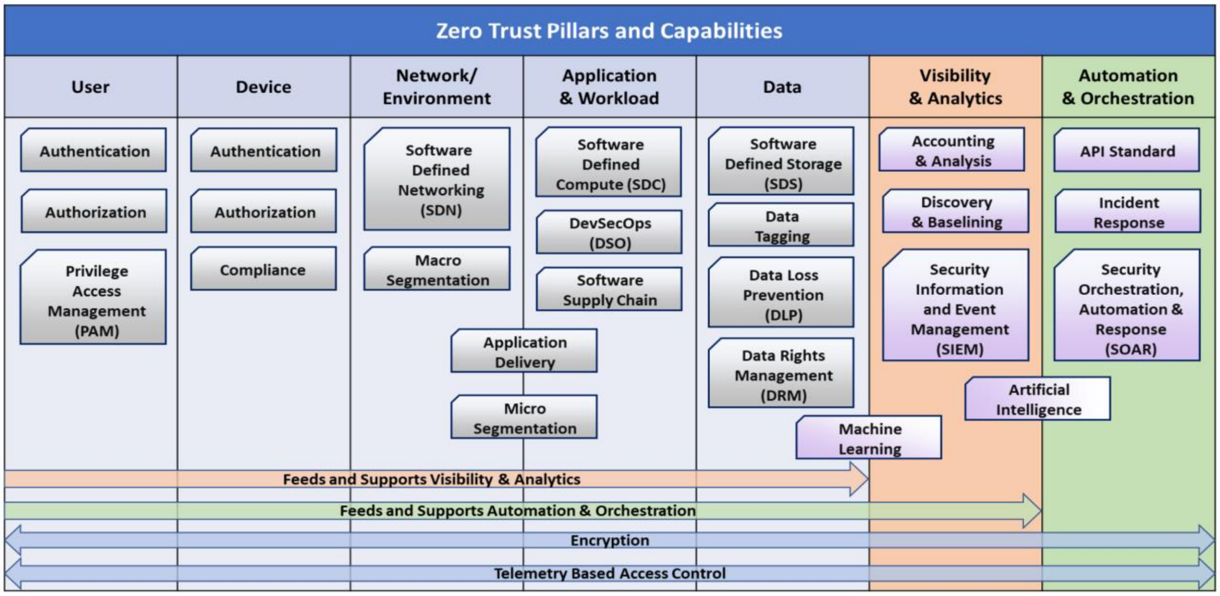
Zero Trust Pillars and Capabilities
User
Device
Network/ Environment
Application & Workload
Data
Visibility & Analytics
Automation & Orchestration
Authentication
Authorization
Privilege Access Management (PAM)
Authentication
Authorization
Compliance
Software Defined Networking (SDN)
Macro Segmentation
Software Defined Compute (SDC)
DevSecOps (DSO)
Software Supply Chain
Application Delivery
Micro Segmentation
Software Defined Storage (SDS)
Data Tagging
Data Loss Prevention (DLP)
Data Rights Management (DRM)
Machine Learning
Accounting & Analysis
Discovery & Baselining
Security Information and Event Management (SIEM)
Artificial Intelligence
API Standard
Incident Response
Security Orchestration, Automation & Response (SOAR)
Feeds and Supports Visibility & Analytics
Feeds and Supports Automation & Orchestration
Encryption
Telemetry Based Access Control

Figure 1. DoD Zero Trust Pillars

[1] https://www.whitehouse.gov/briefing-room/presidential-actions/2021/05/12/executive-order-on-improving-the-nations-cybersecurity/

[2] https://csrc.nist.gov/publications/detail/sp/800-207/final

[3] https://dodcio.defense.gov/Portals/0/Documents/Library/(U)ZT_RA_v1.1(U)_Mar21.pdf

Our focus in this phase of research will be on the following DoD Trust Pillar aspects:

- Software Defined Networking
- Macro/Micro network segmentation
- Device Authentication, Authorization, and Compliance

While Security Information and Event Management (SIEM) and compliance scanning such as OpenSCAP is very desirable, due to the limited performance period it is being left for additional study.

1.2 Virtualization and Containerization Approaches

The primary approach for securing the 5G system is the use of service mesh proxies, virtualizing the components (services), and implementing grouping of Virtual Machines [VMs] or a group of Kubernetes containers into pods to segment the networks, with service mesh proxies connecting pods together. This terminology may vary by vendor, and Kubernetes terminology is used.

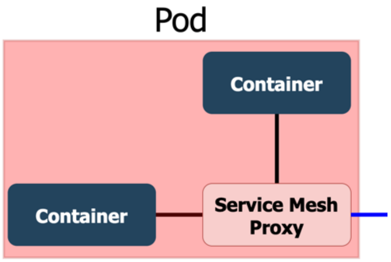

Figure 2. Pods, Containers, and Service Mesh Proxy

Service mesh proxies such as Istio and Consul support both Virtual Machines (VMs) and Kubernetes containers, and the approach documented in this paper addresses Virtual Machines. It is an important caveat that the security approach is dictated by the configuration of the 5G components. As a simplified example, if the 5G core only uses one virtual machine for the entirety of its capabilities, it will not be possible to segment that 5G core.

While Virtual Machines have an operating system, containers typically do not have an operating system. Kubernetes containers have unrestricted network access to the services within a pod, and the IP address is assigned to each pod. While not shown in this paper, an approach has been created for containers as well.

Service mesh proxies target Layer 7 of the OSI model, and this could be a significant obstacle to fully extending the zero trust approach within 5G. Protocols, particularly within the gNodeB, are operating at a lower level than Layer 7. There is at least one solution (Network Service Mesh) that operates down to Layer 3, and this requires further investigation.

Diagrams within this document are in some respects simplified. Service mesh proxies act as “gatekeepers” for the pods, but accurately depicting this causes significant complexities to the diagram while inhibiting its comprehensibility. Therefore, liberties were taken to ensure understanding and the diagrams do not directly reflect how it would be configured in an operational system.

1.3 Architecture Capabilities

Zero Trust Architecture does not exist within 5G standards but is expected to be a core component of 6G. To understand the potential implementation, an overview of 5G capabilities is required.

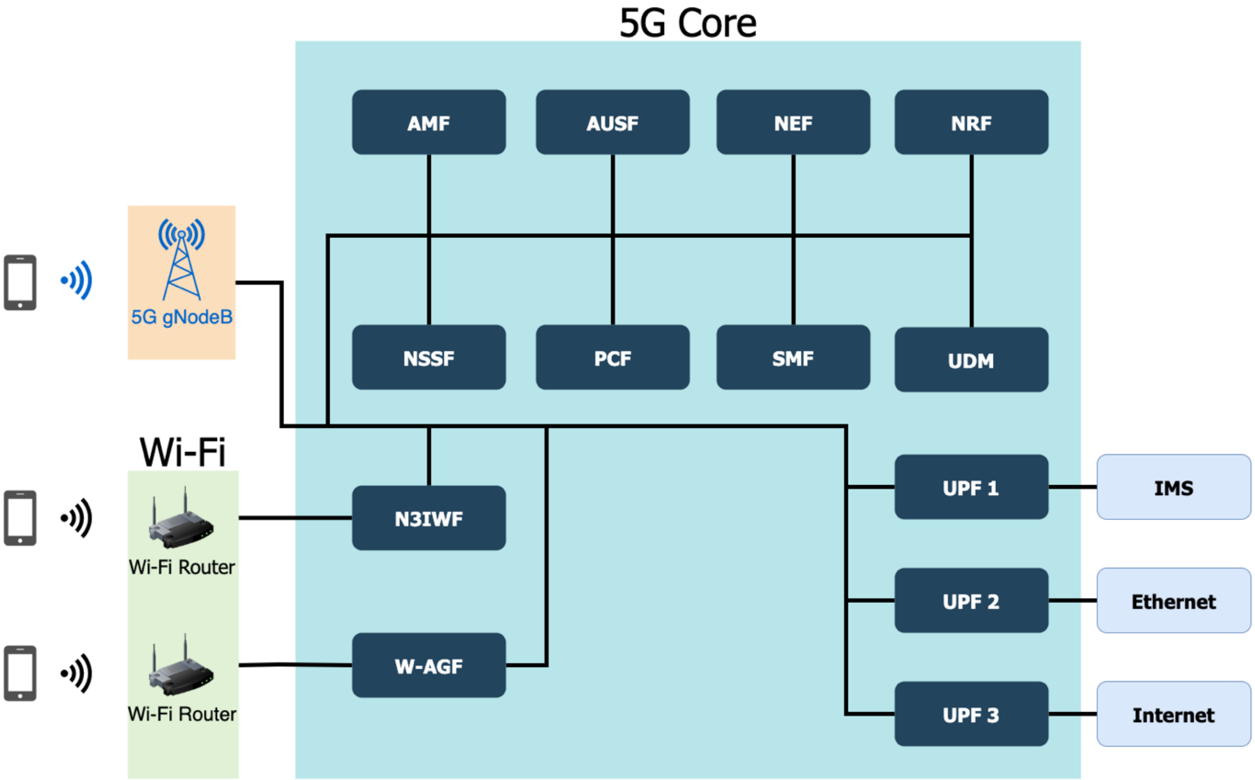

Figure 3. Standard 5G Core Overview

For simplicity, not all components of the 5G core are shown. A mobile device (UE) accesses a 5G SA system through a gNodeB, which provides the air interface. The AUSF validates and authenticates the mobile device, and the 5G device also validates and authenticates the 5G network. The PCF provides the information of the permitted activities and QoS for the UE.

The UPF provides conduits for the user's network traffic, which may be separated based on the category of users. The IMS provides voice/text capabilities and notice that the IMS is located off the User Plane Function (UPF). This is in direct contrast to the 4G approach, which had the IMS within the 4G core. The N3IWF and W-AGF provide access to a connected Wi-Fi network as well as a wireline network. Mission Critical Services is not included within the 5G system as the standards are still nascent, and this requires future study when 5G MC Services standards are finalized.

The gNodeB is simplified and may contain three components: the Radio Unit (RU), the Distributed Unit (DU), and the Centralized Unit (CU). The RU is responsible for RF transmissions, the DU is for local control of the RU, and the CU can coordinate multiple DUs. The initial Zero Trust Architecture phases for gNodeB will be covered after the 5G core.

2.0 Phase Details

To provide an iterative approach to implementation and testing, a spiral approach is detailed, so that baseline zero trust components can be established before adding more features. This accomplishes two things: 1) To test and demonstrate applied concepts, such as the service mesh proxy, and 2) to allow adaptation in further phases based upon lessons learned in earlier phases.

This section details each Phase, and what features and development they entail. Once again, the Virtual Machine (VM) approach is slightly simplified for clarity.

2.1 5G Core Phase 1

The 5G core is split into two primary components (pods), the servers and the UPFs. The server network is not directly exposed to the traffic of any external network. UPF firewalls would either be using the service mesh proxy policies or potentially external to the pod.

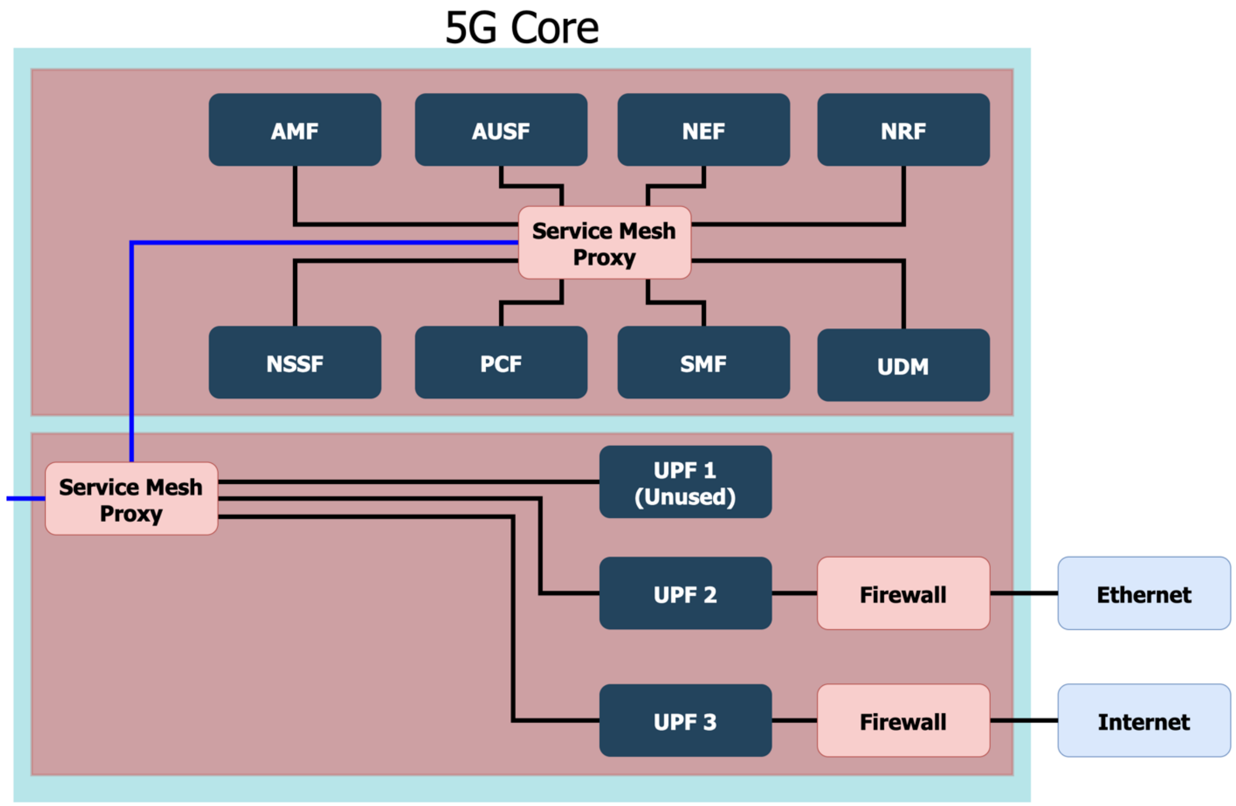

Figure 4. 5G Core Phase 1

This approach is manageable but does have a significant restriction: the service mesh proxy (bottom left) that connect the gNodeBs to the core must process significant amounts of traffic (potentially 20 Gbps per sector antenna) with minimal delay (milliseconds for URLLC). This restriction does not make the design approach impossible, merely much more expensive.

One additional issue that occurs is a single point of failure with the service mesh proxy serving the gNodeBs. By dividing the gNodeB sites over multiple service mesh proxies, this removes the single point of failure for this particular component (assuming there isn't a common mode failure such as configuration). It can also act as a load balancer between the service mesh proxies. This also reduces the overall cost of the mesh proxy solutions, as lower performance firewalls can be used instead of a single, extremely high-performance firewall.

2.2 gNodeB Phase 1

The essential structure of a gNodeB is shown above and is split into three components: the Radio Unit (RU), the Distributed Unit (DU), and the Centralized Unit (CU). Not all gNodeBs will have the capabilities split into three components, and it may be one device performing all functions.

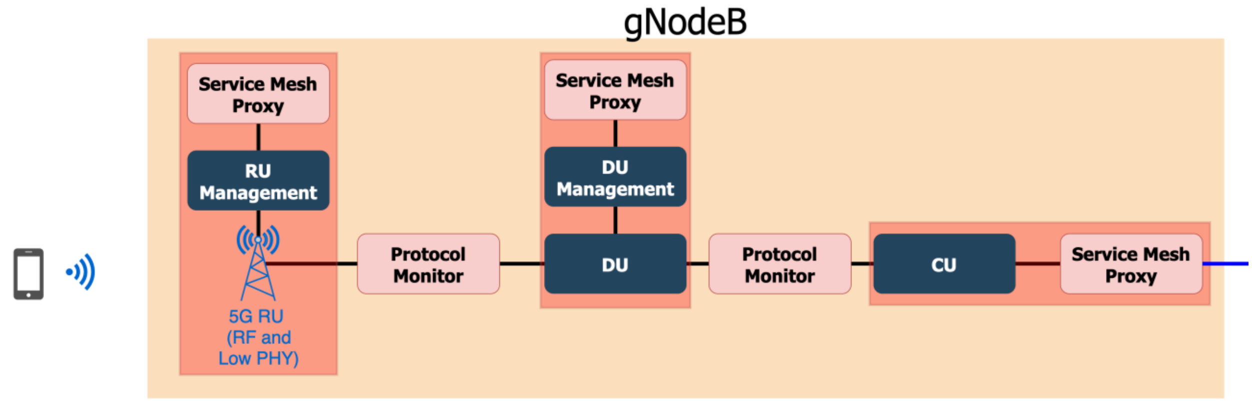

Figure 5. gNodeB Phase 1

There are multiple points where the RU, CU, and DU responsibilities can be split, and certain configurations are more complex. For our configuration, we will assume the most “complex” split, which is Option 7.X. This aligns with the approach that various vendors are using as well as aligning with O-RAN specifications (known as Option 7.2) but introduces potential difficulties. In this configuration, the communication layers may be lower than service mesh proxies typically operate.

Each device (RU, CU, and DU) is untrusted from the other. While each is shown in the same light orange box, these components do not need to be installed at the same location. As an example, the CU is not required to be installed at the gNodeB site and can also control multiple DUs. As well, the DU could also be installed separately from the RU, though significant latency restrictions exist. These configurations require additional security requirements as the devices may be physically exposed to the general public.

It’s worth noting that there is traditionally a separate router at the cell site and at the core; these are not included in the diagrams. The assumption is made that the Service Mesh Proxy is acting as the site router.

2.3 5G Core Phase 2

Phase 2 of the core incorporates the IP Multimedia Subsystem (IMS), which handles voice calls and SMS/MMS. It is partially untrusted in this phase.

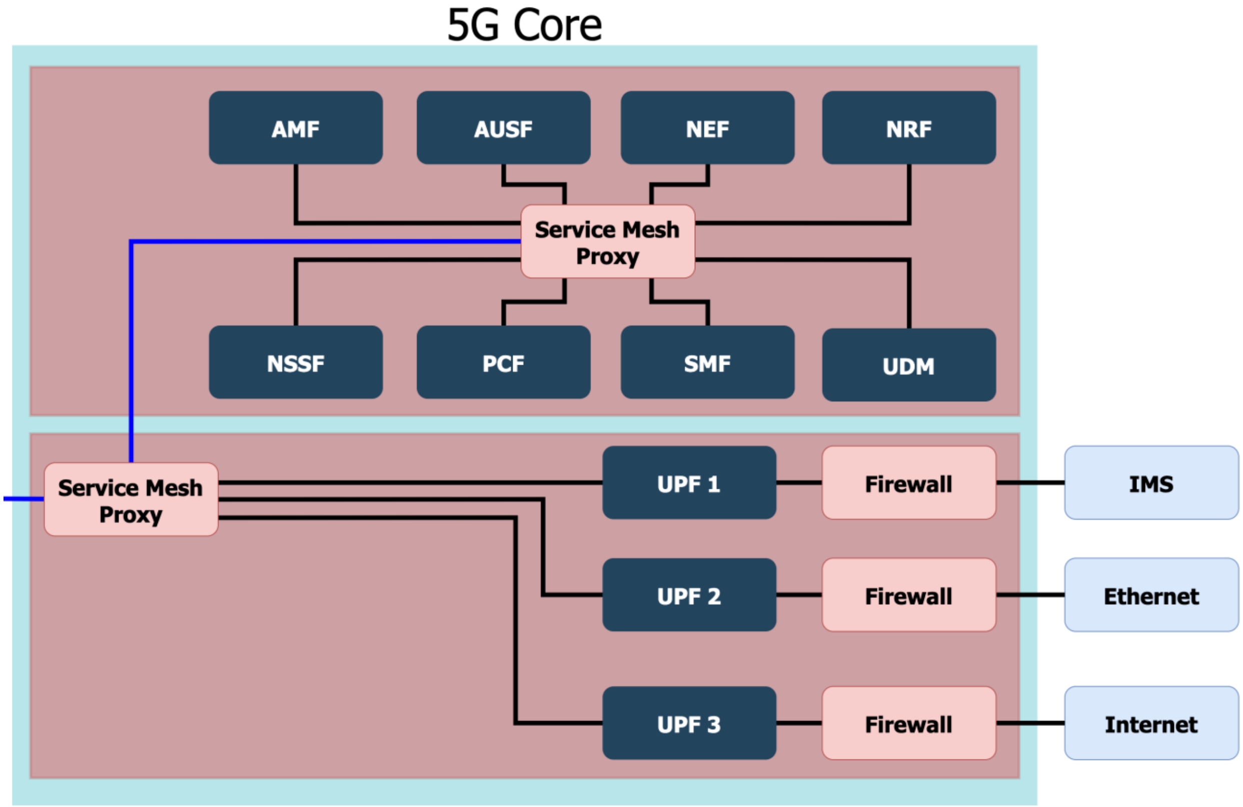

Figure 6. 5G Core Phase 2

The IMS will provide telephony and texting capabilities from UE to UE as well as integration into the PNNL telephone system. The IMS is shown abstracted, which obscures the significant number of components within the IMS. The implementation of the IMS depends not only on the vendor but the configuration as well.

2.4 gNodeB Phase 2

Introductions of a Protocol Monitor between the RU/DU/CU provide some external analysis to ensure that communications or the devices themselves have not been compromised.

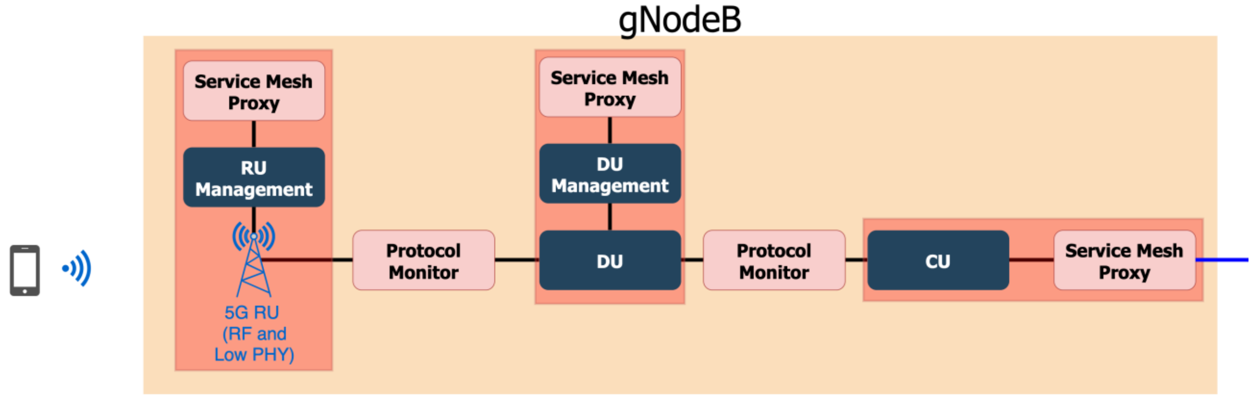

Figure 7. gNodeB Phase 2

The Protocol Monitor challenges develop as the levels of communication become lower (i.e. more difficult moving from right to left). This is due to the lower level of communications occurring between the devices as well as the increasing sensitivity to latency.

Latencies typically are within the following parameters:

- RU to DU – 75 us one way
- RU+DU to CU – 5 ms one way

The RU to the DU represents a significant limitation, and therefore requires minimally invasive processing. The protocol monitors will most likely not be able to take direct action against malicious communications but will send an alarm.

2.5 Phase 3

The primary focus will be on the IMS Zero Trust Architecture. This is a significant undertaking and is entirely dependent on the vendor used. As an example, some functionality, such as the P-CSCF/I-CSCF/S-CSCF or BGCF/I-CSCF/S-CSCF, is combined in some vendors.

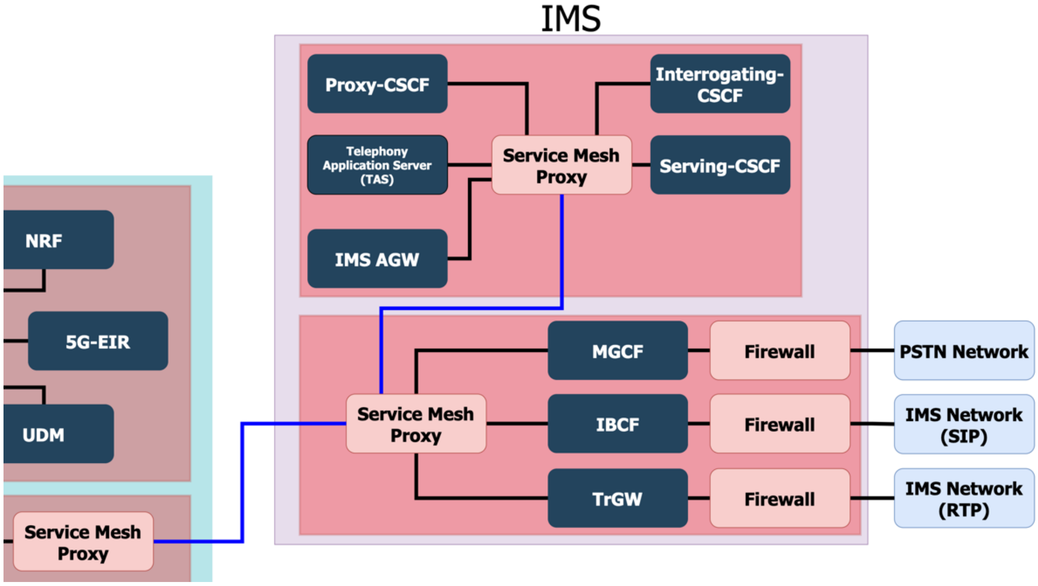

Figure 8. 5G Core Phase 3

In addition, Wi-Fi is introduced and will be incorporated into the 5G core in a Zero Trust Architecture approach. Wi-Fi devices will be capable of accessing data as well as text/voice.

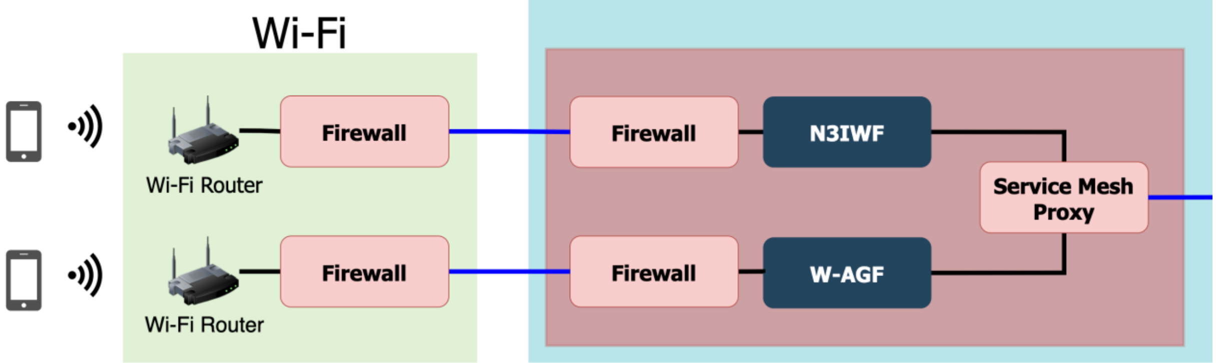

Figure 9. Wi-Fi Phase 3

Incorporating both methods of integrating Wi-Fi methods allows for additional security, even in “trusted” Wi-Fi networks.

Phase 3 also introduces support for multiple gNodeBs as well as interoperability between two gNodeBs for handoff purposes.

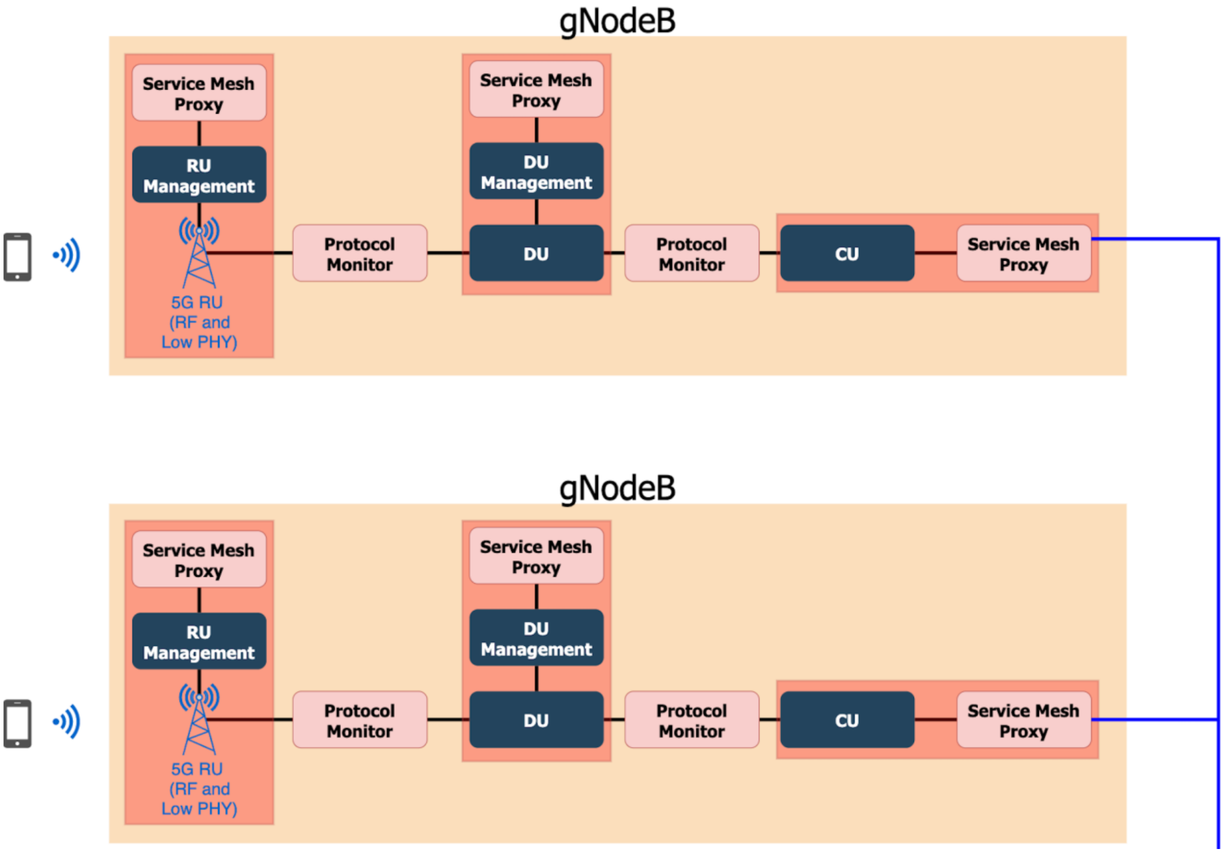

Figure 10. gNodeB Phase 3

The interactions between the gNodeBs are handled by the Xn interface and represents an important milestone to a deployable system.

2.6 Phase 4

Multi-Access Edge Computing (MEC) is introduced and will be incorporated into the 5G core in a Zero Trust Architecture approach. MEC allows for more local data applications, reducing latency.

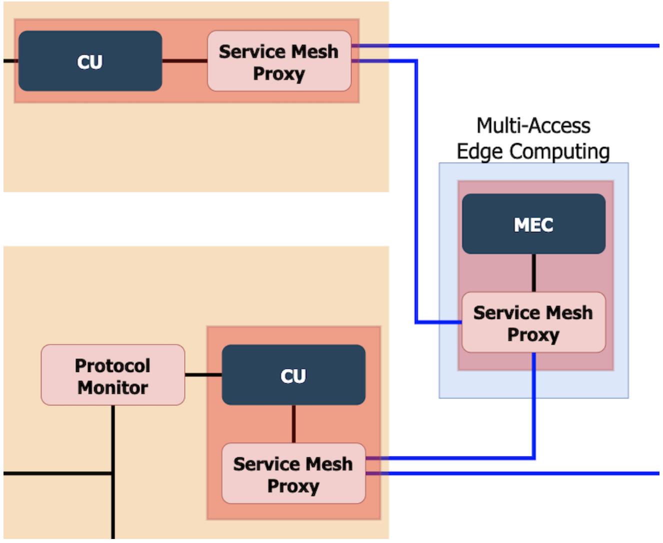

Figure 11. 5G Core Multi-Access Edge Computing (MEC)

MEC is abstracted at this point, and the assumption in the diagram above is that the network is completely local to this area. Should an external network connection be required, it will require additional security measures for the remote servers.

V2X servers will also be integrated into the system. However, the V2X applications will not be internally configured for Zero Trust Architecture within the included phases.

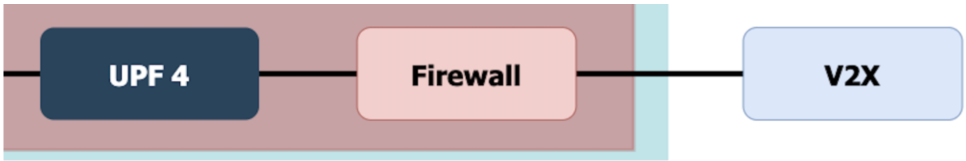

Figure 12. V2X Servers

This phase takes an existing CU and uses it to control another DU/RU, which is similar to how commercial deployments operate.

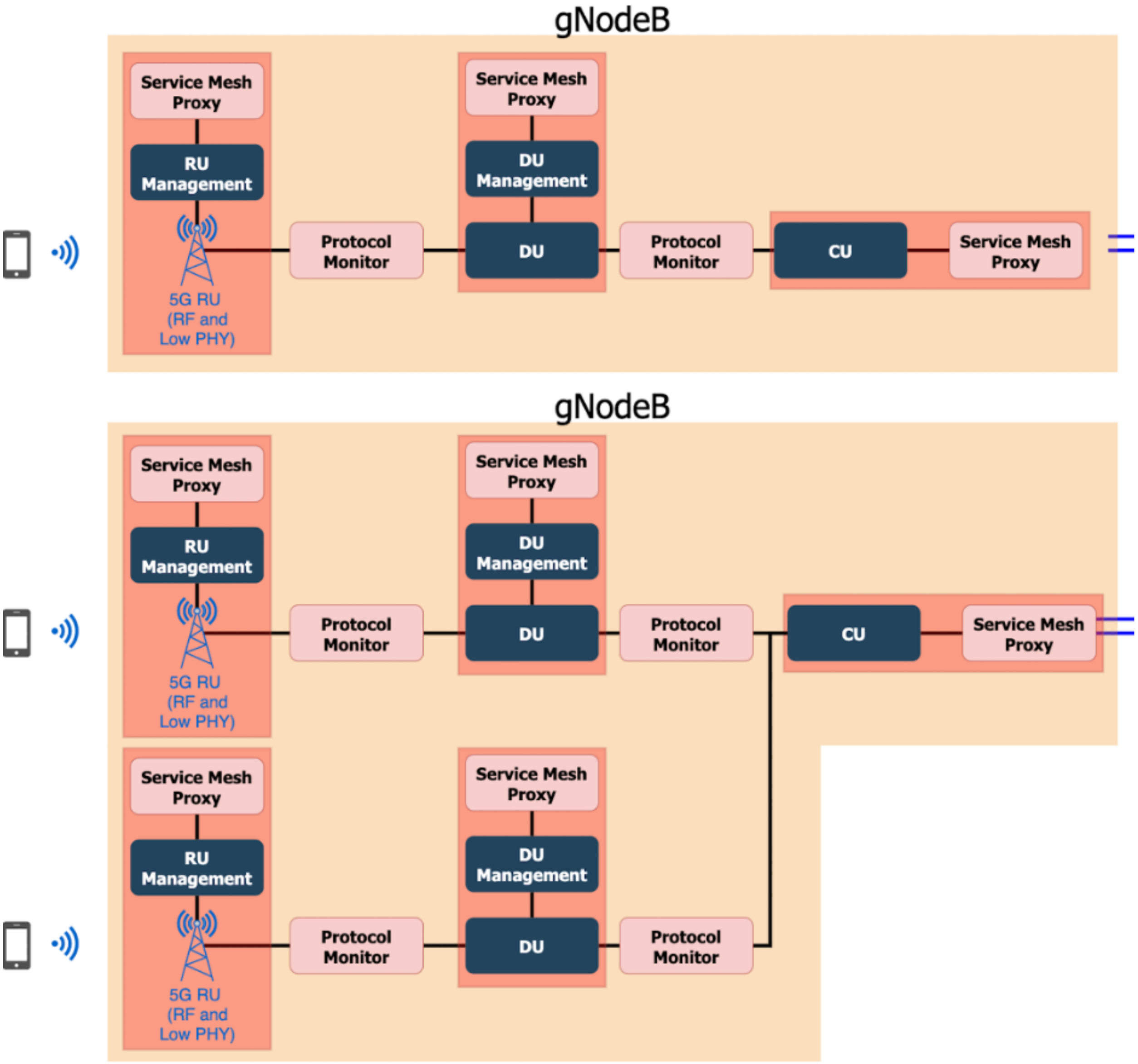

Figure 13. gNodeB Phase 4

While only one protocol monitor is shown in each path between the DU and the CU, an additional protocol monitor before the CU would be included in a real environment. This would be due to the difference in location between the DU and the CU, with one residing at each end.

2.7 Beyond Phase 4

The final phases will include O-RAN integration with basic Zero Trust Architecture.

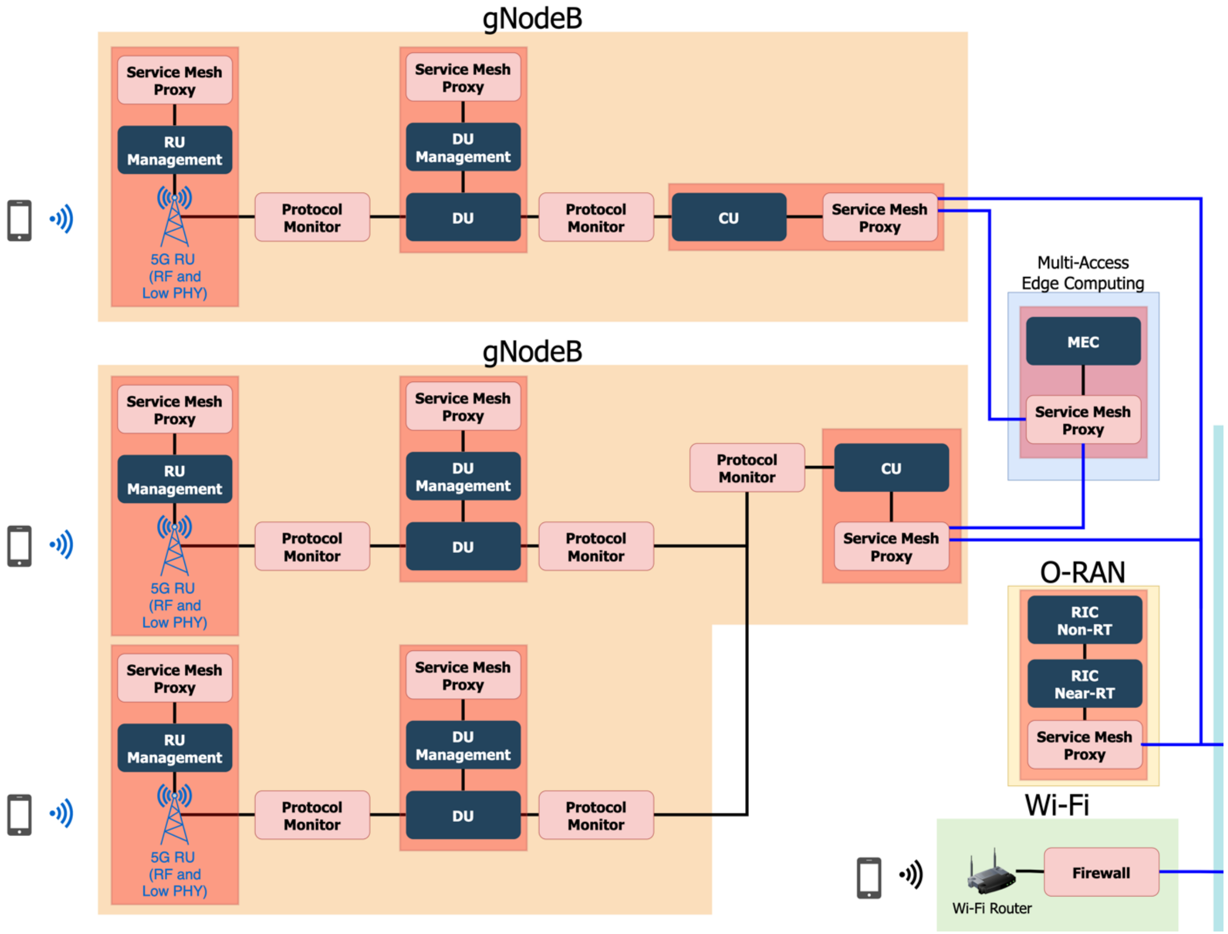

Figure 14. 5G Core and gNodeB with O-RAN, Phase 5

Protocols at the top gNodeB between the RU/DU/CU are using O-RAN standard protocols, and protocol monitors would be updated as required to support this. Fujitsu has a white paper[1] available about security within O-RAN and includes components of Zero Trust Architecture.

Complexity is introduced by including O-RAN; both the Non-Real Time RIC and the Near-Real Time RIC communicate to the RU, the DU, and the CU. Therefore, the primary efforts are to ensure proper communication between the gNodeB components and the O-RAN framework.

[1] https://fujitsu.lookbookhq.com/light_reading/open-ran-security-wp?utm_source=promo&utm_medium=d-ad

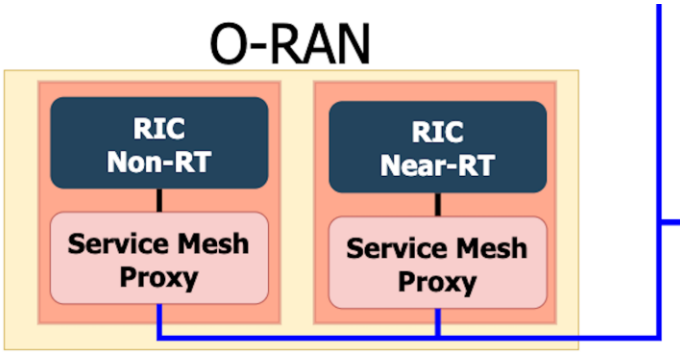

Figure 15. O-RAN, Phase 6

Within Phase 6, Advanced Zero Trust O-RAN integration will occur, segmenting the O-RAN control architecture further. This diagram is still simplified, but the protocols and communications that can occur are complex. All gNodeBs would be using O-RAN standard protocols.

The Non-Real Time RIC and the Near-Real Time RIC communicate between each other, and the separation of the two functions requires careful consideration. The Non-Real Time RIC is capable of AI/ML, and the behavior of the device may not be predictable.

While the 5G core has not been directly modified for some phases, it would be ideal to isolate each of the servers (which are network functions) individually. This will require working with the appropriate 5G core vendor to place each network function in its own pod. It may be possible to take advantage of the control plane and user plane separation to enhance security, and this requires further study.

3.0 Next Steps

As mentioned towards the beginning of the document, it is an important caveat that the approach is dictated by the configuration of the 5G components. As an example, if the 5G core only uses one virtual machine for the entirety of its capabilities, it will not be possible to segment that 5G core. This means that the generalized approach demonstrated will require being tailored to the software components available.

The next initial steps would be to implement this approach based on the core, IMS, and gNodeB equipment available to PNNL. Some vendors are more virtualized and/or containerized than others, and procuring a core along with an IMS that is more aligned with this technology approach would simplify the steps while maximizing the value of PNNL efforts. The process would be further refined after each successful iteration, isolating more functionality as vendors allow.

Pacific Northwest National Laboratory

902 Battelle Boulevard
P.O. Box 999
Richland, WA 99354
1-888-375-PNNL (7665)

www.pnnl.gov