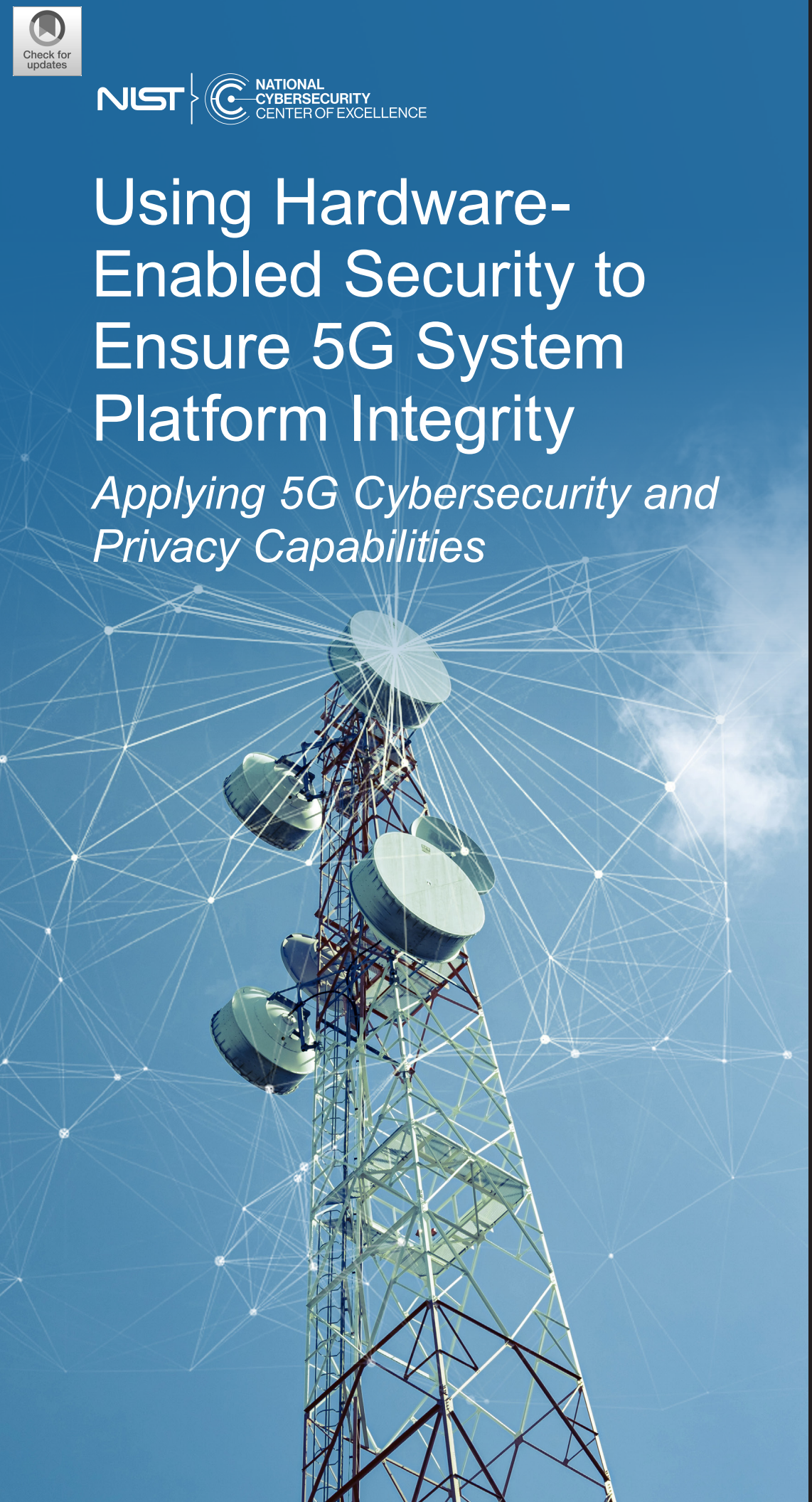
NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

Using Hardware-Enabled Security to Ensure 5G System Platform Integrity

Applying 5G Cybersecurity and Privacy Capabilities



NIST Cybersecurity White Paper

NIST CSWP 36B

Michael Bartock
Jeffrey Cichonski
Murugiah Souppaya *
Information Technology Laboratory

Karen Kent
Trusted Cyber Annex

Parisa Grayeli
Sanjeev Sharma
The MITRE Corporation

**Former NIST employee; all work for this publication was done while at NIST*

March 2026
Final

This publication is available free of charge from:
https://doi.org/10.6028/NIST.CSWP.36B

Abstract

This white paper provides an overview and an example of employing hardware-enabled [1] security capabilities to provision, measure, attest to, and enforce the integrity of the compute platform to foster trust in a 5G system's server infrastructure. It discusses security threats within computing environments and how leveraging hardware roots of trust (HRoT) and remote attestation can help mitigate specific threats. This white paper is part of a series called Applying 5G Cybersecurity and Privacy Capabilities, which covers 5G cybersecurity- and privacy- supporting capabilities that were demonstrated as part of the 5G Cybersecurity project at the National Cybersecurity Center of Excellence (NCCoE).

Audience

Technology, cybersecurity, and privacy professionals who are involved in using, managing, or providing 5G-enabled services and products. This includes potential private 5G network operators, commercial mobile network operators, and end-user organizations. Readers should already be familiar with the basics of mobile network architectures and components.

Keywords

3GPP, 5G, cybersecurity, firmware, hardware, hardware-enabled security, hardware root of trust

Acknowledgments

We are grateful to the following individuals for their generous contributions.

AMI: Haripriya Bashyam, Thomas McCarthy, Muthukkumaran Ramalingam, Presanna Raman, Stefano Righi

AT&T: Jitendra Patel, Bogdan Ungureanu CableLabs: Tao Wan

Cisco: Matt Hyatt, Kori Rongey, Steve Vetter *

Dell Technologies: Dan Carroll

Intel Corporation: Steve Orrin

Keysight Technologies: Corey Piggott

MiTAC Computing Technology Corp.: Simon Hwang

The MITRE Corporation: Sallie Edwards *, John Kent, Mary Raguso *, Theresa Suloway

NIST: Cherilyn Pascoe, Adam Sedgewick *, Kevin Stine

Nokia: Gary Atkinson, Rajasekhar Bodanki, Robert Cranston, Jorge Escobar, Don McBride

Palo Alto Networks: Aarin Buskirk, Bryan Wenger

T-Mobile: Todd Gibson

** Former employee; all work for this publication was done while at that organization*

Overview

3rd Generation Partnership Project (3GPP) standards specify 5G systems as service-based architectures (SBAs) with a design that works well when implemented with cloud-native technologies leveraging microservices and container technology. In short, for the first time, the cellular system's core network can operate like a modern cloud application that is scalable and resilient. Previously, in LTE (4G) network functions (NFs) were virtual machines and before that in 3G, NFs were physical dedicated telecommunication appliances. Now, a single 5G NF can be comprised of a multitude of software containers running on many distributed servers. Hence, NF operation and management are now largely automated, relying on container orchestration engines to scale up on demand.

This shift enables NFs and 5G core networks to run on commodity servers instead of purpose-built telecommunications equipment. In addition, there is an opportunity to adopt existing advanced cybersecurity capabilities and techniques available in cloud platforms that have supported traditional information technology workloads. Implementing hardware-enabled platform integrity measurements and asset tags, both based on hardware roots of trust, can provide a stronger foundation for cloud-native infrastructures of 5G systems.

What's the problem?

The data center threat landscape has evolved to encompass numerous attack surfaces that can be targeted by attackers seeking persistent access. With organizations paying more attention to software security, attackers are pushing lower in the layers of the platform stack shown in Figure 1. This forces security teams to address attacks that threaten the platform firmware and hardware below the operating system (OS). For example, a rootkit named LoJax [2] discovered in 2018, can be installed in server firmware, which allows it to persist despite OS reboots or reinstallations and to remain invisible to OS malware scans, while providing the attacker with full access to the OS and potentially everything above it.

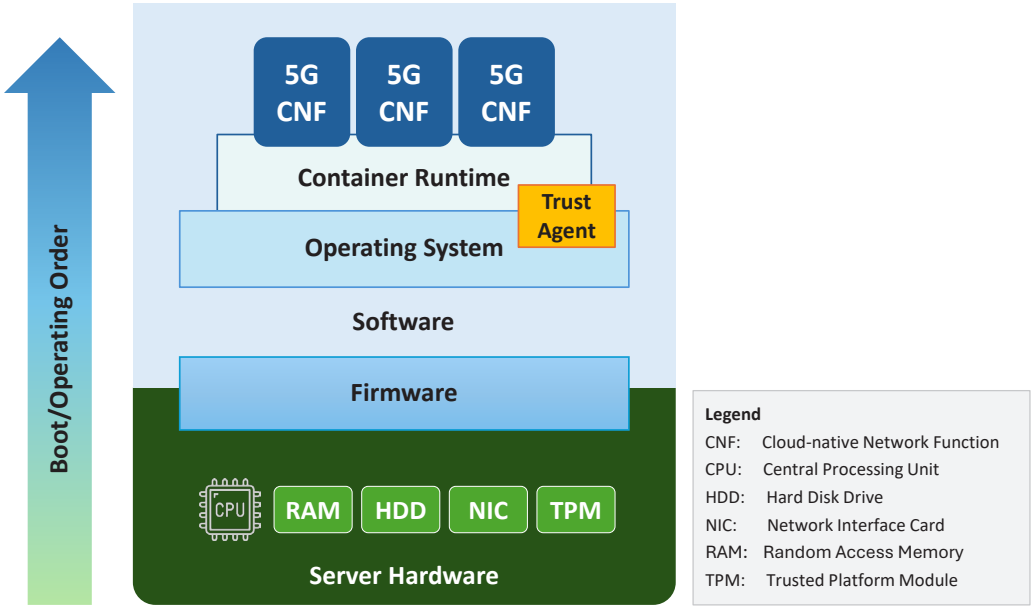


Figure 1. Simplified platform stack

In a 5G environment, these types of attacks could involve modifying configurations to send traffic to unauthorized locations, gaining access to control 5G cloud-native network functions (CNFs), stealing proprietary network software, or exporting the subscriber information. Traditional cybersecurity protections for cloud infrastructures are often rooted in firmware or software. This makes them inadequate to address these attacks because software- and firmware-based protections run at the same layer as the attacks. For example, if firmware can be successfully exploited, the security controls based in that firmware can most likely be compromised in the same fashion and therefore shouldn't be trusted to detect malware. There needs to be a mechanism below the firmware that can help detect and prevent these threats.

A 5G system is composed of the hardware, network, and software components for operating the 5G environment. As 5G systems adopt automated cloud-native applications running within a datacenter, it can be challenging to track every server in the large clusters of systems that support the 5G infrastructure. Because of this, it can be very difficult to ensure that 5G CNFs run on their intended servers for the purpose of isolating and enforcing where the critical functions operate. One approach is to use a tag to label a critical CNF so that it can be provisioned to a particular set of dedicated servers. While software tags in 5G container orchestrators can be one way to achieve this, there is no guarantee that the software tags are associated with the correct physical servers. In the context of this paper, the terms "platform" and "server" are used interchangeably.

How can hardware-enabled security address the problem?

The 5G standards defined by 3GPP do not specify cybersecurity protections for the underlying commodity components that support and operate the 5G system; these aspects are deemed implementation-specific. It is expected that mobile network operators will make a risk-based decision on the countermeasures to mitigate attacks against their commodity components.

In order to identify and protect against firmware and software attacks, each server that makes up a 5G system can employ hardware-enabled security mechanisms. Hardware roots of trust (HRoT) (see Section 3.2 of [1]) can help mitigate threats by establishing and maintaining *platform trust*—an assurance in the integrity of the underlying 5G server configuration, including hardware, firmware, and software. This is achieved by cryptographically measuring those components in a sequence and then saving the measurements to a secure storage element such as a Trusted Platform Module (TPM) [3] on the 5G server. If a server's cryptographic measurements have not changed, then there is assurance that none of the hardware or firmware components have been altered by the attacker through malicious code injection or other means.

Platform integrity measurement and storage within hardware provides a stronger foundation than that offered by software because of the immutability property of hardware: a physical component must be changed to subvert the process, versus a change to software or code if measurements are performed in software. Additionally, performing the measurement and storage of platform integrity within hardware modules happens at the lowest level of the stack, before any malicious software can inject itself into the process.

An additional feature commonly associated with platform trust is the concept of asset tagging. Asset tags are simple key-value attribute pairs that are associated with a 5G platform, like geographic location, company name, division, or department. These key-value attributes are tracked and recorded in a Remote Attestation Server (RAS) (see Section 6 of [1]) and can be provisioned directly to a 5G server through a trust agent. The trust agent can then secure these attribute associations within the host platform by writing a checksum in the form of cryptographic hash measurement data for the asset tag information to a hardware security module, such as the platform TPM. The asset tag hash is then retrieved by the RAS as part of the TPM report and included in the platform trust report evaluation.

While enabling HRoT for 5G systems provides a critical mechanism for ascertaining platform integrity, it's only one piece of the solution. Since the 5G system is made up of tens, hundreds, or even thousands of individual physical servers, there needs to be a way to collect and maintain all of their platform integrity measurements. Each server can send its latest boot-time platform measurements to the RAS. The RAS maintains these up-to-date platform measurements, as well as a list of allowed measurements. A platform is considered "trusted" if its current boot-time measurement matches one from the allowed list. The RAS compares each 5G server's current platform measurement with the list of trusted measurements to see if it can be trusted to host 5G CNFs.

There also needs to be a way to ensure that CNFs are only deployed on servers that are trusted. To achieve this, the 5G CNF orchestrator that is responsible for starting and stopping 5G CNF workloads communicates with the RAS. Each time the orchestrator wants to start a new instance of a 5G CNF, it can first ask the RAS for a list of trusted servers, and then deploy the 5G CNF workload to one of the servers with a current status of trusted.

The HRoTs enable additional security capabilities for the infrastructure supporting 5G beyond what is defined in the 3GPP specifications. These capabilities include hardware-enabled controls to:

- measure platform integrity for each server in the 5G infrastructure at each boot;
- assign specific labels for each server in the infrastructure;
- remotely attest each server's measurements and labels against policies; and
- use the results during 5G workload orchestration to enforce trust status of the host platform.

By providing this assurance, mobile network operators can gain a level of visibility and control over where access to CNFs and data is permitted, and know that the hardware infrastructure where 5G workloads are executing hasn't been tampered with.

How can I use a hardware root of trust in my 5G network?

Since the platform integrity measurements on the servers used for the 5G system are performed by HRoT, they must have an implementation of HRoT on them. Many technologies and vendors implement HRoT, with some examples listed in Section 3.2 of NIST IR 8320 [1].

The network operator needs to verify that their servers have these HRoT capabilities and that they are enabled, as well as having secure storage with cryptographic functions such as TPM on each server for the measurements. The network operator needs to install and enable RAS in the 5G system to ensure the individual platform measurements are collected at each boot and maintained throughout the server's lifecycle. Also, the network operator's 5G CNF orchestrator needs to use platform trust statuses, as determined by the RAS, as part of its compute node selection when deploying instances of CNFs.

What else should I know about hardware-based security?

An HRoT can be leveraged as a starting point that is implicitly trusted. Hardware-enabled controls can provide a foundation for establishing platform integrity assurances. Combining these functions with a means of producing verifiable evidence that these integrity controls are in place and have been executed successfully is the basis of creating a trusted platform.

Platforms that secure their underlying firmware and its configuration provide the opportunity to extend trust higher in the stack. Verified platform firmware can, in turn, verify the OS boot loader, which can then verify other software components all the way up to the OS itself and the hypervisor or container runtime layers. The transitive trust described in NIST IR 8320 is consistent with the concept of the chain of trust (CoT)—where each software module in a system boot process is required to measure the next module before transitioning control. HRoT can be further extended to be used by container orchestrators to ensure that whenever a container instance is instantiated, it is placed on a compute host that has been proven to maintain its platform integrity – in other words, is trusted.

HRoTs for CNF orchestration will not prevent attacks against the compute servers; however, they can detect and prevent the 5G CNFs from running on a compute node if it is compromised. This capability will provide visibility into the lower levels of the computing infrastructure hosting 5G CNFs. Additional hardware-enabled security technologies not discussed in this paper, such as confidential computing and Trusted Execution Environments, can also be leveraged to protect data in use within shared compute environments [1].

Standards developing organizations and 5G-relevant cybersecurity guidance documents describe and recommend the use of these capabilities to protect 5G infrastructures. The European Telecommunications Standards Institute (ETSI) has released over 20 specifications and reports with specific guidance on various aspects of Network Function Virtualization Security. ETSI GR NFV-SEC 007 V1.1.1 (2017-10) describes and recommends the use of these advanced capabilities. The Enduring Security Framework (ESF), a public-private partnership addressing risks to critical infrastructure and National Security Systems, has also recommended the use of these advanced security capabilities in part 2 and part 4 of their series of papers titled Security Guidance for 5G Cloud Infrastructures.

The NCCoE 5G Cybersecurity project followed this ETSI and ESF guidance within its environment and implemented these advanced security capabilities within the 5G core of a functional commercial-grade 5G network. This working system is an example of how to enable these capabilities in 5G infrastructure, and it provides a blueprint that can assist other implementations. The deployed 5G core network technology stack utilizes mainstream container orchestration software, Kubernetes, which by default has integration capability with RAS. The HRoT and 5G CNF orchestration capabilities were enabled without any 5G vendor product source code or hardware modifications. The enablement of these capabilities only required trust agents and supporting libraries to be installed and configured on each server, and policies to be created within the 5G CNF scheduler.

Summary

As part of the ongoing digital and communications convergence trend, specialized telecommunications hardware is increasingly being replaced by cloud-native network functions that run on commodity servers. 3GPP specifications have significantly improved the architecture and the security posture of 5G systems, but they do not specifically address the underlying computing infrastructure, leaving a potential cybersecurity gap.

This paper provides an example of how introducing hardware-enabled security capabilities helps to fill that gap and mitigate infrastructure threat vectors, which supports a holistic approach to cybersecurity of communication systems. As the expectation for the next generation of communications systems continues the shift towards commodity hardware and cloud-native applications, these HRoT technologies can be applied to those systems as well.

Additional Technical Details

The rest of this white paper is intended for readers seeking more in-depth knowledge of the hardware-enabled security capabilities, particularly HRoT, for measuring the platform, and extending them for asset tagging, RAS, and CNF orchestration functionality. The technical information presented here is implemented and observed in the NCCoE 5G Cybersecurity testbed.

For background information on the NCCoE 5G Cybersecurity project, including the architecture and components of the 5G standalone network built within the demonstration lab environment, see NIST SP 1800-33 Volume B, 5G Cybersecurity, Approach, Architecture, and Security Characteristics [4].

Measuring the Platform

To ensure 5G infrastructure platform integrity, each server that hosts 5G CNFs uses an HRoT to perform a measured launch at boot time. In the NCCoE 5G lab, the HRoT is initialized in the server's central processing unit (CPU). Figure 2 shows the high-level flow for taking the measurements and writing them to the TPM to establish the CoT.

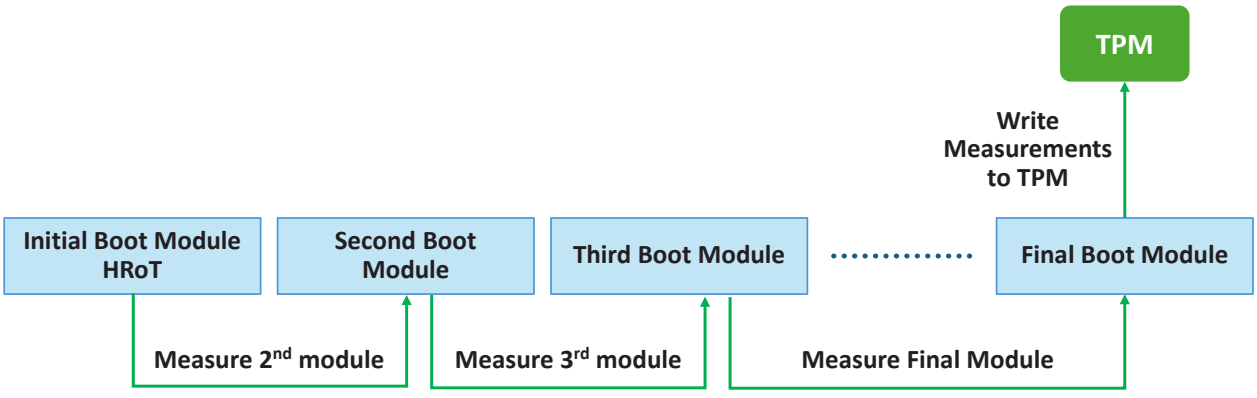


Figure 2. High-level flow for taking and writing measurements

1. The server is powered on and the HRoT begins measuring the first module in the boot process.
2. The HRoT performs a cryptographic measurement, or hash, of the next module in the boot process. This is an important distinction because cryptographic hashes are one-way functions, so their output cannot be guessed and they are commonly used to verify integrity of data.
3. Once the second module in the boot sequence starts, it takes a cryptographic hash of the third module in the boot process before that module can start.
4. This process continues all the way through the entire boot process so that each module is cryptographically measured by the previous one before booting.
5. The result is a chain of cryptographic measurements for each individual module in the boot process, so that it will become apparent if any of the modules have been changed into an unallowed configuration, either inadvertently or maliciously.

The composition of these measurements together is what makes up the trust status of the 5G server platform. Once the trust measurements have been completed, it is critical to store them securely so they cannot be modified. Therefore, a tamper-resistant hardware security module (HSM) should be used. The NCCoE deployment leverages the TPM as its HSM; these TPMs are already installed by server OEMs. Figure 3 shows an example of how the status of the platform integrity measurements is displayed in the NCCoE 5G testbed.

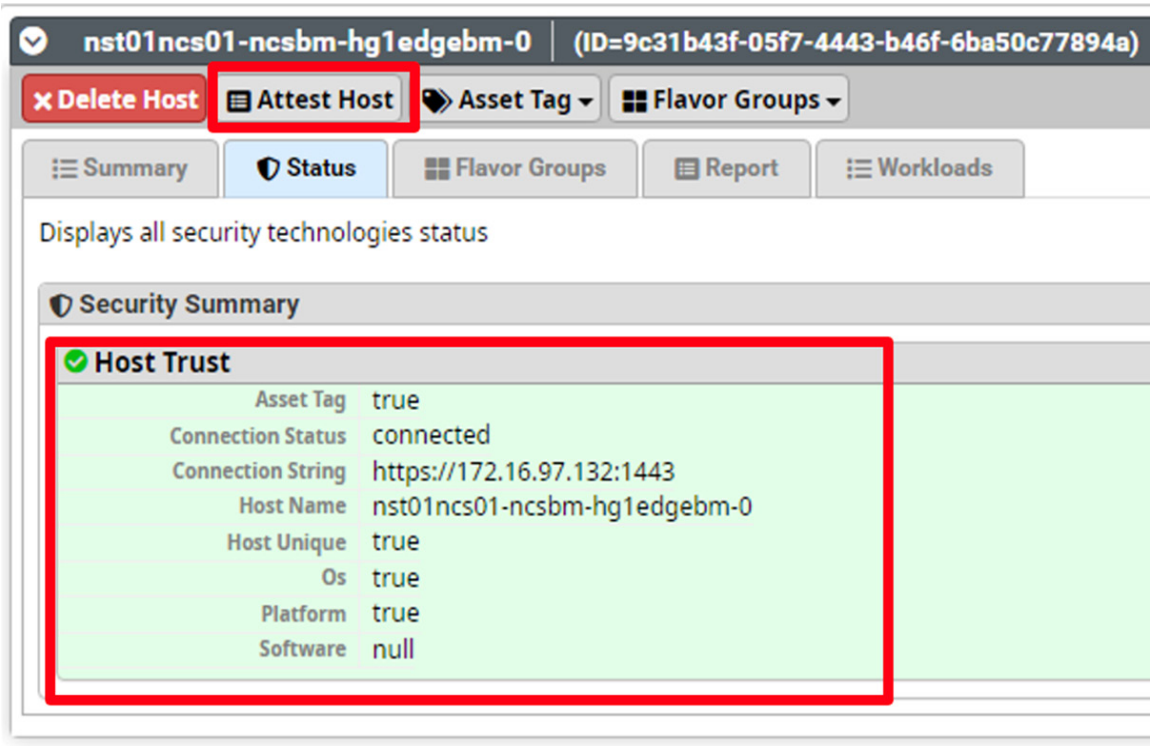


Figure 3. Example of the displayed status of platform integrity measurements implemented in the NCCoE 5G Cybersecurity testbed

Uniquely Identifying 5G Servers with Hardware Asset Tags

The TPM secure storage capabilities can also be leveraged to store unique identifiers for the 5G servers. These identifiers are based on simple key-value pairs that are used as asset tags. The asset tags can be used to describe a multitude of attributes about an individual or group of servers. For example, they can describe things like geolocation, intended type of workload, or intended vendor.

In order to be meaningful, these asset tags should be provisioned from a source other than the server itself. Hence, there is a process in place for the RAS to create the asset tags and push them to the 5G servers via a trust agent on the 5G server that writes to the TPM. This has two benefits: it ensures that the asset tags are coming from a server responsible for managing the environment, and that the asset tags are logically located with the servers' trust measurements. With the trust measurements and asset tags paired, the RAS has a technical mechanism to query the trust status and unique identifiers for each 5G server in the environment. Figure 4 shows the logical workflow for how the RAS creates and pushes asset tags into servers.

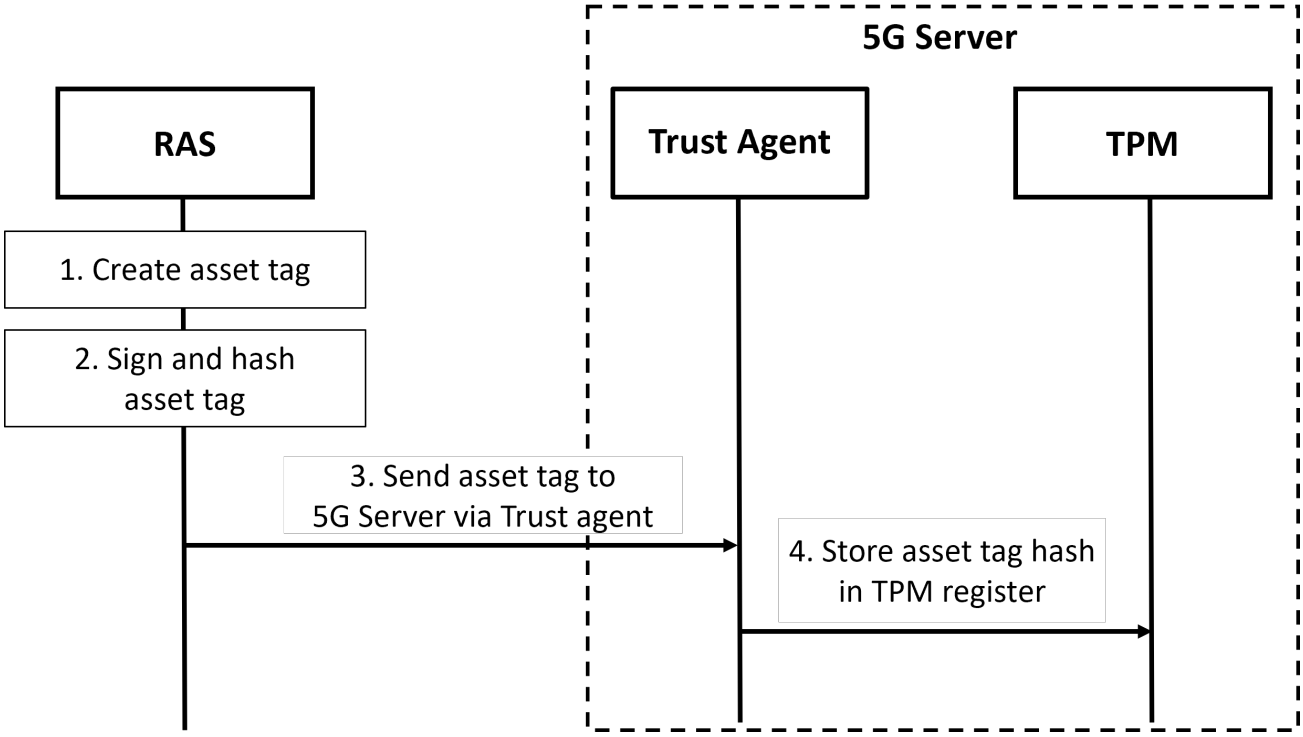


Figure 4. Logical workflow for RAS creating and pushing asset tags to servers

For example, if the 5G Access and Mobility Function (AMF) CNFs are only allowed to run on a subset of servers in the 5G core, a key-value pair in JavaScript Object Notation (JSON) format for these servers such as "{'NodeType':'Worker'}" could be stored in their TPM. Figure 5 shows an example of how the hardware asset tags are displayed in the NCCoE 5G Lab, with Rule 2 specifying the node type as a worker.

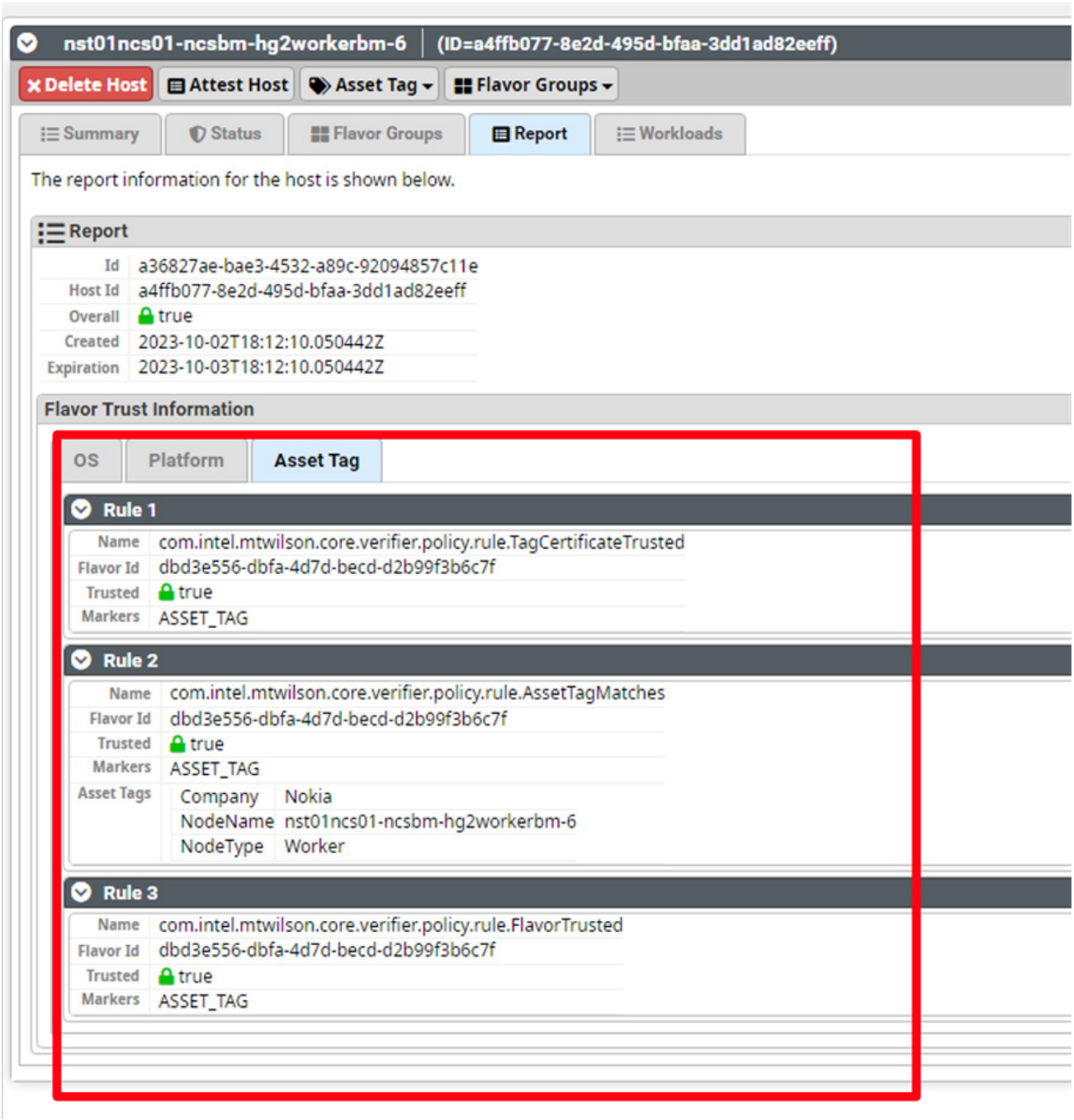


Figure 5. Example of hardware asset tags

Getting the Measurement Ready to Use

Now that there is a technical mechanism to measure trust values for the servers in the 5G environment, to make them useful there must a be a method to read them from the server. A method called remote attestation provides the mechanism for 5G servers to send their trust values to a RAS, which can collect and aggregate the trust measurements from all of the servers in the 5G environment. The Internet Engineering Task Force (IETF) has documented a general remote attestation procedure architecture for networked systems, which is applicable in the 5G environment [5]. This allows for greater visibility of firmware, BIOS, and OS versions installed on systems, as well as creating allowed lists for which versions are acceptable.

The allowed lists are created by knowing which versions of platform modules will be on 5G server platforms and precomputing the entire CoT. Mobile network operators can then create trusted sets of servers to run specific CNFs based on allowed lists of trusted values. When the 5G servers go through their boot process and send their trust measurements to the RAS, the RAS can compare them to the allowed list values. If the trust measurements match, the server can be marked as trusted; however, if the measurements do not match an allowed value, the server can be marked as not trusted. Note that measurements might not match for a number of reasons, not all indicating tampering – for example, a legitimate firmware update that was not included in a new trusted measurement. In this case, proper determination of the attestation status would require forensic analysis of these failures. Figure 6 illustrates how each 5G server sends its platform integrity measurements to the RAS and has its trust status evaluated.

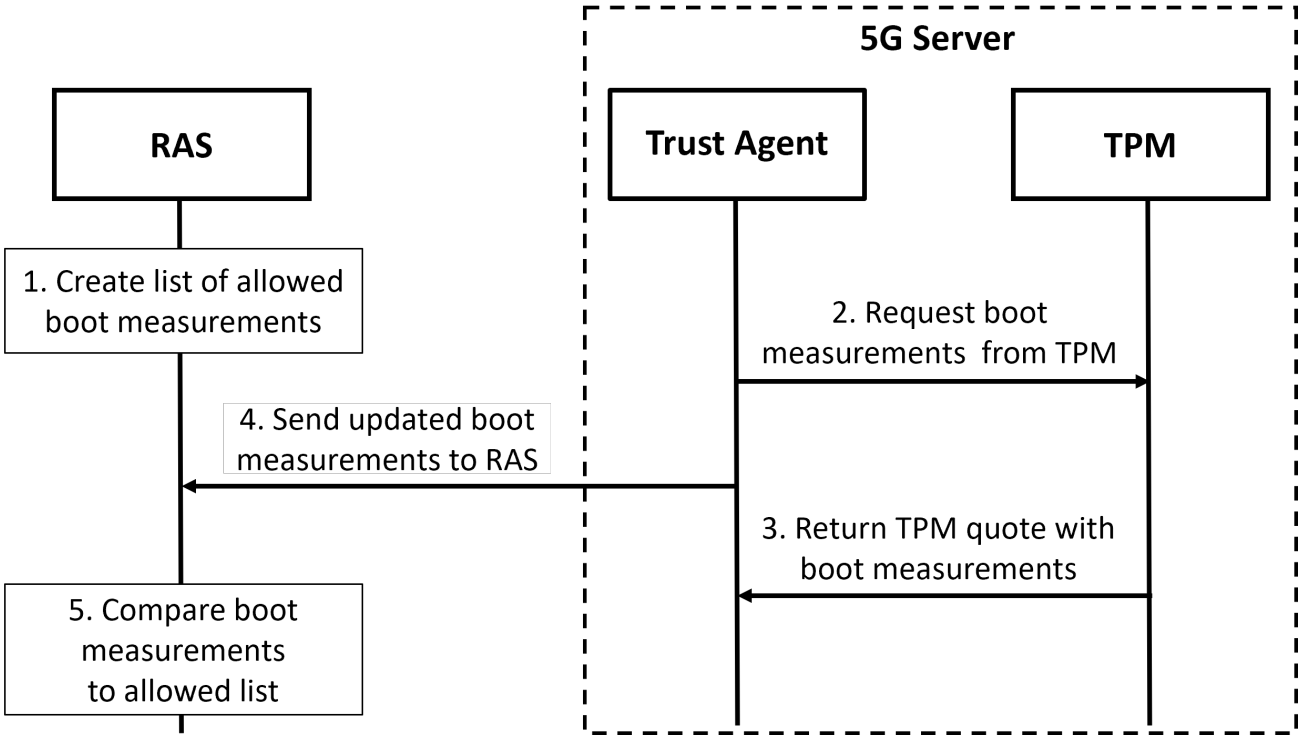


Figure 6. Logical workflow for evaluating platform integrity measurements

Figure 7 shows how remote attestation presents the trust status of all 5G servers in the NCCoE 5G Lab, and how an individual server displays its trust status when selected.

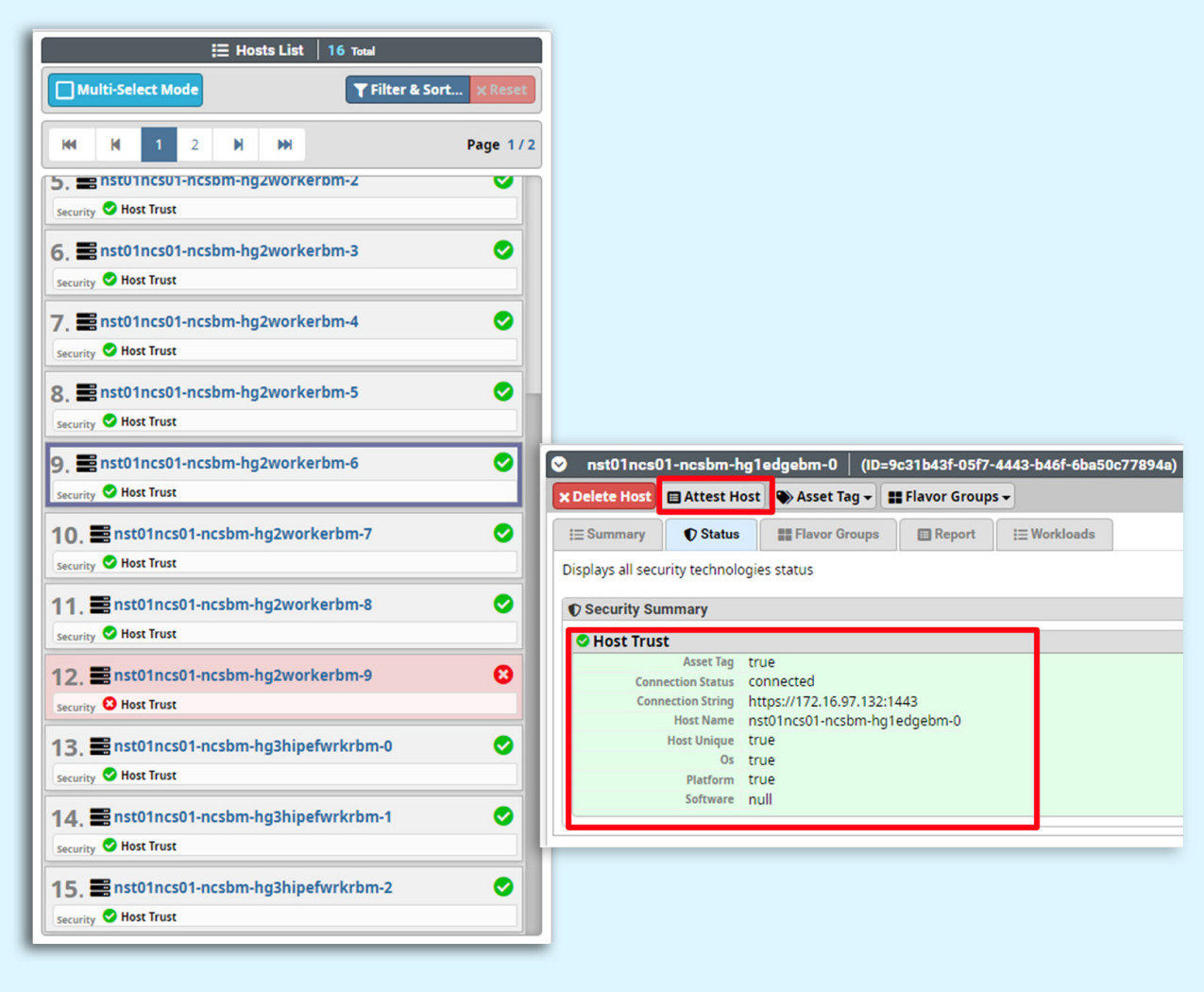


Figure 7. Example of trust status display

Using the Platform Measurement

CNF orchestration within the 5G environment is the process of finding a suitable compute server to place a workload, and once one is found, instantiating the workload on it. Typical orchestration engines look for performance metrics in their algorithms, such as CPU usage, amount of free memory, affinity rules, and workload type to find suitable compute servers. However, with the trust values and asset tags stored in the RAS, these additional two factors can also be used to place 5G CNFs. For example, when deploying an AMF CNF image, the criteria for workload placement can be that the target 5G server's trust values must be marked as trusted and have the "{'NodeType':'Worker'}" asset tag.

With the integration between the remote attestation server and 5G CNF scheduler, the placement of workloads on servers in a known-good state and configured for a specific purpose can be technically enforced. Figure 8 shows the workflow for remote attestation, asset tag provisioning, and workload placement.

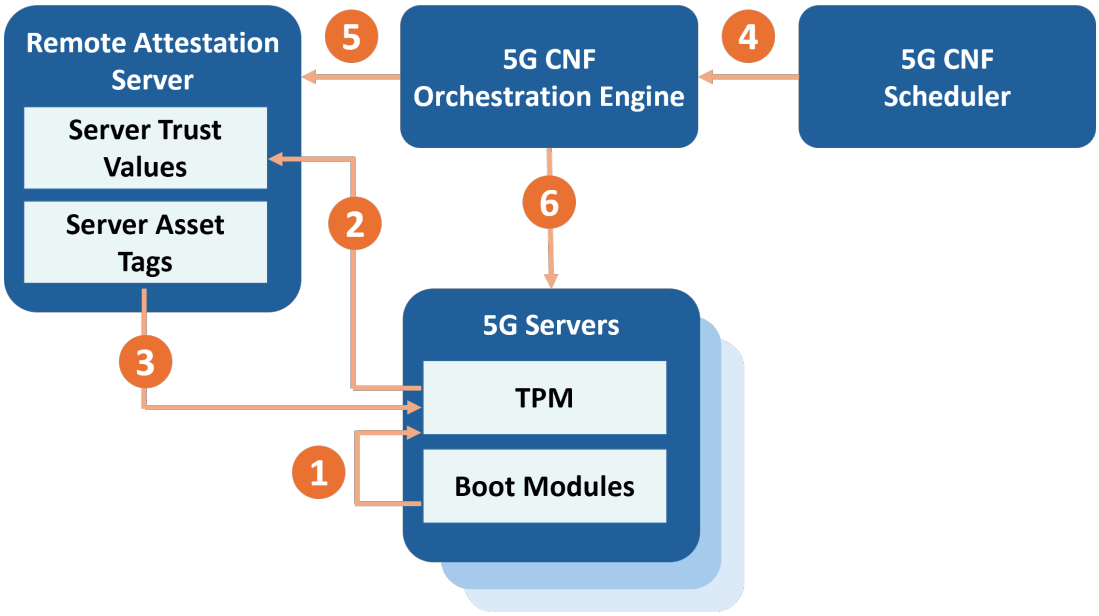


Figure 8. Logical workflow for remote attestation, asset tag provisioning, and workload placement

The following are descriptors for the process flow steps shown in Figure 8:

1. CoT is created from the HRoT, and trust values are stored in the TPM on 5G compute servers.
2. The 5G compute servers push their most recent trust values to the RAS.
3. The RAS pushes asset tags to the 5G compute servers for storage in their TPMs.
4. The 5G CNF scheduler requests deployment of a CNF image.
5. The 5G CNF orchestration engine queries the RAS to find a set of 5G compute servers marked as trusted with applicable asset tag.
6. The 5G CNF scheduler deploys an instance of the requested CNF to a compute server that matches the trust and asset tag policy.

The 5G core implementation in the NCCoE lab uses Nokia Container Services (NCS), which is based on open-source Kubernetes packages. Utilizing HRoT measurements for CNF placement was implemented using built-in Kubernetes scheduler functions so that no modifications had to be made to NCS. Out-of-the-box Kubernetes deployments allow the creation and use of Custom Resource Definitions (CRDs), which may extend the default Kubernetes scheduling policies. Custom CRDs were created in the NCCoE 5G Cybersecurity testbed, which defined specific CNF images that must be run on remotely attested compute servers in a trusted state, along with a specific hardware asset tag. The text below provides an example of a CRD that was created to ensure that the 5G Unified Data Management (UDM) CNF meets this requirement.

```
[root@nst01ncs01-ncsbm-masterbm-2 5g_trusted_workload]# cat
udm_nim1_trusted.yaml
        serviceType: UDMNIM
        vnfType: REGSTR
    spec:
      affinity:
        nodeAffinity:
          requiredDuringSchedulingIgnoredDuringExecution:
            nodeSelectorTerms:
            - matchExpressions:
              - key: isecl.trusted
                operator: In
                values:
                - 'true'
              - key: isecl.TAG_Company
                operator: In
                values:
                - Nokia
```

How Mobile Operators Can Enable HRoT for CNF Orchestration

NIST SP 1800-33B defines four infrastructure security capabilities that, when all are implemented, culminate in the ability to use HRoT for CNF orchestration. These capabilities essentially implement the following:

1. Use compute servers that implement HRoT – see appendices of NIST IR 8320 for an incomplete list of examples
   a. Ensure that HRoT technologies are turned on and configured appropriately.
   b. Measure platform integrity for each server in the infrastructure using hardware-enabled controls.
2. Assign specific asset-tags (labels) for each server in the infrastructure using hardware-enabled controls, which are stored in the HRoT.
3. Perform remote platform attestation for each compute server's trust measurements and asset tags against policies, and allow workload orchestrators to access these findings so the results can be used as factors in workload placement.
4. Configure the workload orchestrator to deploy and migrate CNFs only to servers that match specified platform measurements and labels.

References

[1] Bartock MJ, Souppaya MP, Savino R, Knoll T, Shetty U, Cherfaoui M, Yeluri R, Malhotra A, Banks D, Jordan M, Pendarakis D, Rao JR, Romness P, Scarfone KA (2022) Hardware-Enabled Security: Enabling a Layered Approach to Platform Security for Cloud and Edge Computing Use Cases. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Interagency or Internal Report (IR) 8320. https://doi.org/10.6028/NIST.IR.8320

[2] Wired (2010) Russia's Elite Hackers Have a Clever New Trick That's Very Hard to Fix. Available at https://www.wired.com/story/fancy-bear-hackers-uefi-rootkit/

[3] Trusted Computing Group (2025) TPM 2.0 Library (Trusted Computing Group). Available at https://trustedcomputinggroup.org/resource/tpm-library-specification/

[4] Bartock M, Cichonski J, Souppaya M, Dey S, Grayeli P, Mulugeta B, Sharma S, Teague C, Scarfone K, Righi S, Ramalingam M, Rhea P, Santharam M, Mosley R, Ungureanu B, Patel J, Wan T, Romness P, Hyatt M, Lebel L, Carroll D, Orrin S, Brown L, Zhou Y, Piggott C, Jones M, Yeh M, Atkinson G, Eustace D, Wenger B, Morgan S, Balmakhtar M, Schumacher G (2022) 5G Cybersecurity. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Special Publication (SP) 1800-33B. Available at https://www.nccoe.nist.gov/sites/default/files/2022-04/nist-5G-sp1800-33b-preliminary-draft.pdf

[5] IETF Datatracker (2023) RFC 9334 (IETF). Available at https://datatracker.ietf.org/doc/rfc9334/

Certain commercial equipment, instruments, software, or materials, commercial or non-commercial, are identified in this paper in order to specify the experimental procedure adequately. Such identification does not imply recommendation or endorsement of any product or service by NIST, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

NIST Technical Series Policies

Copyright, Use, and Licensing Statements

NIST Technical Series Publication Identifier Syntax

Author ORCID iDs

Michael Bartock: 0000-0003-0875-4555

Jeffrey Cichonski: 0009-0006-1137-2549

Karen Kent: 0000-0001-6334-9486

Murugiah Souppaya: 0000-0002-8055-8527

How to Cite this NIST Technical Series Publication:

Bartock M, Cichonski J, Souppaya M, Kent K, Grayeli P, Sharma S. (2026) Using Hardware-Enabled Security to Ensure 5G System Platform Integrity: Applying 5G Cybersecurity and Privacy Capabilities. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Cybersecurity White Paper (CSWP) NIST CSWP 36B. https://doi.org/10.6028/NIST.CSWP.36B

Contact Information

5g-security@nist.gov

National Institute of Standards and Technology
Attn: Applied Cybersecurity Division, Information Technology Laboratory
100 Bureau Drive (Mail Stop 2000)
Gaithersburg, MD 20899-2000

Additional Information

Additional information about this publication is available at https://www.nccoe.nist.gov/5g-cybersecurity, including related content, potential updates, and document history.

All comments are subject to release under the Freedom of Information Act (FOIA).

NIST NATIONAL CYBERSECURITY CENTER OF EXCELLENCE