REFERENCE

TCG Trusted Network Communications for Mobile Platforms

Version 1.0
Revision 28
26 September 2018
PUBLISHED

Contact: admin@trustedcomputinggroup.org

TCG PUBLISHED

TCG

Disclaimers, Notices, and License Terms

Acknowledgements

Kathleen McGill, John Hopkins University APL

Charles Schmidt, The MITRE Corporation

Ira McDonald, High North Inc.

Carlin Covey, NXP Semiconductor

André Rein, Huawei

Wael Ibrahim, Amex

Bo Bjerrum, Intel

Table of Contents

1. Introduction

Increasingly, mobile devices have become the primary computer for many people in their day-to-day lives. However, these mobile devices face many of the same threats plaguing traditional endpoints, like personal laptops, and their software is subject to similar types of vulnerabilities that malicious parties can exploit. These vulnerabilities are generally understood in the context of the security and privacy of individual users' personal data. However, mobile devices also leverage common networks and other services in their day-to-day use and thereby expose these common services to attack. This reality poses a challenge to mobile network and service providers, who need to protect operations for their large user base from compromised devices.

In order for network and service providers to offer the desired convenience without sacrificing the security and privacy of their users, those providers require some assurance that mobile devices are healthy before they permit access to their services. This requires not only a way to measure the health of mobile devices, but a way to convey these measurements to authorized parties. These requirements present new challenges to mobile devices and their management.

The Trusted Computing Group's (TCG) Mobile Platform Work Group (MPWG) and Trusted Network Communications (TNC) Work Group define and promote standards-based technologies to address these challenges. Today, the management of mobile devices entails collecting mobile-specific information through proprietary, non-extensible means. Through the adoption of standardized, extensible mechanisms, the TCG enables a mobile device environment that is more secure and interoperable than what is possible today. This paper looks at some examples of challenges in better securing the mobile environment and a few specific TCG technologies that can help address these challenges.

1.1 Mobile Platform Architecture

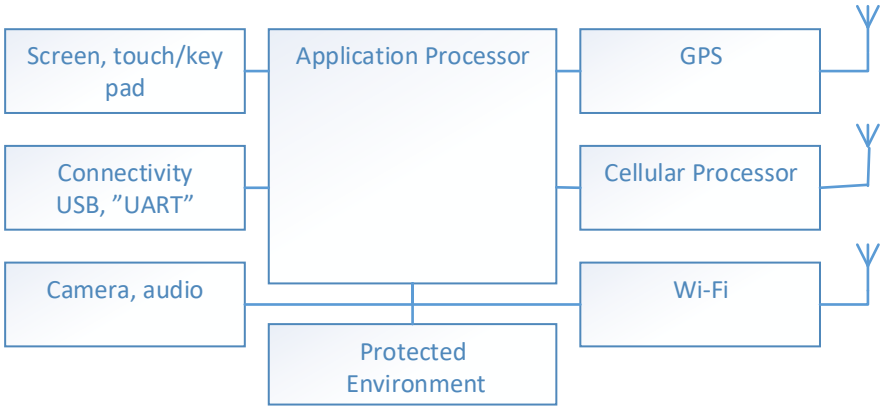

Figure 1 Generic Mobile Device Hardware Architecture

Figure 1 shows a generic mobile device hardware architecture. Mobile platform architectures are centered around the Application Processor (AP), which hosts the main operating system. The AP core handles all interaction with the user (screen, touchpad, biometrics), connectivity (wired data transfers, charging), and image processing. To support the AP there is a set of Service Cores. The Cellular Processor (CP), which runs the radio

protocol stack, is the main service core for a mobile device. GPS and Wi-Fi are other important Service Cores. The Protected Environment is a critical component of the platform architecture that protects different services and their data across the platform. These protected services include data integrity, confidentiality, and cryptography. The Protected Environment can also host a Trusted Platform Module (TPM).

Products can vary in the types and computational capacity of the Services Cores they include. In low cost products, many software components may execute directly on the AP rather than on separate cores. However, the Protected Environment is always implemented as an isolated environment with hardware support for resource separation.

The MPWG has published multiple specifications focused on improving the trustworthiness of mobile devices. In particular, the MPWG has advanced four important concepts for securing mobile devices:

- Mobile Roots of Trust: A Root of Trust (RoT) is a component that performs one or more security-specific functions, such as measurement, storage, reporting, verification, and/or update. It is trusted always to behave in the expected manner. A RoT is the foundation of a trusted computing device and designed to be resistant to practical hardware and software attacks. Higher level device applications that provide modern day convenience can be verified in a trust chain that is ultimately based on the security of these RoTs.
- Secure Boot: Secure Boot [2] is a process in which every software image is validated before execution. Specifically, the first software module loaded is verified by the RoT. That module then verifies the next loaded software module and so on, forming a transitive chain of verification tracing back to the RoT. Secure Boot is the de-facto standard boot protocol for mobile devices.
- Measured Boot: In Measured Boot [2], as each code module is loaded it is measured and then the measurements are securely stored in or protected by a TPM. Once in the TPM, those measurements can be securely attested to other parties, as defined by the TPM 2.0 Library Specification [3].
- Protected Environment: A Protected Environment is a functional element that has execution and memory resources that are isolated from other components of a mobile device in order to host sensitive applications that require security and privacy properties [2]. The Mobile Reference Architecture includes an informative example of a Protected Environment as an implementation of the Global Platform Trusted Execution Environment (TEE) [4]. It is not necessary that the Protected Environment comply with Global Platform standards, but a Global Platform TEE which includes the integrity protections and anti-rollback features should be able to satisfy the requirements of a Protected Environment.

Together, these capabilities are critical to the security of mobile devices because they support important security functions, including device integrity, attestation, isolation, and confidentiality.

1.2 TCG's Trusted Network Communication Architecture

TNC is an open architecture to support network access control and security automation. TNC standards integrate security tools across endpoints, networks, and servers into an intelligent, responsive, coordinated defense.

The TNC Architecture [1] offers three primary capabilities:

- a Compliance capability, which evaluates an endpoint's adherence to network policy both at the time of connection and on an ongoing basis while it is connected to the network;
- an Orchestration capability, which provides a dynamic repository and notification service for real-time state and events; and
- an Access Control capability, which controls access to protected resources and networks based on endpoint posture and other factors.

These capabilities can be used for many purposes, including security automation, continuous monitoring, asset management, endpoint compliance assessment and enforcement, protection of critical resources, leveraging of shared information, event correlation and assessment, and a variety of other functions. Application of these capabilities enables trusted network communications - the ability to understand the trustworthiness of an endpoint before and while it is allowed to communicate on the network.

The TNC Architecture provides an extensible base for the collection and distribution of information about endpoints within a network. The core TNC protocols standardize mechanisms for secure communications between authorized parties, and support routing of information to agents capable of servicing a given request or handling a particular type of collected data. This framework is designed to be extensible to allow new collection and data processing agents to be added easily. Implementers can create new data collectors and have them immediately make use of the underlying TNC protocols. This means that, if one needs some new type of endpoint information collected, one can purchase or build a TNC-compliant collector and have that collector immediately make use of the TNC transport mechanisms. This gives enterprises complete, vendor-agnostic control over what data they are collecting from managed endpoints. Similarly, new tools that consume and process collected endpoint information can also be easily added to a TNC deployment, giving enterprises flexibility and control over how collected information is used or exposed to other tools.

In summary, TNC provides an extensible, standards-based framework that allows enterprises to control and manage the information collected from endpoints. Enterprise managers can combine TNC-compliant data collectors from different vendors, and even create collectors of their own, to ensure that they can gather the information they need. The underlying TNC protocols ensure that this collected data is delivered securely to authorized parties. The collected data can then be put to a range of uses, including compliance assessment, access control decisions, and exposure to other authorized tools. The result is a powerful tool for enterprise management, situational awareness, and security.

2. Use Case: Device Health Reporting in Mobile Networks

When a mobile device attaches to a cellular base station, the base station establishes (or refreshes) a device-specific Security Context for that mobile device within a Security Anchor function in the mobile network infrastructure. The Security Context is a record that identifies and tracks security characteristics of a mobile device that is connected to a mobile network. The Security Anchor is a functional element of the back-end mobile network that creates, updates, and manages these Security Contexts. This Security Context is transferred to a new base station every time there is a handoff, and occasionally may also be transferred to a new Security Anchor. This is the basis of the mobile network's determination of its level of trust in the mobile device.

2.1 Challenge

Currently the established Security Context does not include any platform health information about the mobile device. Platform health could include information about whether the operating system is patched, whether certain security services are enabled, or whether certain apps known to be sources of spam or malware are present. This level of assessment is fairly straightforward on traditional, managed computing devices. Unfortunately, current cellular network technologies do not support that level of granular assessment of attached mobile devices. Secure integration of the mobile device's platform health information into its Security Context would allow more fine-grained decision making about the mobile network resources and functional services that the mobile device is allowed to access. For example, a device with known spam apps might be given reduced non-emergency data bandwidth to reduce its ability send spam to other devices.

3. Use Case: Device Recovery of Configuration

Mobile devices contain a variety of configuration data that is critical to their correct operation. For example, all mobile devices contain calibration files that control their use of the licensed spectrum space. Corruption of these files could cause incorrect calibration and result in degraded operation for users and even violation of laws. While there are measures in place to maintain the integrity of these sensitive configuration files, deliberate or accidental corruption is always possible. Mobile devices require secure mechanisms to restore the integrity of this configuration data for proper operation.

3.1 Challenge

There are two challenges in this use case. The first is similar to the Device Health Reporting Use Case. For regular operation, the mobile device may need to provide evidence of the integrity of sensitive data to certain authorized parties. For example, a mobile device may need to demonstrate that its calibration files are not corrupted as a condition of access to services. To do this, the device must have a means to measure the state of data such that the authorized party can verify the authenticity and integrity of the measurement and of the sender.

In addition, if the device requires “resetting” of its sensitive data (e.g., if its calibration files have been corrupted), it needs to be able to verify the authenticity of the “clean” data source and the integrity of the “clean” data. Given the severity of the consequences should this process be subverted, the controls and “proof” offered by the providing party need to be strong enough to withstand concerted efforts to compromise them.

4. Use Case: Streaming Services for Mobile Devices

Increasingly, mobile devices are the means by which users consume streaming services. (E.g., Netflix, HBO, etc.) As many of the providers of these services operate on a paid subscription model, providers of streaming services wish to ensure that content is only provided to authorized devices. (I.e., devices whose owners have paid for the content.) This is done by provisioning authorized devices with a Device Specific Key (DSK). The DSK is a software token stored on the device and its provisioning generally occurs when the device is manufactured. When the user purchases content, the DSK is registered with the content provider. Only devices in possession of the device specific key can authenticate to the service provider's license server. Later, this DSK is used as the cryptographic basis of a content key, issued by the license server, which can be used to secure transactions to download media.

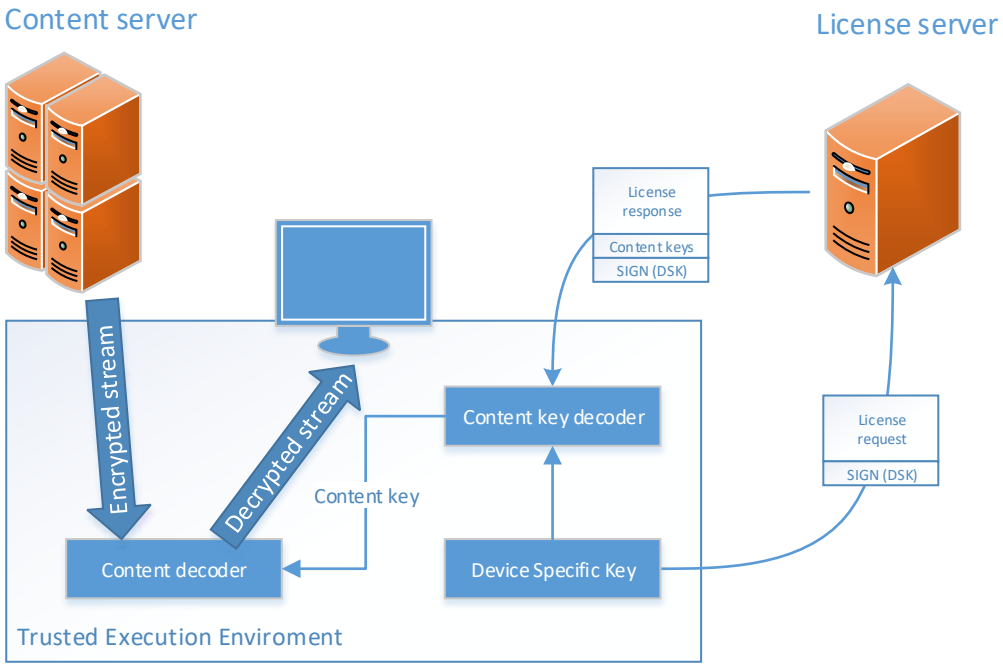

Figure 2 Provisioning Mobile Devices for Streaming Services

4.1 Challenge

The biggest concern for content providers is the security of the DSK, which is the root of trust for the exchange of content keys. As the DSK is only a software token, it is possible for copies of the DSK to be created that could then be shared with unauthorized devices, giving those devices access to the paid user's content.

To address this threat, content providers would like to have some assurance that the device to which they give a content key is healthy, and that safeguards to protect the confidentiality of its DSK are intact. Currently, content providers base this judgement on the general architecture of the mobile device. A better solution would be to have a trusted assessment of the device, including both its architecture and relevant software components, delivered to the service provider to determine the device's eligibility to receive a content key. At the same time, this assessment capability needs to be controlled to ensure that user privacy is not violated by this assessment.

5. Solutions

Over the past few years, TCG has developed a wide range of standards that help address challenging security problems. In particular, standards produced by TCG's TNC and MPWG have a direct bearing on the aforementioned challenges facing mobile devices and networks. The following sections provide an overview of these standards and technologies and describe how they can be applied to mobile network challenges.

5.1 Trusted Network Communications

TNC describes a set of architectural components that can be composed into a flexible, extensible framework supporting measurement of endpoints and delivery of those measurements to a central server. Once delivered to the server, these measurements can drive a wide range of security activities, including but not limited to, determining the level of network access granted to a device, triggering remediation actions for non-compliant devices, archiving for long-term trend analysis, and managed exposure to other security tools that can perform additional analyses.

5.1.1 Components Supported by TNC

An overview of the possible components supported in the TNC architecture appears in Figure 3:

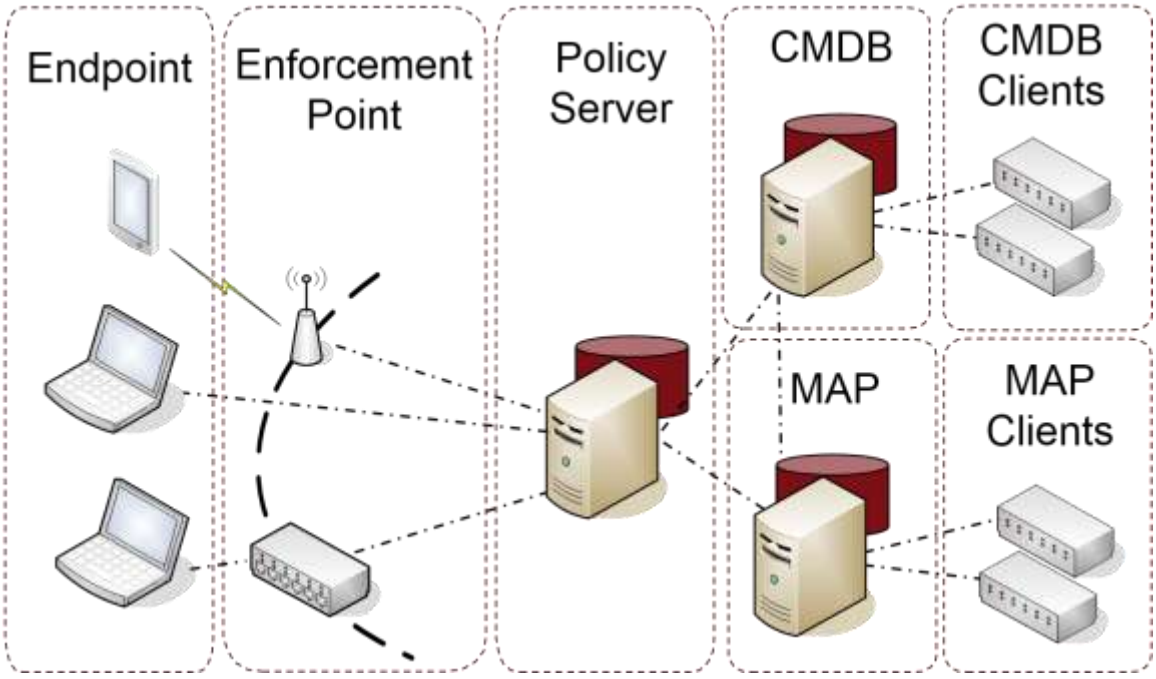

Figure 3 The Trusted Network Communication Architecture

At the far left are endpoints whose state is being measured. This information is sent to the policy server in the middle of the graphic. From there, the policy server can:

1) Send instructions to network Enforcement Points, changing the endpoint's network access

2) Store the information in a Configuration Management Database (CMDB) to support long-term trending and analytics
3) Share collected information with enterprise management and analysis tools via the Metadata Access Point (MAP), which maintains real-time information and supports event-driven notifications

All of this is done over encrypted, integrity-protected network connections between mutually authenticated communicants. In short, TNC supports secure collection of endpoint measurements and controlled exposure of this information to authorized parties to support a range of security and network management use cases. Enterprises can support as many or as few components as they desire – only the policy server and endpoints are strictly necessary, and the other components can be used or ignored depending on the needs of the enterprise. However, should the enterprise wish to add one of these components at a later time, the TNC standards ensure that this can be accomplished easily.

5.1.2 Extensible Collection Mechanisms

Two key components of TNC are the Integrity Measurement Collectors (IMCs) and Integrity Measurement Validators (IMVs). IMCs are located on endpoints and IMVs are located on the policy server. The former manages the collection and sending of endpoint measurements while the latter can query IMCs for these measurements, receive the results, and perform initial processing on measurements.

The TNC architecture is designed to support the easy addition or replacement of IMC-IMV pairs. This means that, should an enterprise need to collect a new or different type of measurement from endpoints, a new IMC-IMV pair can be purchased or custom built to perform this task. These IMC/IMVs only need to concern themselves with data collection and structuring, leaving the details of transport, encryption, and authentication to other layers of the TNC architecture. This makes it simple to expand the measurements collected from endpoints and have them delivered to where they are needed, all within the same infrastructure.

5.1.3 TNC on Mobile Networks

Today, TNC is primarily employed in traditional networking environments. However, the capabilities TNC provides align well to the challenges facing mobile environments.

With regard to mobile device assessment, TNC could facilitate fine-grained measurement and evaluation of mobile device health when connecting to a cellular network. Access could be allocated based on what measurements, if any, a device provided.

Similarly, TNC can easily support device recovery. Should mobile device measurements indicate that any aspect of the device's core configuration is corrupted or out of compliance, the policy server can push remediation instructions back to the device. The fact that these communications channels are encrypted and integrity protected, and that all participants are authenticated and known, minimizes the chance of this channel being hijacked for malicious purposes.

Finally, TNC also can facilitate more secure streaming services. The state of DSKs and the security features that protect those keys can be monitored by IMCs and reported when the device seeks a new content key. This allows content providers not only to verify that the device possesses the requisite key, but that the device is protecting the key against unauthorized duplication.

5.2 Trusted Mobile Devices

The TPM 2.0 Mobile Reference Architecture Specification [2] describes an architecture that enables trustworthy mobile devices through the construction of a secure environment from both shared and dedicated resources. This architecture enables a TPM executed within a Protected Environment on mobile devices. The specification includes examples of several implementation approaches that provide capabilities to support the use cases presented in this paper.

5.2.1 Protected Environment

A Protected Environment is a functional element that has execution and memory resources that are isolated from other components of a mobile device. The Protected Environment provides a secure environment in which sensitive applications can execute and can store sensitive data. These sensitive applications and data are protected from malware and other untrusted software in the device's richer execution environment. There are many possible implementations of a Protected Environment. In most cases, a Global Platform certified TEE would be an example of a Protected Environment.

The Protected Environment relies on sound mobile RoTs. Through secure boot, the RoTs extend a chain of trust to the Protected Environment, enforcing the integrity of its code and data. Once established, the Protected Environment preserves integrity of sensitive applications executing within it. In addition, the Protected Environment protects the integrity, confidentiality, and availability, as appropriate, of data stored within it through a highly restrictive interface to the environment, access controls, and encryption.

The Mobile Reference Architecture specifies a set of requirements that a Protected Environment must satisfy. These requirements define the necessary boot sequences and integrity protections for the code and data of the Protected Environment itself, for isolation of execution resources, and for protection of non-volatile storage. These requirements ensure that applications hosted by the Protected Environment have security and privacy guarantees. Further, the Mobile Reference Architecture provides informative examples, with device architecture and boot protocols that are representative of mobile devices at the time of publication. These examples provide insight on how the requirements can be realized in devices without dedicated hardware roots of trust.

5.2.2 Measurement

In TCG, measurement refers to the collection of evidence of a device's state. The TCG has defined and matured the concepts of device measurement and provided specifications to ensure the integrity and confidentiality protection of those measurements. The MPWG has further adapted those concepts to mobile devices. For example, in measured boot, the code and data of a mobile device's boot cycle is measured and stored securely. The measurements obtained in measured boot can serve as evidence as to the health of the device.

5.2.3 Applying the Mobile Reference Architecture

These building blocks of trusted mobile platforms provide a solid foundation for solutions to many mobile device security challenges.

For device health reporting, measurement is the means by which evidence is collected for reporting. Measurements collected during measured boot can provide the starting point for this evidence. Further, the Protected Environment provides secure execution resources for measurement agents (such as an IMC) to operate, and provides secure storage for the measurements and any keys or cryptographic identities. The latter could be used to prove the device's identity or authenticate remote parties, such as during a TNC exchange.

Similarly, device configuration recovery requires the ability to measure sensitive data and to protect and authenticate those measurements for submission to authorized parties. The Protected Environment provides secure storage for the sensitive data to reduce the likelihood of corruption. Finally, particularly in this case in which there is a remote, privileged stakeholder, the Protected Environment provides a safe environment for the remote stakeholder to deploy a trusted application on the device, if appropriate. This might entail authentication of messages from the remote stakeholder using cryptographic identity on the device, or even a local execution agent to control updates.

Finally, these concepts support mobile secure streaming by providing secure storage within the Protected Environment to protect the DSK and by providing measurement capabilities (through measured boot and the protection of measurement tools such as IMCs) to support attestation of the device health and the ability to protect sensitive data to the streaming service provider.

6. Conclusion

These use cases represent just a few of the challenges in the field of mobile communication and computing to which TCG technologies can be applied. TCG continues to develop and expand on these technologies to better align them with the unique constraints and needs of the mobile environment. Parties interested in these solutions are invited to join the TCG so they can contribute their expertise and perspectives to the development of these solutions.

The TCG encourages product developers and service providers to consider the ways in which TCG standards can contribute to the securing of their products and services. TNC, Mobile, and other freely available TCG standards not only provide requirements that lead to more secure interactions, but also better interoperability between products and servers. Through the application of trusted computing to the mobile communications domain, we can all enjoy more robust, reliable, and secure mobile computing.

7. References

[1] Trusted Computing Group, *TNC Architecture for Interoperability*, Version 2.0, October 2017

[2] Trusted Computing Group, *TPM 2.0 Mobile Reference Architecture*, Rev 142, December 2014

[3] Trusted Computing Group, *Trusted Platform Module Library, Part 1: Architecture*, Rev 1.38, September 2016

[4] Global Platform, *TEE System Architecture*, Version 1.1, January 2017