mode51 Software

Using Hardware Security Modules in Mobile Networks

10/02/2021

1. Introduction

This document discusses the integration of Hardware Security Modules (HSMs) in key Mobile Network infrastructure including the Mobile Core Network.

2. Telecoms Industry Evolution

Few core network vendors have used HSMs even as recently as for 4G core networks. Beyond the fundamental technical benefits such as the protection of keys, the tamper proofing of critical algorithms and the presence of a True Random Number Generator (TRNG), in 2021 there are a number of converging factors that significantly enhance the business case for the inclusion of HSMs:

2.1. Government Regulation in the UK and Germany

The need to specify requirements for a high baseline level of security for telecoms equipment vendors has been established in late 2020. The details of what this means are currently being developed.

In the UK, this government article dated 24th November 2020 indicates:

- *New legal duties on telecoms firms to increase the security of entire UK network*
- *Fines up to ten per cent of turnover or £100,000 a day for failing to meet standards*

In Germany, according to this article dated 16th December 2020:

- *Companies will be required to submit a "guarantee" that contains details on how they ensure that components of critical systems can't be misused for illegal purposes*
- *A vendor that fails to meet the threshold for trustworthiness can be banned from operating equipment.*

2.2. Virtualization Including the Integration of Legacy Stacks

NFV has been around for several years. Large incumbent vendors are using virtualization as part of their overall 5G core rollout. This also includes the option to migrate legacy core network functions on dedicated legacy equipment into new container based architectures. This will be deployed for 5G SA (Stand Alone), whereas 5G NSA (Non Stand Alone) is a stop gap that still uses the existing 4G core.

This integration of the 2G, 3G and 4G functions into the virtualized stacks as well as the new 5G functions provides an opportunity to also integrate older algorithms such as COMP-128 implementations inside an HSM. Milenage has been used for 3G, 4G and 5G though TUAK is also now a good option because the eSIM spec mandates the presence of this newer algorithm. Using an HSM to provide these algorithms provides a portable method that essentially outsources the security and responsibility of the sensitive processing.

Virtualization introduces two new problems that are relevant to the Authentication Vector (AV) generation that are solved by the use of HSMs:

1. **Poor entropy** - virtualized nodes suffer from low entropy (material needed for random number generation) because the underlying source is shared by all the instances running on top of it. HSMs include an approved True Random Number Generator (TRNG) that can both provide sufficient entropy and additionally, because the randomness is *True* as opposed to *Pseudo* (PRNG), operate as a significantly higher quality source of entropy.

2. **Automated Replication Exposes Sensitive Keys and Credentials** - automatically scaling up computing resources is both fantastically powerful and also a security problem. If the OP keys are stored in a file on the HSS server then multiplying the instances also multiplies the number of copies of the vulnerable keystore. Storing the keys in an HSM mitigates this. Additionally there are platforms such as HashiCorp's Vault that can also be used to protect eg. database credentials used by application servers and the revocation of SSH credentials when support personnel leave.

It is likely that the new Telecoms regulation will require that these side effects of virtualization are addressed with appropriate security measures, ie. the use of HSMs.

2.3. Cloud Based HSM Services

Thales's SafeNet Luna HSM is now available as a cloud based service, called *Data Protection on Demand*.

Entrust's nShield HSM is also available as a cloud based service.

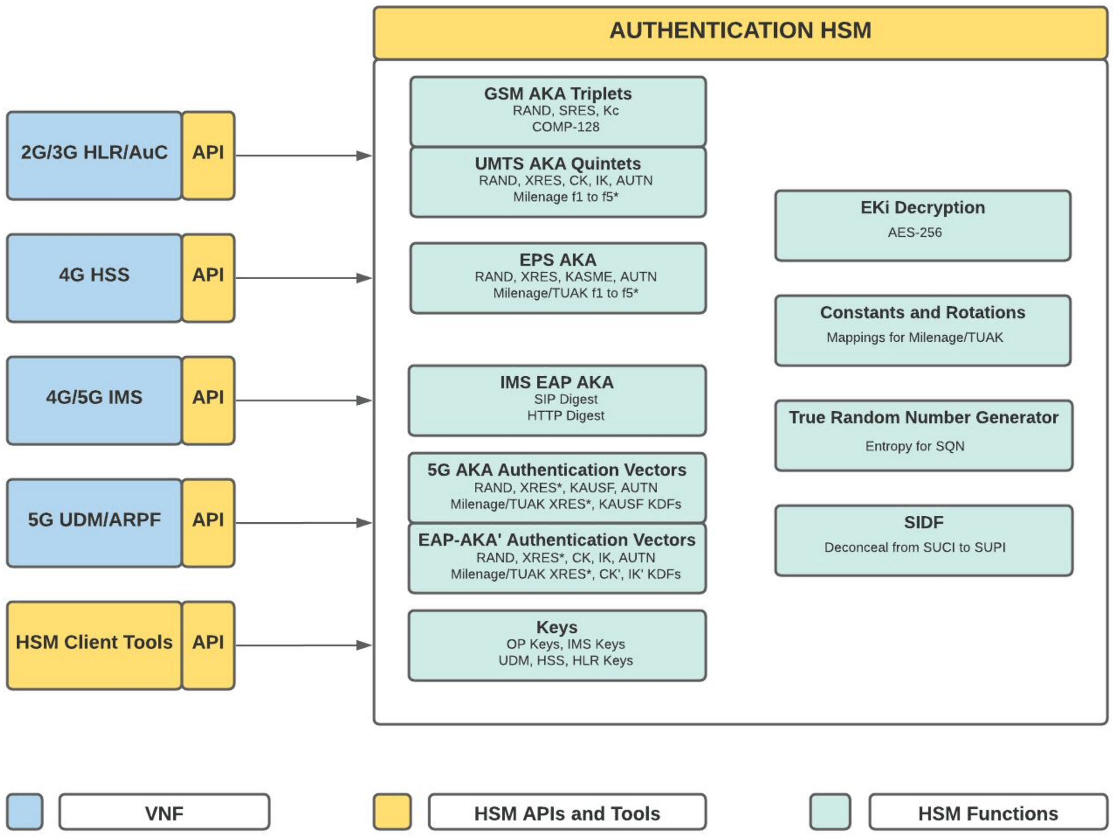
In these examples the vendor will be responsible for the maintenance and upgrade of the hardware seamlessly and the physical HSMs will be hosted in their own data centres.

For smaller networks with lower load requirements these cloud based instances may be ideal. However, there will always be a latency cost incurred by connecting from the HSS into a remote data centre, and as the performance of the HSS is critical, it is likely that physical HSMs located in close physical proximity to the HSS machines will be needed where there is any significant load. Some customers may also require that the core network elements are in a data centre of their choosing, in which case an on premises HSM deployment will be needed.

3. Use Cases for Deploying HSMs

3.1. Using HSMs in the Mobile Core Network for the UDM/ARPF/HSS/HLR/AuC

3.1.1. Generation of Authentication Vectors



The HSM can be used to:

- Store OP keys (Operator Variant Algorithm Configured Field) used to generate the OPc needed for SIM provisioning and the generation of authentication vectors for the UDM/ARPF/HSS/HLR/AuC

- Store EKi keys used to encrypt and decrypt the Ki, so that the UDR/HSS/HLR will only deal with the Ki in it's encrypted form and never hold it in memory in the clear

- Provide entropy for sequence number generation, though the SQN will still be managed by the UDM/HSS/HLR

The UDR/HSS will:

- Not store OP keys

- Only store and handle the EKi, without the keys needed to decrypt back to a Ki

- Not be responsible for the Milenage or TUAK algorithms or SUCI de-concealment

- Still be responsible for the Sequence Number (SQN) management

There are a number of different algorithms and functions that can be managed by the HSM as well as the storage of the OP keys. This also means that the logic is protected from tampering;

- UMTS, EPS, 5G AKA and EAP-AKA' all use Milenage. The 256 bit version can also be supported as well as the newer TUAK algorithm. The presence of the latter is now mandatory for eSIMs though Milenage can still be used

- IMS AKA includes digest functions that can be run by the HSM

- 5G AKA and EAP’ AKA include a number of key derivation functions that can also be run by the HSM

- GSM AKA and COMP-128 derivations. Potentially a build harness could be provided for this so that any MNO specific customisation can be kept secret, ie. inserted and built by the customer

3.1.2. True Random Number Generation

HSMs include a True Random Number Generator (TRNG). Normal machines usually include a Pseudo Random Number Generator (PRNG) which produces lower quality entropy. The availability of entropy directly affects the quality of the cryptographic operations. Lower entropy results in more predictable random numbers.

Virtualisation also dilutes the finite amount of entropy available between multiple virtual instances.

It is extremely likely that regulation will require the use of TRNGs, particularly for virtualized environments.

3.1.3. UDR/HSS EKi Database Field Encryption

Including an HSM means that it also makes sense to protect the storage of the sensitive Ki in the UDR/HSS. It can be stored in encrypted form and handled in such a way that it is never decrypted externally to the HSM.

Commonly it is stored as an EKi in the HSS along with the kdbId to identify the key. Rather than decrypting in the HSS’s memory though, instead the EKi and kdbId can be submitted to the HSM during the AV generation procedure which can both decrypt it and make use of it without needing to return and reveal it.

3.1.4. Constants and Rotations

The Milenage algorithm supports further customisation for the constants c1 to c5 and the rotations r1 to r5. Commonly operators have used default values, as these also need to be inserted into the SIM profile at the point of manufacture. With the advent of eSIM it is possible that operators may begin to make use of this additional layer of security on a per subscriber level.

Sets of constants and rotations could be inserted into the HSM and the HSS could store the designated set identifier per subscriber without having visibility of the actual values.

3.1.5. Subscription Identifier De-concealing Function (SIDF)

In 5G the Subscription Identifier De-concealing Function (SIDF) is part of the UDM and de-conceals the SUPI from the SUCI.

3.2. Using HSMs During SIM Provisioning

The HSM can be used to:

- Generate new Ki values using a TRNG (True Random Number Generator)
- Store OP keys (Operator Variant Algorithm Configured Field) used to generate the OPc needed for SIM provisioning
- Store keys used to encrypt and decrypt the Ki, as the EKi will be provisioned into the HSS database. The Ki will only be used within the HSM's memory and won't be returned to the HSS

Legacy systems that produce data for physical SIMs (ie. pre-eSIM) will commonly use a keystore on the local filesystem. Using an HSM to store the OP keys and run Milenage obviates any risks surrounding access to the SIM data generation machine from IT support staff that may be transient.

The newer RSP (Remote SIM Provisioning) standards mandate the use of HSMs but it is likely that legacy systems run by MNOs will continue to run for several years alongside the emergence of RSP's provisioning mechanisms.

3.3. Using HSMs for SIM OTA

SIM OTA (GSM 03.48) is the process of sending Remote File Management (RFM) and Remote Application Management (RAM) commands to SIM profiles. This can be used to change files for eg. roaming steering, forbidden lists, and the deployment of applets.

Originally these payloads were sent as concatenated SMS with a different class to standard messages. Modern SIMs (and eSIMs) may use a push to wake mechanism where an SMS push message provides an indication that a payload is available to download and then the data can be downloaded over HTTPS.

SIM OTA messages contain a command packet header featuring a Redundancy Check, a Cryptographic Checksum or a Digital Signature.

Proprietary algorithms may be used to produce a derived key, the KiC. This may be formed, eg. from the ICCID (SIM profile serial number). This algorithm could be placed within the HSM.

Note that SIM OTA is still in use with modern RSP (Remote SIM Provisioning).

3.4. Using HSMs in Remote SIM Provisioning (RSP)

RSP is a relatively recent set of specifications that facilitate the deployment of SIM profiles to eSIMs. The eUICC is a secure element installed on a device that can contain multiple SIM profiles, as opposed to the traditional slot for a physical SIM card with one resident profile. An RSP server, external to any single MNO's network, can request SIM profiles and package them up for delivery to eUICCs.

The use of HSMs as part of this process is mandated in the specs.

4. Performance Benefits of Customised HSMs

4.1. Custom Firmware Functions vs Direct Algorithms

The advantages of custom functions implemented inside the HSM are:

- Performance benefits derived from running a multi-step process in a single request to the HSM rather than sending multiple requests from the client. For example, the steps could be:
    - Decrypt EKi
    - Regenerate OPc once for sharing with the two functions
    - Run Milenage's f1 and f2345 functions reusing the OPc
    - Or run the resynchronization variants, f1* and f5*
- Flexibility for the management of constants and rotations
- Seamless management of the connectivity expressed as straightforward functions/methods that can be called directly from client code, built on top of HSM vendors' own client libraries that implement eg. their specific value added load distribution features

Some HSM vendors are beginning to provide Milenage support. In the case of SafeNet PKCS#11 can be used with two new mechanism types, CKM_MILENAGE_SIGN for f1, f1* and f2, and CKM_MILENAGE_DERIVE for f3, f4, f5 and f5*. The Ki needs to be stored as a concrete key inside the HSM which may be cumbersome to manage for many millions of subscribers. The client needs to make multiple requests to first run f1 and f2, then subsequently to run f3, f4 and f5.

Performing the Authentication Vector generation in multiple parts to the HSM is always going to incur a higher performance cost compared to the single request approach of the custom functions.

5. Automating Secrets Management with HashiCorp's Vault

Vault provides automation for the management of common secrets in backend infrastructure such as database credentials, API keys, ssh user credentials and TLS certificates.

The need for this type of management is amplified by the automatic scaling that is part of virtualization, as embedded secrets will be replicated into multiple instances.

Vault introduces the concept of *dynamic secrets*, where a secret is very short lived and can even be produced on demand. This increases the security of the overall system by minimising the attack surface available to attackers.

The automatic renewal of TLS certificates is also a key benefit of Vault. Think of Ericsson's expensive outage with O2 where a hard coded TLS certificate expired and brought the house down.

mode51 Software is working with HashiCorp to add HSM based CA signing to Vault.

6. Glossary

AKA	Authentication and Key Agreement
ARPF	Authentication credential Repository and Processing Function in 5G
AuC	Authentication Centre produces the triplets in 2G AKA
AV	Authentication Vectors, also termed triplets in 2G AKA, quintets in 3G AKA
COMP-128	2G AKA algorithm(s)
EKi	Encrypted Ki
HLR	Home Location Register subscriber database used in 2G
HSM	Hardware Security Module
HSS	Home Subscriber Server replaces the HLR/AuC in 4G
kdbId	Key index ID indicating which key has been used for the EKi
Ki	In 2G the Ki is the subscriber key. This is renamed to the K in 3G onwards. All instances of *Ki* in this document refer to the *K*
Milenage	UMTS AKA algorithm also used for EPS, 5G and EAP' AKA, successor to COMP-128
MNO	Mobile Network Operator
NFV	Network Function Virtualization
OP	Operator Variant Configuration Field, essentially a key
OPc	Computed result of passing Ki and OP into the Milenage function
OTA	Over The Air. SIM OTA is a provisioning mechanism for updating files and applets on SIM cards and now also in eSIM Profiles
PRNG	Pseudo Random Number Generator

TRNG	True Random Number Generator
TUAK	Successor to the Milenage algorithm. Now mandatory for eSIMs as an option alongside Milenage, ie. the eUICC contains both algorithms, whereas older physical SIMs will most likely only contain Milenage
UDM	Unified Data Manager (5G version of the 4G HSS)
UDR	Unified Data Repository as an optionally split component from the UDM housing the subscriber database