Challenges and Solutions for Industry-Grade Secure Connectivity

Harald Klaus
T-Labs (Research & Innovation)
Deutsche Telekom AG
Berlin, Germany
Harald.Klaus@telekom.de

Felicitas Hetzelt
Security in Telecommunications
Technical University of Berlin
Berlin, Germany
file@sect.tu-berlin.de

Peter Hofmann
Telekom Security
T-Systems International GmbH
Berlin, Germany
P.Hofmann@t-systems.com

Andreas Blecker
Telekom Security
T-Systems International GmbH
Frankfurt am Main, Germany
Andreas.Blecker@t-systems.com

Daniela Schwaiger
T-Labs (Research & Innovation)
Deutsche Telekom AG
Berlin, Germany
Daniela.Schwaiger@telekom.de

***Abstract* — This contribution discusses challenges and solutions for secure communication that are relevant for the success of the Industry 4.0 paradigm. It shows that IT security has to be solved with respect to the technical challenges of hardware or software components to be used in the Industry 4.0 context. Aditionally, the paper highlights the organizational challenge to reduce security risks to a minimum during the complete industrial life cycle.**

Keywords — security, secure connectivity, secure connectivity gateway, secure boot, secure, TCOS

I. Motivation

The fourth industrial revolution (I4.0) enables a simpler networked production and maintenance process through a coherent communication architecture, which has an enormous impact on the economy and efficiency of the entire production chain. In principle, Industry 4.0 implies cross-company networking at all levels of traditional production. Traditional boundaries between individual production systems become less important, whereas flexibility in the design and arrangement of component, process or system groups are subject to the dynamic of value-added networks.

IT security risks, however, arise from the extensive introduction of networking technologies into the production process in the context of Industry 4.0, leading to a wealth of new and highly complex security requirements. The number of possible weaknesses is increasing because, apart from attacks enabled by physical access, remote attacks have also to be considered. For example, remote maintenance facilitates the maintenance and control of systems enormously. At the same time, this advantage requires a higher level of security in order to prevent system manipulation. Since threats cannot be eliminated entirely, the goal must be to reduce them to an acceptable level. Security techniques are continuously developed, but the attack tactics are becoming more and more sophisticated. The objective must be to improve the degree of security as far as possible so that attacks are adequately countered without completely shutting down the system, and successful attacks become so complicated and expensive that they are also uninteresting for criminal or terrorist actors. Furthermore, it is important to consider the interoperability. The required level of security must be achieved with the components of different manufacturers and must not prevent vertical or horizontal interoperability.

Within the project "Industrial Communications for Factories" (IC4F), a mature consortium of industrial players, academic research institutes together with SME's work together to develop appropriate solutions for Industrie 4.0 challenges and demonstrate the results with selected use cases, which are described in more details in [1].

II. The Integration Challenge

Industrial manufacturing plants usually are, compared to IT or mobile devices, long-term investments with an endurance of many years or decades. Many factories still fulfil their intended production purposes, although the technological and economical ambience changed revolutionary due to the Industry 4.0 paradigm. This require comprehensive secure connectivity and networking abilities of all industrial devices that would allow to benefit from the advantages of new connectivity services like:

- Monitoring of live production data, machine parameters and deterioriation gradients
- Remote access to production devices for flexible re-configuration, maintenance and repair
- Monitoring and documentation of quality-relevant information
- Predictive maintenance processes
- Simulation and optimization of production processes derived from realistic live data

The list of new services is large, and the economic opportunities are enormous, if the integration challenge is solved. Therefore, it is necessary to solve another central challenge in order to protect machine-specific data and parameters during the long lifetime of industrial machines:

III. The Industry-Grade Certificate Exchange Challenge

In large organizations with many unknown partners, trust is one of the most important factors for successful business and technical co-operation. Human societies developed certificate mechanisms and management processes for personal identities, producing reliable processes and documents to ensure trust, such as identity cards, notarial acts, document sealing, etc.

978-1-7281-0568-0/19/$31.00

In technical or industrial domains, all elements in the communication chain have to be trusted in order to ensure trusted secure communication, i.e. for every data source, each data transmission facility including the communication method as well as the entire used encryption and decryption algorithms being used, a commonly accepted identity or certificate should be assigned. Otherwise, neither production data could be reliably acquired nor security policies be enforced. Furthermore, automatic control and monitoring of device interactions become rather complicated or impossible, if large amounts of sensors and actors work on the shop floor.

IV. Connectivity Services based on Secure Gateways

Within the IC4F project, a simple but challenging approach is realized: In order to connect existing production plants, each machine or device will be equipped with a secure gateway that establishes a secure connectivity path.

The secure gateway

- Collects machine data and production parameters and enables remote connectivity
- Provides a unique ID using a TCOS[1] chip-set
- Enables a standardized communication interface using the OPC unified architecture protocol

The software architecture of this additional low-cost device is shown in Fig. 1:

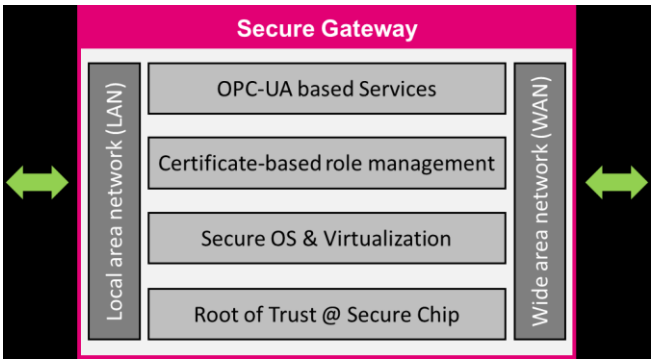

Figure 1: Software architecture of the secure gateway.

In order to fulfil essential requirements to withstand security threats, the secure gateway has been developed by a team of security experts of Telekom Innovation Laboratories, of Telekom Security of T-Systems and the TU Berlin chair "security in telecommunications".

A. *Threat Models*

To motivate the current design of the secure gateway we first introduce the relevant threat models. To provide protection against remote as well as local attackers which will have physical access to the remote gateway module, two **threat models** were considered:

Threat model for remote attacks: A remote attacker is able to intercept any connection between a valid user and the secure gateway and modify or omit any data sent via this connection. Note that connections between the secure gateway and the industrial component are wired and can therefore be considered secure regarding remote attack vectors only. A valid user is any party, which has obtained possession of a valid client certificate (cf. Section V) and uses this certificate to authenticate itself towards the secure gateway.

Threat model for local attacks: A local attacker is capable of physically accessing and therefore modifying the secure gateway device. Modifications in this scenario include the removal and replacement of the attached TCOS card (see below) or the installation medium. The physical security of the industrial component as well as the connection between the secure gateway and the industrial component has to be taken into consideration during onsite integration, but is considered out of scope for this paper.

In the next step, we describe the **design** of the remote gateway and motivate important design decisions with respect to our threat models in detail. The security of the gateway is based on multiple and to some extend orthogonal security features. In order to establish any connection with the secure gateway module and thereby the industrial component, a bidirectional authentication of the client as well as the server has to be performed. The client in this case is a mobile end user using a smartphone or tablet that is equipped with a valid authentication token. The client connects to the secure gateway server, which is also provisioned with a unique authentication token.

B. *Telekom Card Operating System (TCOS)*

The secure gateway offers for the server-sided certificate to be stored using the Telekom Card Operating System (TCOS) which is mounted directly on the secure gateway hardware. The customer can protect the TCOS module with a PIN stored in an encrypted region of the gateway's file-system and thereby bind the TCOS module to the specific gateway device.

Once the client connects to the secure gateway module it asks the module to identify itself based on its unique authentication token. If the gateway module is able to provide a valid authentication token to the client, the client proceeds to authenticate itself towards the secure gateway module using its own authentication token.

C. *High Assurance Boot (HAB) Mechanism*

The Secure Gateway prototype is built using an NXP evaluation board based on an NXP Freescale ARM processor featuring a High Assurance Boot (HAB) mechanism [3-5] that has been developed to mitigate attacks based on physical tampering. A local attacker who is capable of breaching the casing and obtain physical access to the secure gateway would be able to remove the TCOS smart-card component. HAB is preventing that the stolen component can be misused to impersonate a valid server, or manipulate the installation medium.

To that end, HAB ensures that only software signed by a trusted party can be executed on the device. Therefore, even if the gateway or TCOS module is stolen from the factory, it cannot be used towards malicious intents, if doing so requires the modification of the software running on the device. The HAB features are integrated into the SOC design and can therefore not be physically removed from the Secure Gateway. Note that these features enable the platform to

[1] TCOS = **T**elesec **C**hipcard **O**perating **S**ystem [2].

securely store an encrypted PIN which is required to unlock the TCOS module thereby mitigate the threat of theft of the module itself.

D. Evaluation

In order to evaluate the security design against the threat models described in Section B, we first analyze the threat emerging from a remote party.

A remote party is able to intercept and modify any communication exchanged between client and server. An attacker might attempt to fool the client into exposing sensitive data with respect to the industrial component, by intercepting the connection and posing as the secure gateway module. In this case, the attack is mitigated by the bidirectional authentication mechanism which ensures that the client will only maintain the connection if the server authentication token received is valid. Similarly, the server will only allow persisting connections with clients that are able to provide a valid authentication token. Further, once bilateral trust has been established, all communication is encrypted following a standard encryption protocol depending on the specific application level protocol (HTTP, OPC-UA …). An attacker is therefore unable to ex-filtrate sensitive data by eavesdropping on the connection.

Note that even if the attacker is able to gain remote access to the secure gateway module, e.g. by exploiting a local service, she will not be able to ex-filtrate the local secrets in order to impersonate the remote gateway module, since the secrets are only accessible to isolated components. An attacker with remote access to the gateway module would however be able to eavesdrop on various incoming and outgoing communication channels, depending on the already obtained privileges and the specific makeup of the attached industrial component. This threat should be mitigated by limiting the privileges of a potential remote attacker by isolating remotely reachable software components and ensuring the continuous integrity of the secure gateway module.

E. Platform Integrity and Secure Firmware Update

As mentioned in the previous section the security of the secure gateway verifies and therefore relies on the integrity of the software and data. In this section, we will detail the integrity protection mechanisms in place and evaluate their potential to mitigate attacks according to the thread models described above.

The software and data components we aim to protect are

(1) the boot loader which is the first dynamic software component responsible for loading the operating system;

(2) the operating system kernel along with;

(3) the kernel device tree which describes the hardware configuration of the board;

(4) the initial ram file system,

(5) the root file system,

(6) user applications and

(7) the application data.

The boot loader integrity is verified though a signed hash of the boot-loader binary. The hash is signed with a private key which is only accessible to privileged parties. The validity of the signature is verified through a public key appended to the boot loader binary, whereas the validity of the public key itself is verified by comparing it to a hash value written to one time programmable (OTP) memory on the secure gateway module. It is important to note that the initial verification code is also loaded from OTP memory and can therefore not be modified. In addition to verifying the boot loader's integrity, we also encrypt the boot loader binary that enables us to use it as a safe storage space for additional key material.

The boot loader and any keys stored within the boot loader binary can now be trusted and will in turn validate the integrity of the operating system kernel, the kernel device tree and the initial ram file system in the same manner as described above.

Due to the storage size of the gateway's root file system we have to rely on an external storage medium (instead of RAM) and therefore use block level encryption and integrity verification mechanisms provided by the underlying linux operating system (dm-crypt [6] and dm-verity [7]). We use the verified initial ram-disk image to verify, decrypt and finally mount the root file system partition. The encryption key is derived from a master key which is unique to each secure gateway device and is stored as part of the encrypted boot-loader. Since file system integrity verification requires a read-only partition, volatile user applications and data are stored in separate partitions and not integrity protected. Mandatory access control policies stored in the integrity protected regions of the file system in addition to standard file system permissions ensure that the integrity of the whole module cannot be compromised by introducing unverified components.

The described mechanisms provide the user of the secure gateway module with a strong assurance regarding the device integrity right after the system has been started. To ensure system integrity during run-time the secure gateway relies on established mechanisms. These include the containerization of different services running on the gateway as well as enforcement of mandatory access controls. In combination with the security mechanisms described in the previous sections the secure gateway is able to assert system integrity throughout the operation of the secure gateway given the described threat model.

The secure firmware update mechanism builds upon the signature verification feature described in the previous paragraphs. End users will be given access to custom firmware signing keys that are to be used to sign the components of the new firmware image. These keys will correspond to the public key hashes fused into the OTP of the respective secure gateway models. The users must use the provided signing keys to generate valid update images, the validity of which will be verified first by the verified update software in order to prevent incapacitating the gateway and later by the secure boot firmware when the new image is loaded.

V. Certificate Management and Public Key Infrastructures

A. Certificate Management by Public Key Infrastructures

In the past, a PKI (Public Key Infrastructure) has proven successful as a classic centralistic management approach for the organizational technical mapping to management. Similar to the ID card system of a country, devices are identified, registered, and receive a digital identity in form of a certificate as a result of a registration process. In order to ensure that this process for issuing certificates is not arbitrary, the PKI is operating as technical vicarious agent and is subject to a clearly defined organizational, processual guided set of rules.

This set of rules mainly comprises the Certificate Policy (CP), Certification Practice Statement (CPS) of the PKI service - safety regulations as well as guidelines regarding technical and organizational aspects - and describe the Trust Center activities using the roles of a Certification Authority (CA) and a Registration Authority (RA).

Within one PKI mainly two different roles will be active within the running core processes:

Registration Authority (RA): The Registration Authority receives Certificate Signing Request (CSR) and verifies them content-related according to underlying requirements. Furthermore, the RA constitutes the responsible authority within the PKI process, which ensures the applicant's identity by means of identity verification. This is the only way to guarantee that a unique mapping of issued certificate and applicant is ensured.

Certification Authority (CA): The Certification Authority (CA) constitutes the Trust Anchor for all underlying issued certificates. Certificate requests previously approved by RA are signed by the CA using private keys and are turned into certificate themselves. Through this certification, all issued certificates by CA become a member in their "confidence interval".

B. PKI Trust Model and Design Approaches

Through "Third-party (PKI)" integration the PKI trust model enables different actors within a PKI architecture, whose position of trust is initially unknown between them, to establish a position of trust among each other. To this extent, certificate holders use the joint position of trust towards the issuing PKI for proof of identity.

Within PKI, design centralistic arranged PKI structures with up to two levels underneath a common PKI-ROOT-CA are considered as still controllable. The higher the degree of complexity within PKI architecture, the more complex and extensive the PKI processes turn out to be. In particular, lengthy certificate chains comprising multiple hierarchically structured CA systems contribute to higher expenses within certificate check, provided that the certificate check is made over the whole certificate chain up to the ROOT-CA.

Similarly, complexities arise if CA systems within different PKI structures formed positions of trust to one another by means of cross-certification. At this point certificate chains could shorten, what seems to be advantageous at first, but after all the complexity within the validation of certificates would be increased significantly through meshing of positions of trust. Thus, in contrast to an architecture with one ROOT within the centralistic approach with a "tree structure", there has to be found a way within meshed positions of trust.

Following the picture of a centralistic structured PKI architecture in form of a tree structure (cf. fig. 2), the validity model of the respective certificates is derived from a shell model. This means, that originating from the ROOT-CA, within the underlying levels only certificates are issued, which have a certificate lifetime not exceeding all issuing CA systems in the certificate chain.

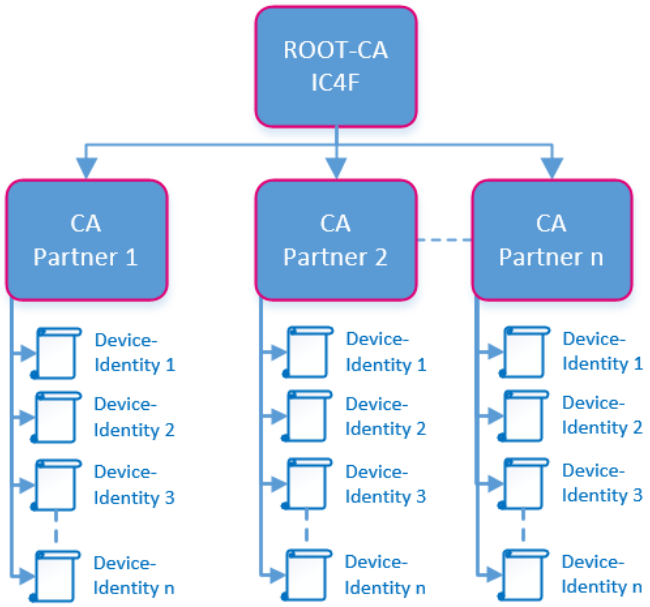

Figure 2: Proposed PKI architecture used by IC4F.

VI. Future Work

Within the IC4F project, the secure connectivity gateway as well as the proposed PKI solutions will be prototypically integrated in several scenarios described in [1] and tested extensively in 2019.

Acknowledgment

The IC4F project is funded by the German Federal Ministry of Economics Affairs and Energy.

References

[1] E. Zielinski et al.: Industrial Communications for Factories. Workshop "Advanced Communication Networks for Industrial Applications." NetSys Conference. Garching, March 2019 (to be published)

[2] https://www.telesec.de/en/tcos-en, accessed on November 29, 2018

[3] NXP Semiconductors, "Secure Boot on i.MX 50, i.MX 53, i.MX 6 and i.MX 7 Series using HABv4", Document Number: AN4581, Rev. 2, 2018

[4] NXP Semiconductors, "Security Reference Manual for i.MX 7Dual and 7 Solo Applications Processors", Document Number: IMX7DSSRM, Rev. 0, 2017

[5] NXP Semiconductors, "i.MX Yocto Project User's Guide", Document Number: IMXLXYOCTOUG, Rev. Rev. L4.9.88_2.0.0-ga, 2018

[6] C. Fruhwirth, "New Methods in Hard Disk Encryption", Institute for Computer Languages Theory and Logic Group Vienna University of Technology, 2005

[7] https://gitlab.com/cryptsetup/cryptsetup/wikis/DMVerity, accessed November 2018