Secure Reconfiguration in Future Mobile Communication Systems

Rainer Falk, Radha Krishna Atukula, Ulf Lücking

Abstract— **Reconfiguration is the subject of many investigative activities performed by industrial consortia and collaborative research projects. In the IST End-to-End Reconfigurability (E^2R) project [1], a perspective on reconfiguration is taken which is not restricted to terminals, but also takes network aspects of architecture into account. Reconfiguration will only be accepted and hence become a success in the market if the security requirements of all stakeholders are satisfied adequately. Secure reconfiguration involves device protection, secure download to defend against potentially malicious software and secure reconfiguration signalling to prevent manipulation of the reconfiguration process.**

I. Introduction

RECONFIGURATION allows changing properties of communication equipment that have previously been fixed by their mere design. It allows the flexible adaptation of reconfigurable equipment to user and operator preferences and the dynamic adaptation to changing network conditions as utilization, radio channel quality, or interference (software defined radio). The improved flexibility poses the threat that changes are made to the configuration of a device that contradict the interests and expectation of end users, network operators and service providers, equipment manufacturers and also regulatory authorities. Malicious radio software could invalidate essential conformance properties, and it could also lead to other types of harm. Without suitable protection mechanisms, other security mechanisms required for secure network access could be circumvented, a user's private data could be sent to unauthorized parties, or call premium rate numbers could be called in the background.

Figure 1 illustrates the main reconfiguration security issues: The overall goal is to ensure a secure and compliant operation. Only authorized software shall be accepted for critical functions, the reconfiguration may be controlled only by authorized reconfiguration managers (RM), and conformity of reconfigurable terminals has to be ensured.

This paper starts in section II with a summary of reconfiguration security requirements. Section III covers security aspects local to the reconfigurable equipment. Section IV describes concepts for authorization (certification) of reconfiguration software. The reconfiguration process is investigated in section V, covering in particular decentralised reconfiguration control. Section VI describes network-based configuration validation to detect configurations that would not be accepted by the target device or not work correctly. Section VII concludes the paper.

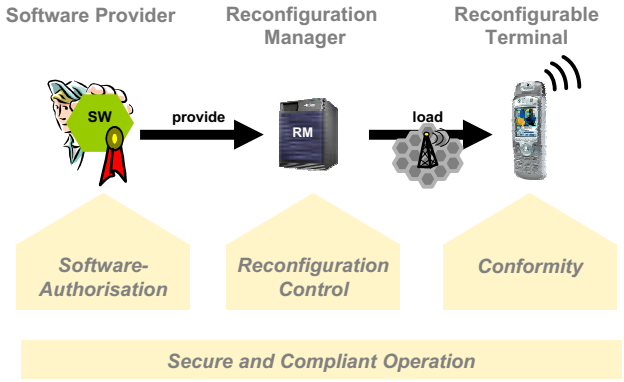

Fig.1: Reconfiguration Security Issues

II. Security Requirements

To ensure reliable, correct operation despite flexibility and openness introduced by reconfigurability, additional reconfiguration-specific security requirements have to be covered in addition to mobile security requirements in general. From equipment perspective it is necessary to provide a protected reconfiguration environment, which shall use privilege modes to grant access to resources only to authorised entities (depending on trust levels). The targeted system shall enforce flexible authorisation policies over resources distributed within the equipment. It shall ensure secure transport of reconfiguration data and sensitive information over a protected reconfiguration communication channel. Exchanged data shall be verified for success of integrity check and kept secure in the equipment with the support of intrusion/tamper detection mechanisms, in order to guarantee the correctness and the privacy of profile and context data. The architecture shall allow the implementation of flexible access control strategies using a hardened 'Real-time Operating System' (RTOS) with respect to security incorporating secure

Rainer Falk is with Siemens AG, Corporate Technology, D-81730 Munich, Germany (e-mail: rainer.falk@siemens.com). Radha Krishna Atukula, and Ulf Lücking are with Nokia Research Center, D-44807 Bochum, Germany (e-mail: radha.atukula@nokia.com, ulf.lucking@nokia.com).

bootstrap methods in order to safeguard the equipment starting in a predefined healthy initial state where all characteristics of the equipment are predictable and under control.

In a vertical market model, hardware and software originate from a single entity that is responsible for conformity. In a horizontal market model, hardware and software originate from independent providers. To support horizontal market model it is mandatory to implement authorisation policies that use a robust certification model. These certificates are validated to the root, where highest trust is ensured, before download and installation of software from third party vendors. The equipment shall also consider fault management by careful monitoring of software installation and invoke recovery and/or rollback procedures in case of equipment failure to perform as expected.

III. Reconfigurable Equipment

Device manufacturers have shown interests in creating closed software environments where only manufacturer accepted code is able to execute. Anyhow, attackers succeeded in getting unauthorised access enabling them to execute their own applications. The same threat could affect reconfigurable equipment (terminals, basestations, access points...), where an attacker could reconfigure the equipment according to his/her own needs.

To be able to create a system with practical applications, secure equipment would require robustness of protocol, operating system (OS) and hardware against all possible threats [4]. A 'Secure Environment' (SE) provides a persistent storage for sensitive data, mechanisms to guarantee memory integrity, random number generation, and cryptographic support. It offers a possibility to load and execute operational software in a secure RAM, which resides as a part of an on-chip memory unit and is visible only if the system is in a secure OS mode. A secure OS mode is characterised by having permission to execute small amounts of security critical code to run as a monitored process. Protected reconfiguration software, which is a trusted piece of code, would also run in the secure OS mode.

The purpose of this mode is to offer mechanisms for expanding the SE services beyond the prefixed features mentioned above. Authentication of trusted software shall be performed before loading and secure enforcement shall guarantee the execution of software in the secure environment. The following paragraphs describe the developed architecture and involved entities as shown in Figure 2.

Configuration Management Module – This module manages the reconfiguration processes according to specified semantic, protocols and configuration data model.

Installation Manager – This entity resides within the Configuration Management Module and coordinates the sequencing of payload installation. It also records progress information as the installation proceeds to allow recovery or rollback after various installation errors.

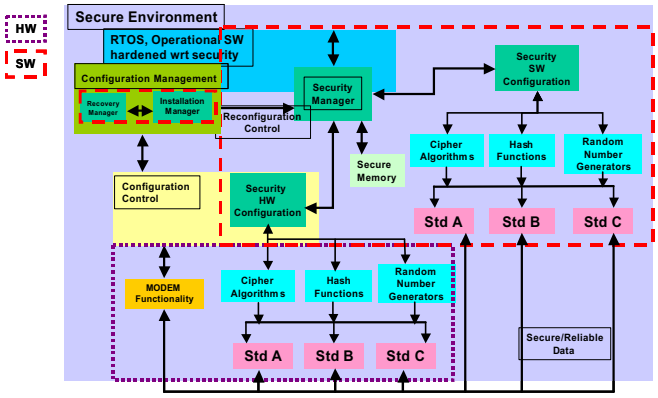

Fig.2: Secure Environment local to Equipment

Recovery Manager – This module is invoked after an installation error has occurred, and its functions include determining the appropriate action to take to recover from an installation error, restoring the context of the install process at the point of the error, and instigating the recovery or rollback actions.

Security Manager – Communicates to *Configuration Management Module*, respective *Installation Recovery Manager* scheduling the correct data delivery, error correction and recovery procedures.

Security HW Configuration – May exist as a specialized entity in *Configuration Control Module* and handles all security related hardware configuration.

Security SW Configuration – Is responsible for configuring algorithms, which exist in software.

RTOS – Will accommodate secure procedures to ensure integrity and specialized bootstrap methods.

The synergy of these entities results in the realisation of the 'Secure Environment' of the reconfigurable equipment.

IV. Reconfiguration Software Authorisation

Reconfiguration software can be classified according to the stakeholder that is the origin of restrictions for reconfiguration software to be acceptable:

- *Regulator* (radio software relevant for conformity of radio emissions, e.g. transmission frequency, emission power, product responsibility),
- *Network operator* (e.g. monitoring and selection of most suitable radio technology, handover decisions, and medium access algorithms),
- *Service provider* (e.g. "branding" of user interface, software needed for service-provider specific services),
- *End user* (e.g. applications, user interface themes, background images, ring tones).

A well known and widely used security mechanism to protect software download is signed content. The provider of a software module attaches a digital signature to the module that can be verified by the receiving device. The digital signature ensures that the module has not been modified (integrity) and

attests its provider (authentication of origin). The receiving device validates the signature of a received software module ensuring that it has not been tampered with and it checks whether it originates from a trusted provider. For managed runtime environments, also the granted permissions are determined. Correct root keys allowing to verify digital signatures of authorised software providers have to be available on the device. The kind of the provider determines where these root keys can be stored, i.e. on the device itself (e.g. for keys belonging to the device manufacturer or keys corresponding to a well-known radio software approval authority) or on a pluggable user module (e.g. for keys corresponding to or being defined by the user's service provider), and who may update them. A chain of digital certificates allows to verify the public key of the signer back to a stored trusted root key. Certificate status checks using e.g. the Online Certificate Status Protocol allow to detect revoked certificates. This may be relevant when potentially a large number of software providers can authorise reconfiguration software as it allows to revoke certificates issued by mistake or issued to unreliable software providers. However, in practice certificate status checks are not widely used.

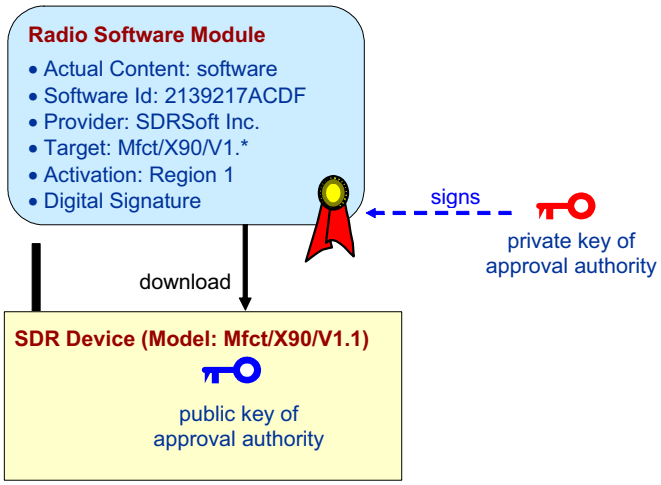

Fig.3: Certification of Radio Software Module using a Digital Signature

While the basic mechanisms for secure software download are well known, specific for radio reconfiguration software is the policy, i.e. which party has to create the signature and thereby indicate towards a terminal that the module may be accepted (authorisation, approval) thereby assuring that conformance properties are not invalidated. Accessibility to the radio download security solution needs to be restricted to ensure that its properties cannot be overridden. Even when a radio software module is authorised to be downloaded on a reconfigurable device, there may be further restrictions concerning the conditions under which it may be activated: In particular, different regulations depending on the region/location have to be respected, and possibly a radio software module requires even a dynamic authorisation by the currently used network stating that its features are supported by the network. In the vertical market model, radio-related software is accepted only if it is authorised by the device manufacturer. Here, the device manufacturer can not only ensure that conformance properties are met, but also ensure a proper operation as he is still in control on which radio software is accepted on devices he brought into the market. Alternative approaches suitable for horizontal market models are a current research topic, as combined or separate approval for radio hardware and radio software, moving the responsibility to validate online a radio configuration to a network-based function, or to supervise radio emissions and to perform reactive measures if a malfunction is detected.

Figure 3 shows an example for radio software authorisation: An approval authority authorises a radio software module by computing and attaching its digital signature using its private key. The module's meta information contains entries to identify the module, the authorised target device, and restrictions on the activation. The device verifies the signature using the public key of the approval authority. Depending on the policy, this approach can be used to realise a vertical market model where only software authorised by the device manufacturer is accepted as well as to realise a horizontal market model where each hardware-software-combination requires authorisation from a trusted approval authority. Should independent approval of radio hardware and radio software be deemed acceptable, it could be realised in a similar way. The target devices a radio software module is authorised to be used on can be encoded as part of the meta-information. A change history allows identifying the responsible party in case of malfunctioning configurations.

Secure software download can be complemented by a restricted radio execution environment. Control parameters as frequency, output power, and bandwidth driving reconfigurable radio hardware can be validated to lie within an authorised range. Actual radio emissions can be monitored and compared with reference data, in particular a spectral power density mask. Reference data could be fixed or changeable only with special restrictions. These would relate to the conformance constraints that the device enforces independently of the currently executed radio software. The device itself or a communication network can monitor correct protocol behaviour, e.g. obeying power control commands. When a malfunction is detected, the device would reconfigure to the previous working configuration or switch back to a fixed failure-mode configuration.

V. Reconfiguration Process

Reconfiguration process has to be controlled by the network when the nature of a reconfiguration is not understandable to end users, when required information is not available to them, or when reconfigurations occur so often that it would be inconvenient for end users to be directly involved (e.g. dynamic radio reconfiguration to adapt to local network conditions). Using a single, centralised reconfiguration

manager leads to clear responsibilities. For example, the user's service provider (home network) wants to define and modify configuration parameters and install software updates (customer care). From a security perspective, it needs to be ensured that only the single trusted reconfiguration manager can actually define and modify the device configuration. This requires protection of the communication between the reconfiguration manager and the reconfigurable device (authentication of reconfiguration manager, optionally also authentication of reconfigurable device; integrity, possibly confidentiality of communication). Security protocols like IPsec or SSL/TLS can be used for this purpose.

One objective of reconfiguration is the dynamic adaptation of the device configuration to changing network conditions. The configuration should correspond to the capabilities, preferences, and dynamic properties (load, radio link properties) of the currently used network. This implies that the currently used network needs to have the possibility to modify the current configuration of accessing devices. A visited network could either signal intended changes towards the central reconfiguration manager placed in the home domain, or it could be granted direct access enabling it to perform intended reconfigurations without involving the home reconfiguration manager. Considering roaming users, this leads to a decentralised reconfiguration control where not only a single reconfiguration manager can modify the configuration, but where also a visited network itself can implement local adaptations. Furthermore, this allows the visited network to perform a coordinated reconfiguration of end user and infrastructure equipment.

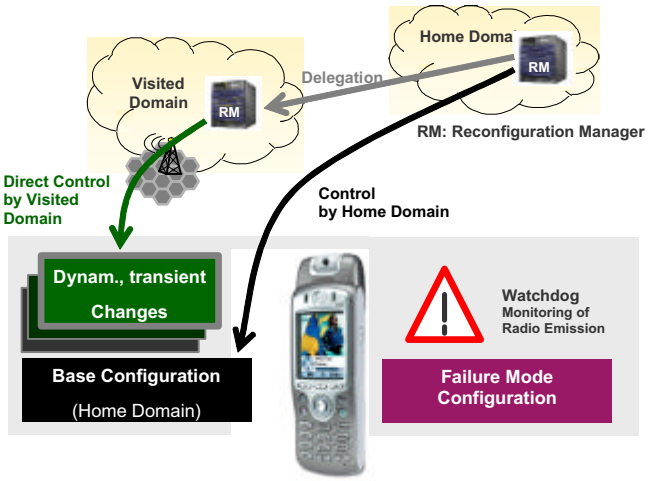

Fig.4: Decentralised Reconfiguration Control

Figure 4 illustrates decentralised control over the device configuration. The user's service provider (home domain) could be at the top of the hierarchy defining the base configuration. He can delegate control on terminal reconfiguration partly to roaming partners (visited domain). A visited network could perform some reconfigurations on terminals of roaming users, but only as far as allowed by the users' service provider. Using several, only temporally valid configuration profiles associated with specific networks ensures that the changes made by a specific network are in effect only as long as that network is actually used. These dynamic, transient changes to the device configuration have no effect on the configuration when a different network is used. The device stores also a fixed failure mode configuration that is activated when watchdog function detects a malfunction.

In general, the same reconfigurable device is used in different environments, e.g. in an operated cellular network that can belong to the user's service provider (home network) or a different network the user is roaming to (visited network), the user's private home network, or the enterprise network of his employer or a business partner. This implies that the device needs to be configured accordingly so that it can be used in these different environments. Figure 5 shows several configuration profiles that a device will have to hold. They correspond to different networking environments and are defined by corresponding stakeholders. Access control checks ensure that only the authorised stakeholder can define and update its associated profile. The device selects the profile corresponding to the currently used network and activates it as current configuration.

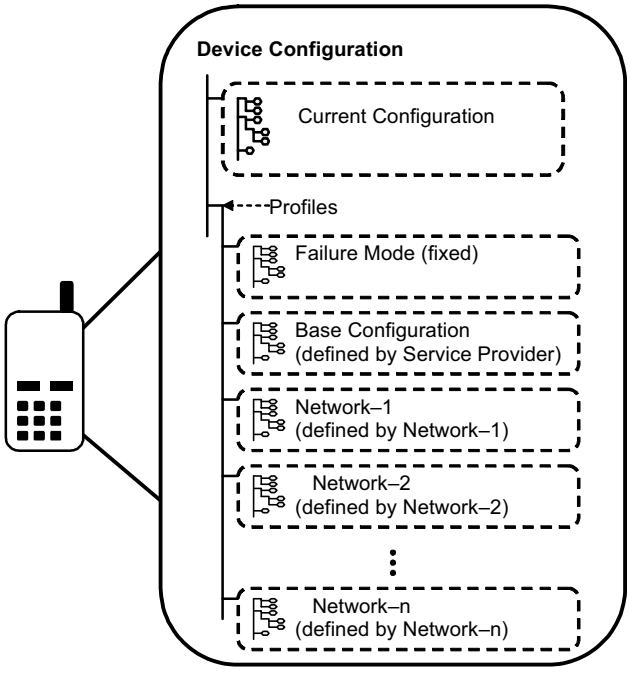

Fig.5: Configuration Profiles

VI. Network Based Configuration Validation

Before the reconfiguration of a target device is actually performed, the intended new configuration (software element or set of parameters) can be validated in the network to check whether it can be expected to operate correctly on the specific target device and in the current network environment. The objective is to detect software/configurations that would not work as expected *before* they are actually downloaded.

The general problem to verify whether the execution of a piece of radio software would lead to non-conformant radio emissions is not decidable, as it would solve also the halting problem [3]. The approach taken within E^2R is to plan for several validation checks that can potentially be performed and select a subset of those that are appropriate and required depending on type of reconfiguration, target device, and kind of software. The objective of these pragmatic checks is to detect and prevent reconfiguration attempts that would fail later on, but they do not "guarantee" the correct operation. The final responsibility for correct and conformant operation lies at the target device.

Possible validation checks include comparison of an identifier of the indicated target execution environment with the actual target environment, scanning the software for calls to manufacturer specific APIs, mimicking the signature validation check as they will be performed by the target device, checking a database containing recommended/tested and/or known problematic software, evaluating statistical data whether a software module has previously been successfully installed and executed on devices of the same type, and the simulation on a device simulator. In roaming cases, a joint validation can be done involving both home and visited domains. A configuration would only be accepted if neither the home not the visited domain detect a problem.

VII. Conclusion

The overall objective of secure reconfiguration is to ensure a reliable operation, despite of the ongoing, flexible download of software and configuration information that could lead easily to severe security problems would suitable protection means not be available. It would be rather easy to address the security issues coming from reconfiguration by applying over-restrictive security measures. For example, only a single, trusted reconfiguration manager could be used, having only the capability to perform very restricted types of reconfiguration for which it is clear that they cannot introduce any security problems. However, the real challenge is to develop security concepts for more open and decentralised approaches to reconfiguration that still ensure a reliable, correct operation according to the expectations of end users and operators and respecting regulatory boundaries. After summarising reconfiguration-related security requirements, this paper described some results from the E^2R project:

- Concepts to provide a "Secure Environment" local to the reconfigurable equipment have been presented.
- A framework for authorisation (certification, approval) of (radio) reconfiguration software has been described that is based on signed content, restricted radio execution environments, and restrictions concerning when a certain software module may be activated.
- Starting with centralised reconfiguration control, approaches for a decentralised control on the reconfiguration were described. Usage of several configuration profiles to support distributed reconfiguration control is also described.
- A flexible validation concept where depending on the target device and the functionality to be modified appropriate checks are selected and performed. Several checks can be included in this framework.

The presented concepts for secure reconfiguration help to prevent the misuse of the flexibility coming with reconfigurability and thereby enable that it is used for the benefit of involved stakeholders as it allows the flexible support of different wireless standards, the fast introduction of new services, and the adaptation and personalization according to the preferences of end users and operators.

Acknowledgment

This work has been performed in the framework of the EU funded project E^2R. The authors would like to acknowledge the contributions of their colleagues from E^2R consortium.

References

[1] IST-E²R "End-to-End Reconfigurability", Project Web Site, http://www.e2r.motlabs.com

[2] R. Oppliger: "Security Technologies for the World Wide Web", 2nd Edition, Artech House, Norwood MA, 2003

[3] IST-2001-34091 SCOUT "Architecture, Functions and Security Analysis and Traffic Management Schemes for IP-Based Mobile Networks and Re-configurable Terminals in Cellular and Ad-Hoc Networks", Deliverable D4.1.2, Jan. 2004, http://www.ist-scout.org

[4] TS 21.133 V410 3GPP, 3G Security, Security Threats and Requirements