A SECURITY ARCHITECTURE FOR ACCESSING HEALTH RECORDS ON MOBILE PHONES

Alexandra Dmitrienko, Zecir Hadzic, Hans Löh, Marcel Winandy
Horst Görtz Institute for IT Security, Ruhr-University Bochum, Bochum, Germany
Ahmad-Reza Sadeghi
Fraunhofer-Institut SIT Darmstadt, Technische Universität Darmstadt, Darmstadt, Germany

Keywords: Health records, Mobile computing, Smartphone, Security architecture, Trusted computing.

Abstract: Using mobile phones to access healthcare data is an upcoming application scenario of increasing importance in the near future. However, important aspects to consider in this context are the high security and privacy requirements for sensitive medical data. Current mobile phones using standard operating systems and software cannot offer appropriate protection for sensitive data, although the hardware platform often offers dedicated security features. Malicious software (malware) like Trojan horses on the mobile phone could gain unauthorized access to sensitive medical data.

In this paper, we propose a complete security framework to protect medical data (such as electronic health records) and authentication credentials that are used to access e-health servers. Derived from a generic architecture that can be used for PCs, we introduce a security architecture specifically for mobile phones, based on existing hardware security extensions. We describe security building blocks, including trusted hardware features, a security kernel providing isolated application environments as well as a secure graphical user interface, and a trusted wallet (TruWallet) for secure authentication to e-health servers. Moreover, we present a prototype implementation of the trusted wallet on a current smartphone: the Nokia N900. Based on our architecture, health care professionals can safely and securely process medical data on their mobile phones without the risk of disclosing sensitive information as compared to commodity mobile operating systems.

1 INTRODUCTION

The usage of mobile phones as multi-purpose assistant device in healthcare has been proposed in several application scenarios. Its usefulness is derived from its mobility and flexibility, i.e., today’s smartphones offer appropriate computing and storage capacity allowing the realization of various applications that can be used basically from everywhere. For instance, healthcare professionals can use a mobile phones to download and share electronic health records of their patients (Benelli and Pozzebon, 2010). In other scenarios, patients use their mobile phones to provide personal health data, e.g., taken from additional biosensors, to a medical information and diagnosis system (Han et al., 2008).

While smartphones are very flexible and cost-efficient computing devices, they generally do not offer sufficient security mechanisms to protect the data they operate on. This is mainly due to the architectural shortcomings of their operating systems, which are derived from the same (security) architecture as desktop operating systems. Typical examples are Google Android (Android Open Source Project, 2010), Apple iOS (Apple Inc., 2010), Symbian (Symbian Foundation Community, 2010), and Windows Mobile (Microsoft, 2010). Although, some of them provide more sophisticated security mechanisms than their desktop counterparts, e.g., application-oriented access control in Android (Google Android, 2010), they still suffer from fundamental security problems due to their large code base and complexity, lacking of strong isolation of applications (secure execution) and insufficient protection of stored data (secure storage). Recent attacks on smartphones demonstrate their vulnerability (Iozzo and Weinmann, 2010; Ven-
The security kernel relies upon. The concept of a trusted computing base, i.e., the code that all security critical execution and to have less complexity for the have shown how to achieve strong isolation for separating systems, e.g., (Fraim, 1983; Karger et al., 1990), which are isolated based on hardware mechanism execution environments for security-critical functions, cure boot components M-Shield (Azema and Fayad, 2008) offer security framework to protect the credentials from malware on the device, and, hence, could be accessed by unauthorized parties.

However, modern smartphone hardware offers advanced security functionality, which are included in their processors, but generally not used by the mainstream mobile operating systems. For instance, ARM TrustZone (Tiago Alves, 2004) and Texas Instruments M-Shield (Azema and Fayad, 2008) offer secure boot\(^1\) functionality, secure storage and secure execution environments for security-critical functions, which are isolated based on hardware mechanism from other processes running on the phone.

On the other hand, previous works on secure operating systems, e.g., (Fraim, 1983; Karger et al., 1990), have shown how to achieve strong isolation for secure execution and to have less complexity for the trusted computing base, i.e., the code that all security relies upon. The concept of a security kernel (Anderson, 1972) incorporates all relevant functionality needed to enforce the security into a kernel that is isolated and protected from tampering by other software and small enough to be verifiable for its correctness and security. While earlier systems suffered mostly from poor performance in those days, recent CPU hardware technology, especially their virtualization support, and the development of efficient microkernel software architectures (Liedtke, 1995) allow for the realization of security kernels with low performance overhead while maintaining compatibility to existing applications. For example, Turaya (EMSCB Project Consortium, 2008) and the OpenTC security

1\(^{\text{Secure boot means that a system terminates the boot process in case the integrity check of a component to be loaded fails.}}\)
Mobile TruWallet prototype. Finally, we conclude in Section 5.

2 PROBLEM SCENARIO

We consider a scenario in which electronic health records (EHRs) of patients are stored on a local server of a healthcare provider, e.g., in a hospital. Health care professionals, like doctors and nurses, are equipped with mobile computing devices (smartphones) on which they can create, edit, and view EHRs. The EHRs are stored on the e-health server, and the smartphones communicate with the server via wireless network connections. For instance, the access of medical data can be realized with web-based applications, using standard web browser software on mobile devices. Figure 1 depicts the scenario we consider.

Since EHRs are very security-sensitive private data, and in most countries protected under strong privacy laws, unauthorized access to these data must be prevented. An adversary may try to eavesdrop or manipulate the sensitive data. As mentioned before, end-user devices are typically the least specified and least secured devices in healthcare infrastructures. Hence, an adversary would most likely try to attack the mobile phone and its communication connection to the server in order to illegitimately access medical data.

Studies like (Vouyioukas et al., 2008) have analyzed how to secure the data transfer, i.e., via encryption (for confidentiality), digital signatures (for integrity and authenticity), and user authentication (for legitimacy of access). However, the protection of the critical cryptographic keys that are needed for those mechanisms is not addressed appropriately. Hence, an attacker who gains access to these keys can circumvent any other protection mechanism.

Therefore, in this paper we concentrate on an adversary model in which the attacker targets the mobile computing device of health care professionals in order to obtain the secret login credentials or keys that are needed to access the EHR server. Once the adversary has access to these credentials, he can download or modify all medical data from the server to which the credentials allow access to. To achieve this goal, the adversary can follow two strategies:

- **Direct Access.** The adversary tries to directly access the sensitive data or keys. He could try to manipulate software running on the phone to access the data, or he could steal the device and try to access the data. The former could be achieved by letting the users install malicious software (malware, such as trojan horses) without their notice, e.g., when they browse to a website containing malicious code that exploits a vulnerability of the phone’s software to install the malware. Doctors may use their phones also for other tasks and they may want to download additional applications to run them on the phone, which could create the vulnerability for such an attack.

- **Indirect Access.** The adversary tries to trick the users to enter their passwords into a faked EHR viewer application. The faked application looks like the real one, but instead of logging into the server, it sends the passwords to the adversary. The faked application could be installed on the phone in the same way as malware described above.

The problem with a commodity mobile phone operating system (OS) is that it cannot provide a sufficient level of protection for the applications or stored credentials. A mobile phone OS that is directly derived from a desktop OS (e.g., Linux or Windows) has limited protection capabilities, i.e., simple process isolation and access control. However, malicious applications can modify or eavesdrop data of other applications since they are running with the same user privileges as other applications.

A more advanced OS, e.g., like SELinux (Loscocco and Smalley, 2001), can enforce mandatory access rules, which provide a stronger isolation of different applications from each other. For instance, a text editor could only edit text files, whereas an audio application could not modify text files. The application of such a system in a mobile e-health scenario has been shown earlier (Agreiter et al., 2007). However, SELinux is a very complex system with security policies that are hard to configure correctly even for moderately complicated scenarios. Moreover, due to a relatively large code base, it is infeasible to perform a comprehensive formal (or even semi-formal) verification of the correctness and security of SELinux. Another example is Android (Google Android, 2010), which provides a similar application-oriented access control, i.e., it defines for each application different access rules and privileges — in contrast to user-oriented access
control as in normal Linux and Windows, where all programs of one user share the same access rights.

Nevertheless, even advanced mobile phone OS’s still suffer from ineffective protection against unauthorized modifications of programs or even modifications of the OS itself. An adversary could install on the user’s phone additional (faked) programs or replace existing programs. The user has seldom a chance to notice the modification, and critical data like credentials could be transferred to the adversary.

3 WALLET ARCHITECTURE

3.1 General Idea

Our security architecture aims to protect against the attacks described above. To counter direct access attacks, our architecture is based on a security kernel that isolates different applications, supports secure boot, and provides secure storage. Hence, authentication data is stored encrypted, and can only be accessed by the legitimate application (TruWallet) when the correct (unmodified) system has been booted.

Our wallet architecture aims to prevent indirect attacks by letting the wallet handle all authentication procedures. During a normal authentication, users do not enter passwords (this is automatically done by the wallet), hence they cannot accidentally disclose them towards a fake application that tries to spoof the look and feel of the legitimate EHR viewer or another application trusted by the user.

Figure 2: General idea of TruWallet.

Our wallet-based security architecture provides two levels of protection (cf. Figure 2):

1. Protection of Authentication Data. TruWallet protects the user’s credentials (username and password) against unauthorized access. This approach is generic, and can be used also for other scenarios (e.g., web applications like eBay or Amazon). Indeed, TruWallet can be used simultaneously by different applications, yet it only authenticates each application to the server where it has been registered as legitimate application before.

2. Protection of Medical Data. An isolated EHR viewer (which can be a special-purpose browser) is used to view EHRs. This viewer cannot be modified because a fixed program image is executed, which is measured by the security kernel by computing a cryptographic hash and compared to a known-good reference value. This ensures that all modifications of the EHR viewer can be detected. In case a browser is used as EHR viewer, this browser is only allowed to contact the EHR server and cannot connect to other sites.

3.2 System Model

Our system model for TruWallet consists of several parties (see Figure 3): A user interacts with a computing platform through a secure graphical user interface secure GUI. An EHR viewer is used to render content that it gets from the wallet, which is acting as a proxy. The wallet obtains the requested content from the server, blinds security-sensitive fields (e.g., password) on the pages presented to the browser, and fills in login credentials when logging into the system. For this, TruWallet has to handle two different SSL sessions: one between wallet and EHR viewer, and one between wallet and server. The secure GUI controls the input/output devices and multiplexes the screen output of the EHR viewer and of the wallet. Moreover, it always indicates the name of the application the user is currently interacting with via a reserved area on the screen, hence providing a trusted path between user and application. Moreover, our architecture includes a compartment for non-medical data and applications. This compartment is strictly separated from the EHR viewer and can be used for arbitrary applications.

3.3 Generic Wallet-based e-Health Architecture

The generic TruWallet architecture is based on a security kernel, which is a small trusted software layer, providing trusted services and isolated compartments. Thus, the security kernel ensures runtime security of the system. Compartments contain arbitrary software, e.g., a complete legacy operating system (Linux in our case), and may communicate only via well-defined interfaces. In particular, a malicious compartment cannot read arbitrary memory of other compartments. In our solution, EHR viewer, non-medical applications and wallet run in different compartments, and we assume that arbitrary software (including malware like Trojan horses and viruses) may be running in the non-medical compartment. There-
fore, the security of our solution is based on trusted components (wallet and EHR viewer) that are executed in separated compartments, isolated from untrusted software that might be running simultaneously on the same platform.

In an earlier work (Gajek et al., 2009), we have demonstrated the feasibility of the wallet architecture on a PC platform. In the PC-based implementation, the compartmentalization was realized by using the isolation property of virtual machines combined with the resource sharing control of an underlying microkernel. The wallet compartment is trusted, which is motivated by the fact that the complexity of the wallet is much lower than that of an EHR viewer or a compartment containing several different applications. Moreover, users cannot install arbitrary software (which may be malicious or flawed) in the wallet compartment, but they may install arbitrary viewers or other tools into other compartments. To prevent unauthorized access by other users to the platform and, hence, the sensitive data, the security kernel requires an overall user authentication (e.g., a user password) to login into the whole system. In this way, the credentials stored by the wallet are bound to the corresponding user.

**Trusted Computing Support.** Trusted Computing (TC) hardware and TC-enabled software is used to provide authenticated boot, i.e., based on a “chain of trust”, the integrity of the software stack including the Trusted Computing Base (TCB) can be verified later, e.g., before access to cryptographic keys is allowed. An alternative to authenticated boot which is usually used on mobile platforms is secure boot: In this case, the system’s integrity state is compared to reference values and can be started only if it has not been modified. Moreover, TC hardware can be used for secure storage, i.e., encryption keys protected by the hardware can only be used if load-time integrity of the system is maintained. Thus, the credentials stored by the wallet are bound to the TCB to prevent an adversary from gaining access to the data by replacing software (e.g., booting a different OS). On the PC platform (Gajek et al., 2009) we used a Trusted Platform Module (TPM) version 1.2 (Trusted Computing Group, 2009) as TC hardware for TruWallet.

The TPM is a dedicated security chip that provides – among other features – cryptographic operations (e.g., key generation, encryption, digital signatures), support for authenticated boot, and the possibility to bind cryptographic keys to the load-time integrity state of the system.

![Figure 3: Generic TruWallet Architecture.](image)

**3.4 Mobile TruWallet**

To implement our security architecture for mobile e-health scenarios, several building blocks for mobile environments are required:

- Trusted hardware for mobile platforms which supports features to protect cryptographic keys and verify the system integrity;
- a secure hypervisor layer for mobile platforms to provide isolated execution environments for applications;
- a security kernel with a secure GUI for mobile platforms to provide a trusted path between the user and applications, and with secure storage for applications;
- a trusted wallet (TruWallet) to handle authentication and protect the user’s credentials.

In the following, we briefly introduce the first three building blocks, before we focus in more detail on the implementation of a trusted wallet on a mobile phone in the next section.

**Trusted Hardware for Mobile Platforms.** The architecture of TruWallet relies on trusted hardware for performing security critical operations. To instantiate TruWallet architecture on a mobile phone, we have to use mobile hardware security extensions instead of a TPM (which is not available on current phones). On mobile platforms, general-purpose secure hardware such as M-Shield (Azema and Fayad, 2008) and TrustZone (Alves and Felton, 2004) is available. In this paper, we focus on M-Shield, because this hardware extension is available in some current mobile devices.

---

2In fact the security kernel has to provide comprehensive user access control as in typical operating systems, including system login and screen lock functionality, in order to prevent unauthorized access to the wallet. However, the details of those mechanisms are out of scope of this paper.

3Of course, it is important that these reference values are stored in a secure location, e.g., protected by security hardware, to avoid manipulations.
phones, including Nokia N900 (which we used for our prototype).

M-Shield provides a small amount of dedicated on-chip ROM and RAM as well as one-time programmable memory for device keys which are accessible only in a special execution mode of the main CPU – the Trusted Execution Environment (TrEE). A secure state machine (SSM) guarantees secure switching between both processor modes, thus the TrEE and normal execution environment are isolated from each other. M-Shield enables the TrEE on a device with the following features: (i) isolated secure code execution; (ii) secure boot; (iii) hardware-based secure storage.

Secure Hypervisor for Mobile Devices. Several microkernels for mobile and embedded devices have been implemented, for instance the commercially available L4 microkernels OKL4 (Open Kernel Labs, 2010) and PikeOS P4 (Brygier et al., 2009). These microkernels provide isolation between user space applications, just like their counterparts on other platforms (e.g., on PCs). Therefore, they can be used for a secure hypervisor layer for a security kernel on mobile phones. In particular, the seL4 microkernel has been formally verified for correctness (Klein et al., 2009), hence taking an important step towards building a formally verifiable security kernel on top of a microkernel.

Security Kernel with Secure GUI for Mobile Devices. The Turaya Trusted Mobile Desktop (Selhorst et al., 2010) implements a security kernel with a secure user interface for mobile devices. Its TCB consists of a hypervisor layer and a trusted software layer. The hypervisor layer is implemented on top of an L4 microkernel, which has been ported to the Nokia N900 mobile phone. The Trusted Software Layer contains a number of security services, such as a secure graphical user interface (called TrustedGUI), a virtual private network (VPN) client, and a file encryption service.

4 WALLET PROTOTYPE ON NOKIA N900

In order to demonstrate the feasibility of running a trusted wallet on a mobile phone, we have implemented Mobile TruWallet, a mobile version of trusted wallet, on a Nokia N900 device.

4.1 Mobile TruWallet Architecture

Instead of realizing the full implementation of a security kernel, for which we refer to the works of (Brygier et al., 2009; Klein et al., 2009; Selhorst et al., 2010), we have implemented the wallet on Maemo (Maemo, 2010), which is based on Linux and provides standard operating system process isolation and discretionary access control.

Figure 4: Mobile TruWallet Architecture.

The architecture of Mobile TruWallet we have implemented is depicted in Figure 4. As it can be seen, TruWallet resides on the operating system side, but also operates on secrets at the same time, e.g., maintains a TLS channel to the web-server and also performs authentication with the user passwords. However, our generic architecture assumes that TruWallet is isolated from the rest of the system. This assumption is reasonable to some extend in the context of existing operating systems for Nokia mobile phones: Symbian OS and Maemo. Their platform security enables, with different degrees, process isolation. The microkernel-based Symbian OS provides process execution isolation and enforces control on inter-process communication via a capability mechanism, while Maemo’s security model is based on Discretionary Access Control (DAC) which enforces security by process ownership.

We achieved process isolation on Maemo by creating a Mobile TruWallet process under a unique UserID and defining restrictive access rights to that UserID. Note that for this prototype, we rely on the standard Unix/Linux discretionary access control security framework, and there is always the threat that an administrator (root account) with the super-user access rights is compromised. However, we implemented the wallet as if it was running on a security kernel. This approach allows us to concentrate on the wallet-specific aspects for the prototype (i.e., performance, user interface, compatibility to the mobile
web browser and web sites, etc.). In a later stage, the wallet can be easily adapted to a security kernel system like the L4-based one on N900 (Selhorst et al., 2010).

Nokia N900 device is based on M-Shield secure hardware. We utilize M-Shield functionality for the secure boot, and we also implement secure storage functionality on the top of M-Shield.

Only authenticated programs, so-called protected applications (PAs), can be executed within the TrEE of M-Shield. However, protected applications have to be authorized, i.e., certified, by the device’s M-Shield stakeholder, most likely the device manufacturer. As we have not been able to get the PA certified by the device manufacturer, we use a different approach: We reuse the general purpose APIs available for M-Shield TrEE. This approach allows third parties to leverage the TrEE. For instance, the On-board Credentials platform (ObC) (Kostiainen et al., 2009) provides the means to develop programs for the TrEE without the involvement of the manufacturer. In our implementation, we build the secure storage functionality of Mobile TruWallet on top of ObC.

### 4.2 ObC Architecture

Figure 5 illustrates the ObC architecture. The core component of the ObC platform – which resides in the dedicated RAM and can be executed in secure environment – is an interpreter. The interpreter provides a virtualized environment where “credential programs”, i.e., scripts developed by third parties, can be executed. The credential programs are written using (a subset of) Lua scripting language (Lua, 2010) or in assembler. When a credential program is executed, the interpreter isolates it from secrets that are stored within the TrEE and from the execution of other credential programs.

The interpreter makes use of a Crypto Library which provides an interface for commonly used cryptographic primitives. It provides a sealing/unsealing function for ObC programs, which can be used to protect secret data stored persistently outside the secure environment. Sealed data is encrypted with a key which is protected by the TrEE, and the ObC platform controls the usage of this key. A device-specific symmetric key called ObC platform key (OPK) is used for the sealing/unsealing functionality.

Credential Manager is a mediator between OS side applications and components that reside within the TrEE. It provides an API for third-party developed applications. Using the API the applications can execute credential programs, and create and use new asymmetric keys. The credential manager maintains a database, in which credentials and credential programs are stored in a cryptographically sealed way.

A more detailed description of the ObC architecture can be found in (Kostiainen et al., 2009; Kostiainen et al., 2010).

### 4.3 Implementation

In our prototype, the wallet is implemented in the C programming language, contains about 2600 lines of code, and runs as separate process with a unique UserID. For the SSL/TLS proxy, we use Paros (Paros, 2010), which is an open-source implementation in Java, and it executes as a process with the same UserID as the wallet. We define restrictive access rights on this UserID so that other processes cannot access the data or code of the wallet.

**Accessing Health Records.** The wallet uses the libxml library to parse web sites and web forms in order to search for password fields. Whenever it finds such fields, these forms are put into a cache and are disabled before they are shown in the web browser. This prevents the user from accidentally typing passwords into a potentially malicious or faked web site. Instead users just provide their user name and simply click the submit or login button in the mobile web browser.

Hence, when doctors want to access a health record from the e-health server, they simply open the EHR viewer browser on the phone and click the login button. The wallet replaces then automatically the disabled password field with the actual password of the doctor’s account on the e-health server. This process runs transparently, so the doctor just sees the EHR viewer application, and when the login is completed he can access the health records on the server.
Registration. Before doctors can use the wallet to login to websites like the e-health server, they have to register their account in the wallet on the phone. Therefore, the wallet also looks for registration forms. When the user tries to access a website with the browser for the first time, the wallet asks the user for an existing password or it can create a new one. Figure 6 shows a corresponding screenshot where the wallet dialog pops up after the user opened a website (with a login field) in the browser for the first time.

Interoperability. We have tested our wallet implementation with several public websites, like web e-mail services, eBay, Amazon, etc. Registration and login work transparently and without noticeable performance overhead for the user. Hence, it should be easy to integrate web-based e-health services on our platform. Special-purpose EHR viewers or other medical applications can be supported as well as long as they use SSL/TLS and web-based login procedures. Other authentication protocols could also be integrated, but may require some effort to adapt the wallet.

5 CONCLUSIONS AND FUTURE WORK

Mobile access to electronic medical data is an emerging application scenario with strong security and privacy requirements that is rapidly gaining practical importance. Existing systems suffer from a lack of appropriate protection for security- and privacy-critical data and applications. Moreover, standard operating systems do not use existing hardware security features to their full extent.

To enable secure mobile access to electronic health records containing privacy-sensitive data, we propose an e-health security architecture which protects the user’s authentication credentials as well as the sensitive medical data. Our architecture is based on commonly available trusted hardware components, a security kernel, and a trusted wallet. In this paper, we introduce our comprehensive security architecture, discuss security building blocks on mobile phones, and present our implementation of Mobile TruWallet on a commodity smartphone.

Since our Mobile TruWallet prototype demonstrates the feasibility of the architecture on mobile phones, future work includes the integration of all security building blocks (i.e., the use of hardware security features, a security kernel consisting of a secure hypervisor and a trusted software layer, and the Mobile TruWallet authentication solution) into one system.

ACKNOWLEDGEMENTS

This work was partially funded by the German federal state North Rhine-Westphalia and supported by the European Regional Development Fund under the project RUBTrust/MediTrust. Further, the author Alexandra Dmitrienko was supported by the Erasmus...
Mundus External Co-operation Window Programme of the European Union.

REFERENCES


