WHITE PAPER

Securing the Future of Authentication with ARM TrustZone-based Trusted Execution Environment and Fast Identity Online (FIDO)

The hardware foundations for simple and strong authentication

Rob Coombs

Date: May 25, 2015

Foreword

Smart connected mobile devices are increasingly used for a wide range of business, financial and entertainment uses. Authentication of the user to a remote Internet-based server is the first step for many applications or cloud-based services. Traditional authentication methods of username and password do not work very well for either the consumer who may have difficulty remembering complex passwords or the service provider (usually referred to as the relying party) who may need to factor in the risk that the password has been compromised. To answer these and other problems with passwords, the FIDO (Fast IDentity Online) Alliance has developed new protocols that enable simple, strong authentication between the user, their device and the relying party. On mobile devices FIDO can be used with biometric authenticators to enable services with the swipe of a fingerprint or the scan of an iris. This vastly improved user experience will benefit consumers, make transactions frictionless and is likely to be quickly adopted by device manufacturers.

Hardware-based security is needed to help protect FIDO from malicious attack. Assets such as cryptographic keys, sensitive processes and the capture of authenticator data should be protected from malicious attack and the integrity of the system needs to be maintained. This paper introduces how ARM® TrustZone® technology provides the hardware isolation necessary for a GlobalPlatform Trusted Execution Environment (TEE) and how this security layer is ideally suited to secure FIDO based authentication.

Copyright © 2015 ARM Limited. All rights reserved.
The ARM logo is a registered trademark of ARM Ltd.
All other trademarks are the property of their respective owners and are acknowledged

Introduction

In 2014 ARM's silicon partners shipped more than 2 Billion ARM® Cortex®-A processor based applications into phones, tablets, DTV and other smart connected devices. These devices are increasingly being used to access cloud-based services and for high-value use cases such as payment and handling of corporate or government information. To protect system assets from attack, modern ARM platforms use a combination of technologies: from the Cortex core Hypervisor mode, to the TrustZone based TEE and tamper proof security processors or secure elements protected with ARM SecurCore® processor IP. This multi-layered or compartmentalized approach increases overall system security and provides the right level of protection that goes beyond the operating system to the different assets within a mobile device.

The TrustZone based TEE was designed to deliver enhanced security from scalable software attacks and common hardware attacks (so called shack attacks) at a lower cost to the market. Its architecture provides isolation between the normal world (Rich Operating System and Applications) and a hidden secure world that can be used for sensitive operations such as crypto, key management and integrity checking. It has become an important hardware security layer for device manufacturers that they have been developing and standardizing over the last ten years to protect valuable system assets. The TEE is standardized by GlobalPlatform who have created a compliance and certification program so that independent test labs can check that platforms are protecting against the threats identified in the protection profile. GlobalPlatform have white papers discussing the TEE: this white paper has been written to add information to their documents covering the FIDO use case and ARM TrustZone technology.

The move to password-less login using biometric authenticators is being accelerated through standardization by the FIDO Alliance. FIDO protocols such as Universal Authentication Framework (UAF) enable local user verification with multiple authenticators such as fingerprint sensors, iris scanners or PIN entry replacing the traditional username and password.

It is often said that security is a chain where security relies on a sequence of linked processes. Using this analogy, the first link is secure hardware that can be isolated using TrustZone technology from the normal world rich execution environment and be the basis for trusted boot. Trusted boot initializes the Trusted OS and therefore the TEE before booting the normal world OS. With the TEE established, a FIDO Trusted App can be provisioned to look after key material, crypto and other sensitive operations. This document looks at why the TrustZone based TEE architecture is an excellent fit to the FIDO security requirements and its role as the de facto base-line security technology used in smart devices with integrated authenticators.

The FIDO UAF Password-less Experience

The consumer with a FIDO enabled smart device can register once with their favorite online shopping site or bank. During the registration process the device creates a public and a private key that is specific to the combination of user, his/her device and the relying party. Subsequent visits to the online store then become much easier for the consumer as they can replace the usual username/password authentication step or confirmation of purchase with a quick swipe of a finger or entering a simple and memorable PIN code [Fig 1]. No common user information is shared by the FIDO protocol as its implementation cannot leak private user information. As the relying party only holds the public key it cannot be used directly by hackers to take over accounts if the website's servers are hacked (currently a major problem in the industry).

Copyright © 2015 ARM Limited. All rights reserved.
The ARM logo is a registered trademark of ARM Ltd.
All other trademarks are the property of their respective owners and are acknowledged

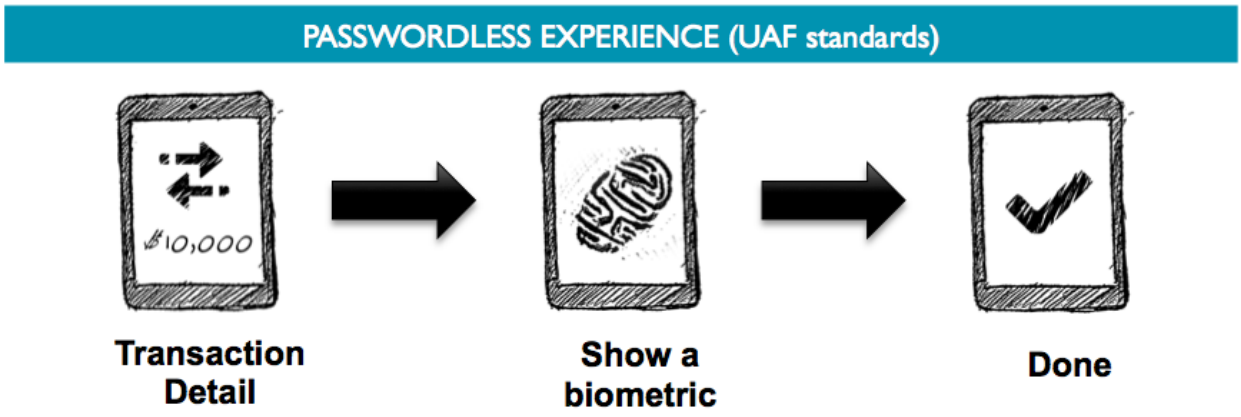

Fig 1. A simple FIDO user experience

Overview of FIDO and the FIDO Alliance

The FIDO alliance comprises more than 180 members covering the whole value chain from major silicon partners (such as Qualcomm), device manufacturers (such as Samsung and Lenovo), operating system companies (such as Microsoft and Google), FIDO server providers (such as Nok Nok Labs) and relying parties (such as Bank of America and PayPal). It develops technical specifications and certification programs to fulfill its mission to enable simpler, stronger authentication. FIDO protocol design is driven from a desire to improve the ease of use of authenticators, built-in privacy, security by design and drive standardization so that relying parties can use any FIDO compliant authenticator. Final FIDO 1.0 specifications are available online and comprise of two alternative user experiences: Universal Authenticator Framework (UAF) provides a password-less experience for devices such as smartphones with built-in authenticators and Universal 2nd Factor (U2F) for a dongle that helps protect traditional username/password against phishing attacks. Work is ongoing to have a unified standard for FIDO 2.0.

Relying parties have been using username and passwords for years but it has proven to be unsatisfactory to both consumers and businesses as passwords have many problems. Consumers like to choose weak passwords or reuse them across websites making it easier for hackers to take over accounts. If they are forced to choose complex passwords, they forget them and transactions may be abandoned. Worse still, passwords are easily phished by scam emails allowing financial fraud on a massive scale: to take one example, Kaspersky estimates that a phishing gang stole $1B from a number of banks last year. Businesses sometimes require use of a second factor such as a One Time Password (OTP) token that typically provides a code to be used alongside the username/password. The often proprietary nature of these tokens has filled consumer's pockets and drawers with hardware: you might have one OTP token for your bank, another for your work email and others for other service providers. Another issue for relying parties using traditional authentication is the need to hold private keys for each customer. These massive databases of credentials create a "honeypot" for hackers who can steal millions of consumer's individual credentials with a single well designed attack. This creates reputational risk for big brand companies who may have to admit to a security breach and ask its customers to quickly reset their passwords.

FIDO mitigates the problems with traditional usernames and passwords and creates a more delightful consumer experience at the same time. For example, on a modern Samsung Galaxy device it is possible to log onto websites or pay for things using your fingerprint. This simple user experience is enabled by FIDO UAF protocol replacing the username/password with a built-in authenticator such as a fingerprint sensor that unlocks a private key on the device that is used in a crypto challenge with the remote server (which holds the public key). The relying party also gains metadata providing some basic information such as the type of authenticator, key protection mechanism used and model of device that can be used in back-end risk analysis. However, no biometric, PIN information or private key is exchanged with the online server. This "Privacy by design" aspect of the FIDO protocols

Copyright © 2015 ARM Limited. All rights reserved.
The ARM logo is a registered trademark of ARM Ltd.
All other trademarks are the property of their respective owners and are acknowledged

provides added protection to the consumer who is less likely to be troubled by security breaches of the stores' server. The crypto challenge is based on well-established Public Key Cryptography principles involving the use of a public/private key pair that is generated on the device for every combination of user/device/relying party. For an overview of the FIDO 1.0 specifications please see here.

The FIDO security requirements can be summarized as:

1. To ensure the integrity of the device
2. To keep key material confidential from unauthorized access
3. To maintain the confidentiality and integrity of sensitive processes

Threat Landscape

Attacks on devices can come in many forms, from malware to social engineering, theft or physical loss of the device, or improperly secured devices either through misuse or by users jail-breaking their devices.

Attacks can be performed by many different methods, and malicious software can be installed by conventional means such as through a rogue app store, via social engineering, trojan or by other attack vectors such as via the browser. When malware is present on a device it has the potential to escape its sandbox or process permissions and any data held or input into the device can then become compromised.

Alternatively, if an attacker can gain physical access to the device, further attacks become possible. If the attacker can access the file system of the device, they can potentially steal data. If the data is encrypted, the attacker could copy the data off the device and perform an offline attack on the encryption. Whilst software attacks are often the main threat, it is important to remember that physical attacks such as opening the device and probing the board become possible if the attacker possesses the phone.

The design of security architecture conventionally relies on two basic concepts: the principle of least privilege, and the partitioning of the system into protected compartments. For example, the TrustZone based TEE is normally designed to maintain its isolation even if the Normal World has been compromised. A malicious hacker may take over the Normal World and spy on communications to the TEE, but the Trusted World will retain its integrity and confidentiality.

TrustZone and the Trusted Execution Environment

GlobalPlatform standardizes the TEE [Fig 3] and generates specifications, compliance programs and certification schemes. They have created white papers providing an insight into the TEE and how it can provide confidentiality and integrity for services such as payment, content protection and dual-persona devices. For the purposes of brevity, only a short description is provided here. A TEE provides a secure enclave to protect sensitive code and data with the security promises of integrity and confidentiality, for example, a malicious application should not be able to read the private keys stored on the device. The TEE is designed to protect against scalable software attacks and if someone has stolen your device, from common hardware attacks sometimes referred to as "shack attacks" (attacks from a knowledgeable attacker with access to normal electronic enthusiast type of equipment).

The TrustZone based TEE provides a "Secure World" where the security boundary is small enough to offer a route to certification and provable security. It is typically used for securing cryptographic keys, credentials and other secure assets. TrustZone offers a number of system security features not available to the hypervisor: it can support secure debug, offer secure bus transactions and take secure interrupts directly into the Trusted World (useful for trusted input). There is an argument to restrict the amount of security functionality in the trusted world to limit the attack surface and make

Copyright © 2015 ARM Limited. All rights reserved.
The ARM logo is a registered trademark of ARM Ltd.
All other trademarks are the property of their respective owners and are acknowledged

certification a practical proposition.

The TrustZone security extensions work by providing the processor with an additional 'secure state' that allows secure application code and data to be isolated from normal operations. This partitioning enables a protected execution environment where trusted code can run and have access to secure hardware resources such as memory or peripherals. Conventionally, the Trusted World is used with its own dedicated secure operating system and a trusted boot flow to form a TEE that works together with the conventional operating system, such as Linux® or Android™, to provide secure services.

Security is as strong as the weakest link in a chain of trust. The starting point of the chain is the Root of Trust (ROT) that is normally implemented in hardware to protect it from modification. Mobile device integrity starts by resetting into Secure World and booting from immutable hardware in the form of a Read-Only Memory and accessing trusted hardware resources such as hardware unique key, random number generators, counters, timers and trusted memory. A carefully designed authenticated trusted boot flow is the basis for device integrity. The Trusted OS is started as part of the trusted boot flow before the Normal World Rich OS is booted.

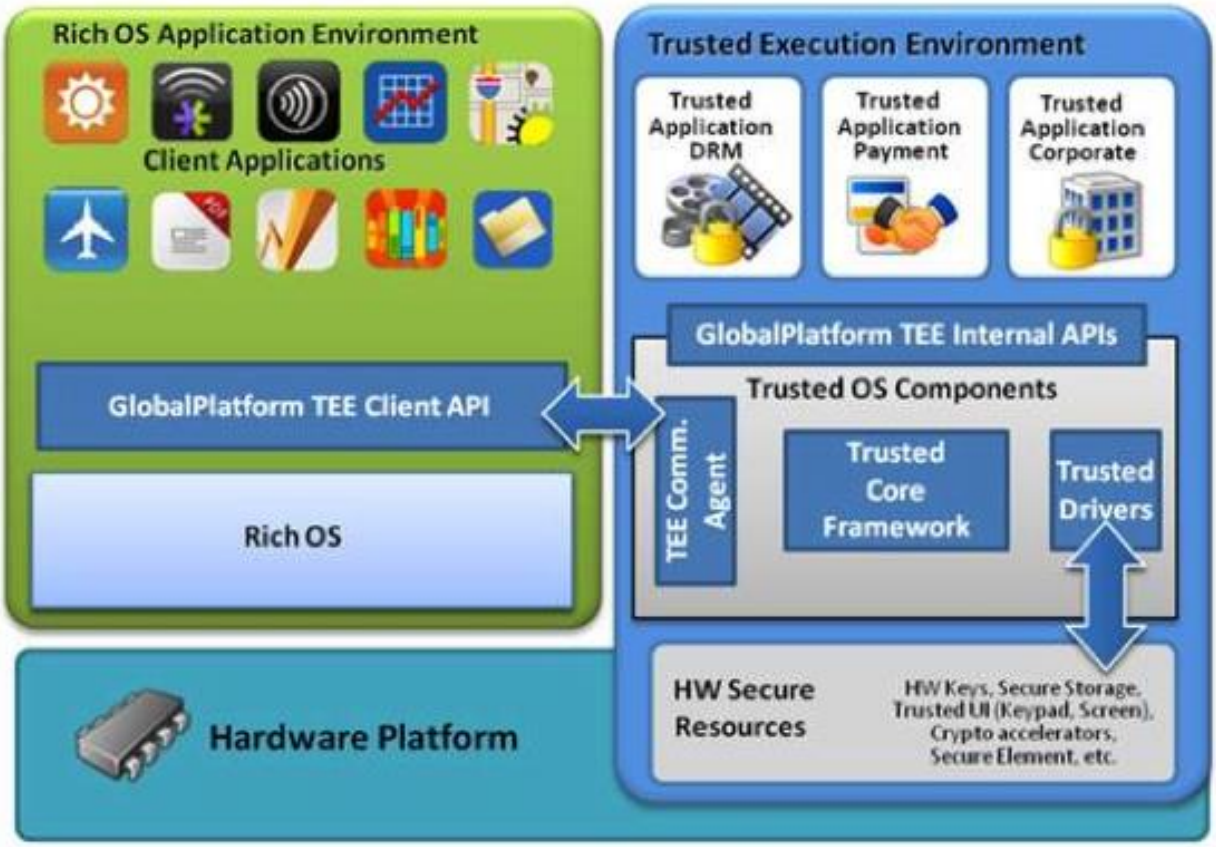

Fig 3. TrustZone can provide the hardware partitioning for a TEE and access to secure resources

Specific role of the TrustZone based TEE in FIDO implementations

The TrustZone (Secure World) based authenticated trusted boot flow and hardware ROT provides the basis for device integrity (a reference implementation of trusted boot can be found here). The Trusted OS can provide trusted services for the FIDO protocol, for example, handling cryptography and user matching algorithms in a hardware protected execution environment. In a typical implementation, nearly all of the FIDO stack will reside in the normal world and only small security sensitive functions are moved into the

Copyright © 2015 ARM Limited. All rights reserved.
The ARM logo is a registered trademark of ARM Ltd.
All other trademarks are the property of their respective owners and are acknowledged

TEE. The code moved to the TEE is referred to as a Trusted App as it benefits from the security promises of confidentiality and integrity. This partitioning builds in resistance to scalable attacks. A major use case of the TEE is to provide a secure key store. Since non-volatile memory is rarely found on an applications processor, FIDO keys are encrypted in the TEE with a hardware unique key burned into the chip. This encrypted and wrapped key is then stored in external memory for storage between boots. Keys would only be decrypted and used within the TEE and never accessible to the Normal World.

A FIDO Trusted App could include the functionality for biometric template storage and matching. This could be handled in a similar way to the storage of crypto keys i.e., encrypted and wrapped within the TEE and stored in external non-volatile storage.

The TrustZone based TEE provides solutions to the FIDO security requirements:

1. To ensure the integrity of the device:
   This is achieved using hardware roots of trust and a TrustZone isolated authenticated trusted boot process.
2. To keep key material confidential from unauthorized access:
   The systemwide hardware isolation provided by the TrustZone architecture extensions enables a small, security certifiable TEE to handle key materials. FIDO keys can themselves be encrypted using strong cryptography and fused Hardware Unique Keys.
3. To maintain the confidentiality and integrity of sensitive processes:
   The TEE provides the security promises of integrity and confidentiality. Typically, small parts of the FIDO process will be statically partitioned into the Trusted World and run as a Trusted Application.

   Please see the Future enhancements section for future devices with a Trusted User Interface (TUI)

4. To maintain the confidentiality of sensitive input data:
   TrustZone enables interrupts from input devices (such as authenticators) to be steered directly to the Trusted World where trusted device drivers can handle them. For example, the TEE can handle touch events from a touchscreen during PIN capture or interrupts from a fingerprint sensor and separate it from malware in the normal world that would not be able to intercept it. When the PIN capture or other input is complete the interrupts can be switched back to the normal world.
5. Protection of sensitive display data:
   TrustZone can be used to protect a Trusted World frame-buffer and its composition. This enables a “what you see is what you sign/buy” feature since the frame-buffer cannot be intercepted, modified or obscured.

Future enhancements

GlobalPlatform has developed a protection profile for the TEE that is being used as the basis for a security certification program. Multiple test labs are establishing programs to test platforms and evaluate the effectiveness of the TEE they contain. Independent testing will assure device manufacturers of the quality of solutions that may be beneficial to the whole value chain. Security certification is expected to be available from the second half of 2015.

Modern ARM-based chips are making increasingly sophisticated use of TrustZone technology. One example is the use of a TUI to protect touchscreen inputs and the display of protected frame buffers [Fig 4]. It is possible to have peripherals that can switch between normal world and secure world: the touchscreen and display are examples where this might be beneficial. In PIN capture mode the TEE may want exclusive trusted access to the touchscreen which can be returned to the normal world when PIN capture is finished. The Display Processor may be acting as a compositor for the various graphics layers and required to display trusted data from the Secure World to ensure “what you see is what you get/sign”. Trusted display data can be generated in a (TrustZone) protected frame buffer and composed as a secure layer with protection against overlay. Adoption of the TUI is expected to increase when standardization

Copyright © 2015 ARM Limited. All rights reserved.
The ARM logo is a registered trademark of ARM Ltd.
All other trademarks are the property of their respective owners and are acknowledged

from GlobalPlatform is completed.

In addition to the TrustZone based TEE, a modern mobile device may have a number of secure elements owned by different parts of the value chain. The SIM card may be owned by the operator, the OEM may have its own SE and the OS may require access to a SE for holding keys or performing system integrity checks. As secure elements do not have access to an input method or display it can be beneficial to establish secure communications with the secure element from the TEE. GlobalPlatform is working on the standardization of communication between a secure element and the TEE.

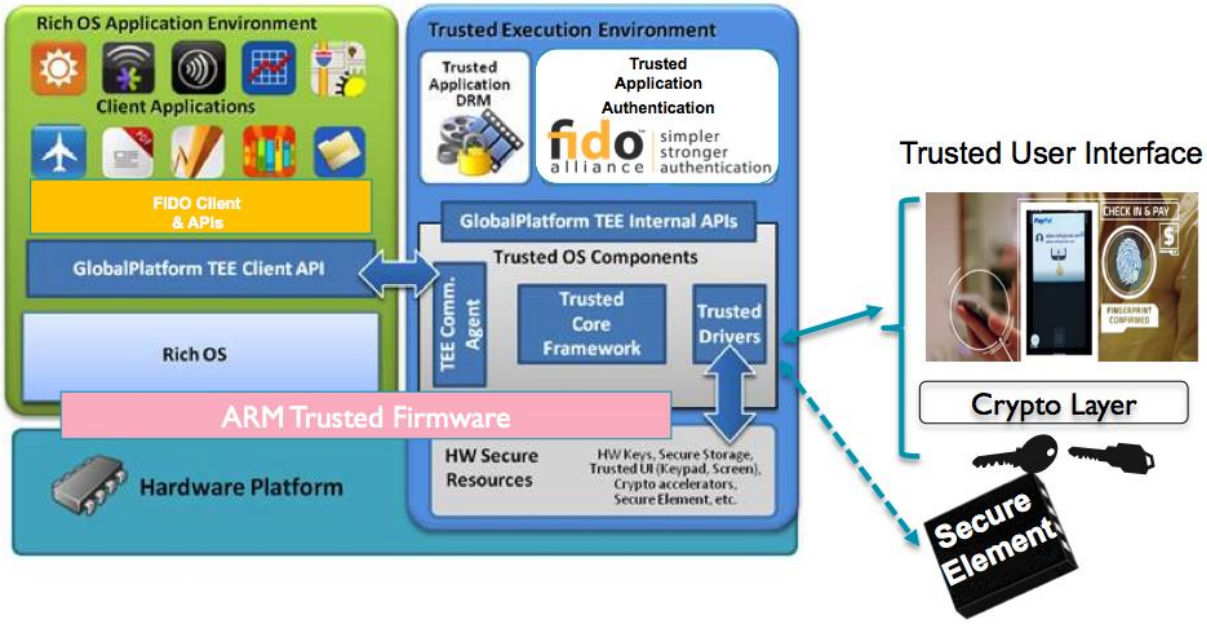

Fig 4. TrustZone based TEE with FIDO Trusted App, Trusted User Interface & encrypted channel to a secure element

Conclusion

The TrustZone based TEE delivers effective system security at low cost for FIDO implementations. A well-designed TEE provides a suitable level of security for FIDO based implementations and is a huge improvement over the username/password normal world methods it is replacing.

In the future we can expect further improvements. Device manufacturers and silicon partners will have the option to have their TEE's security certified by independent test labs. We can also expect TrustZone technology to be extended to cover touchscreen input (for protecting PIN entry) and display output providing a "what you see is what you sign/buy" capability.

FIDO based authentication is already deployed at scale and looks set to become an industry success story by helping consumers move beyond passwords. The TrustZone based TEE demonstrates that when security is well architected it can deliver delightful user experiences.

Copyright © 2015 ARM Limited. All rights reserved.
The ARM logo is a registered trademark of ARM Ltd.
All other trademarks are the property of their respective owners and are acknowledged