THALES

www.thales-esecurity.com

Thales e-Security

Thales Security World

A Secure Key Management Architecture for the
Thales nShield Family of Hardware Security Modules

White Paper

October 2015

Are you:

- Looking for ways to simplify the key management process in your enterprise
- Interested in maximizing the capabilities of your deployed nShield HSM base
- Concerned over how to manage risks and preserve resiliency and scalability

Learn how to:

- Design an HSM deployment architecture that is optimized to your needs and environment
- Map your security policies to a flexible hardware protection infrastructure
- Achieve total lifecycle management of your enterprise keys and cryptographic policies

Contents

1. Introduction

The explosive growth in the volume of data and the rise of distributed IT environments are making organizations today more vulnerable than ever – leading to data breaches that can have profound effects on their bottom-line and reputation. As a result, cryptography has become the foundation for many security initiatives to protect the confidentiality and integrity of critical data assets.
This brings into sharp focus the trustworthiness of those cryptographic operations and the processes by which cryptographic keys are managed. Ultimately, sensitive business data and private information is only as secure as the cryptographic keys and processes used to protect them.

To address this issue, leading manufacturers like Thales have developed dedicated hardware security modules (HSMs) to protect the storage, management and use of keys within a tamper resistant secure environment to achieve a level of security that is unattainable with purely software-based applications. According to John Pescatore in the SANS Whitepaper, *Using Hardware-Enabled Crypto to Thwart Advanced Threats*, September 2015, "To effectively combat continually advancing threats while also meeting business demands, computing environments must evolve to add trusted hardware to protect (and accelerate) critical security functions."

A simplistic view of an HSM is a "walled garden for crypto" or a "crypto lock-box" where keys and the operations that use those keys are hidden and protected, away from prying eyes. This has some appeal but raises important operational challenges. The same issues of availability, scalability and recoverability apply equally to HSM-based operations as they do across the general IT infrastructure. Simply locking keys away will provide security but at the cost of flexibility, performance and control. The challenge is to deliver measurable and certifiable security in a way that is easily integrated and exploits common operational practices, and that can evolve with the needs of the organization. It is this challenge that the Thales Security World key management architecture is designed to address, and has been proven to satisfy.

The Thales Security World architecture supports a specialized key management framework that spans the entire nShield family of general purpose HSMs. Whether deploying high performance, shareable, network-attached HSM appliances, host-embedded HSM cards or USB-attached portable HSMs, the Security World architecture provides a unified administrator and user experience and guaranteed interoperability whether the customer deploys one or hundreds of devices.

Security World overcomes the limitations and management costs that would result from the simplistic lock box concept that considers an HSM to be an isolated, closed device. By adopting a broader perspective, Security World enables HSMs to be deployed with a high degree of flexibility and configured in any combination to meet an organization's operational, security and budgetary needs.

This paper describes the approach undertaken by Thales to provide total key lifecycle and policy management for HSMs and explains in detail the design philosophy and operational benefits of the Security World key management architecture.

The Security World architecture provides a unified administrator and user experience and guaranteed interoperability whether the customer deploys one or hundreds of devices.

2. HSM Key Management Without Compromise

Thales Security World addresses the age old challenge of providing strong protection for keys while at the same time ensuring they are available for use by authorized applications that are deployed over high scale, redundant and distributed server infrastructures.

Security World removes the need to directly access HSMs for many key management functions which reduces operational costs without compromising security.

The Thales Security World architecture provides a business-friendly methodology for securely managing and using keys in real world IT environments. Field proven by thousands of customers over more than a decade, and subject to numerous international security certifications, Security World minimizes the strain on specialist security resources and instead takes advantage of existing data management processes. This drives down the cost of ownership for HSMs while building resiliency and ensuring long term availability of keys. Thales Security World enables security architects to design a comprehensive security system that enforces policies and that governs the creation, use, and, in the event of disaster, recovery of key material – all within the security-certified, tamper-resistant environment of one or more nShield HSMs.

Security World enables key management across an unlimited population of HSMs by creating a logical security boundary that extends beyond the hardware. Within this boundary, keys can be safely managed and provisioned. Very importantly this means that many operational aspects of key management can be safely performed without directly accessing the HSMs – dramatically reducing operational costs without compromising security.

To appreciate the true power and flexibility of this approach it is useful to consider the trust models and management practices that traditionally relate to HSM deployments and the business applications that they serve. HSMs are fundamentally used to create a "trusted layer" where key material can be stored and managed and cryptographic processes can be performed safely. This trusted layer overcomes the fact that the software environment in which applications execute and are managed is not in itself sufficiently trusted. HSMs are used selectively to create a higher level of trust specifically for the most sensitive or most closely scrutinized digital assets and processes.

However, HSMs are specialist security devices requiring a relatively high level of expertise to administer. Management processes are closely supervised, even ceremonial in nature, driving up operational costs and slowing down response rates compared to less secure software deployments. The goal of the Security World architecture is to minimize this impact without compromising security.

The concept of a high-trust “HSM Layer” relative to the “Application Layer” where traditional security controls are in place is illustrated in Figure 1. The Application Layer contains existing server and data storage hardware, virtualization environment, application software and application data – all the elements that would be in place in a software-only deployment where HSMs are not being used. This application zone is managed by existing data and security management tools and processes to provide data backup and recovery, provisioning and replication. The HSM Layer on the other hand, tends to be isolated from these systems and instead is managed by specialists – resources that are generally less readily available in the organization. In order to minimize operational cost it is clearly beneficial, wherever possible, to perform key management tasks within the existing Application Layer rather than in the relatively expensive HSM Layer – providing that the same levels of security can be assured.

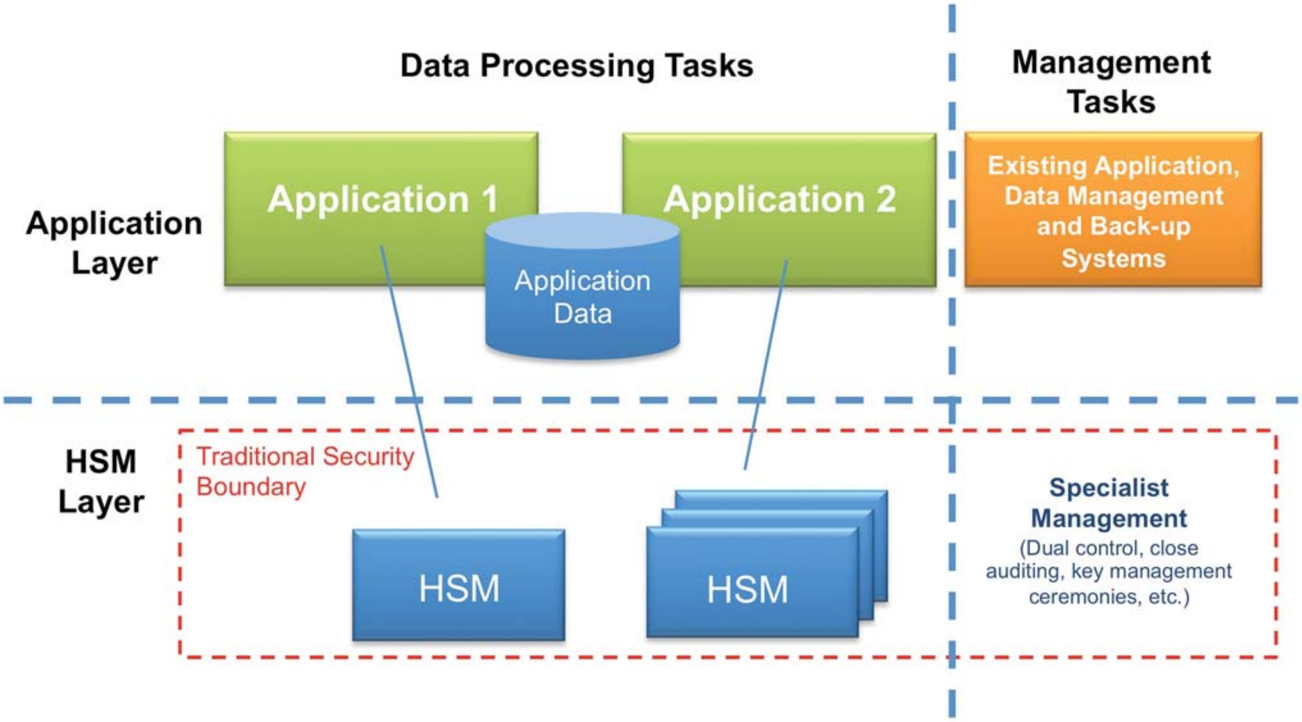

Data Processing Tasks
Management Tasks
Application Layer
Application 1
Application 2
Application Data
Existing Application, Data Management and Back-up Systems
HSM Layer
Traditional Security Boundary
HSM
HSM
Specialist Management
(Dual control, close auditing, key management ceremonies, etc.)

Figure 1 – Distinction between Application and HSM Layers and associated management domains.

Security World enables the bulk of key management tasks to be performed safely within the Application Layer. It achieves this by making a clear distinction between three different classes of keys. These three classes of keys will be referenced throughout this paper and are listed below:

- **Application Keys:** These keys are protected within the HSM security boundary and used by business applications to perform a range of cryptographic operations.
- **Control Keys:** This class of optional keys is used to authorize the use of Application Keys in order to enforce security policies.
- **Infrastructure Keys:** These keys underpin the security of the HSM and Security World architecture and act as the basis of trust between HSMs.

Security World enables the majority of key management tasks to be performed safely within the Application Layer to help minimize operational cost.

3. Security World In Action

The primary aspects of most HSM deployments fall into three general areas – the trustworthiness of the hardware platform, the available tools to manage keys and the methods to enforce security policies – all within traditional, multi-tenant and virtualized environments. The following sections describe the role of the Security World architecture in the practical context of these areas.

Security World minimizes the impact of key management without compromising security.

3.1. Building a Trusted Platform

At the heart of any Security World deployment is the use of nShield HSM hardware. These certified security devices provide a high level of physical security and tamper resistance that is simply unattainable in purely software-based systems. Thales nShield HSMs are available in three form factors (Figure 2) to suit different host scenarios.

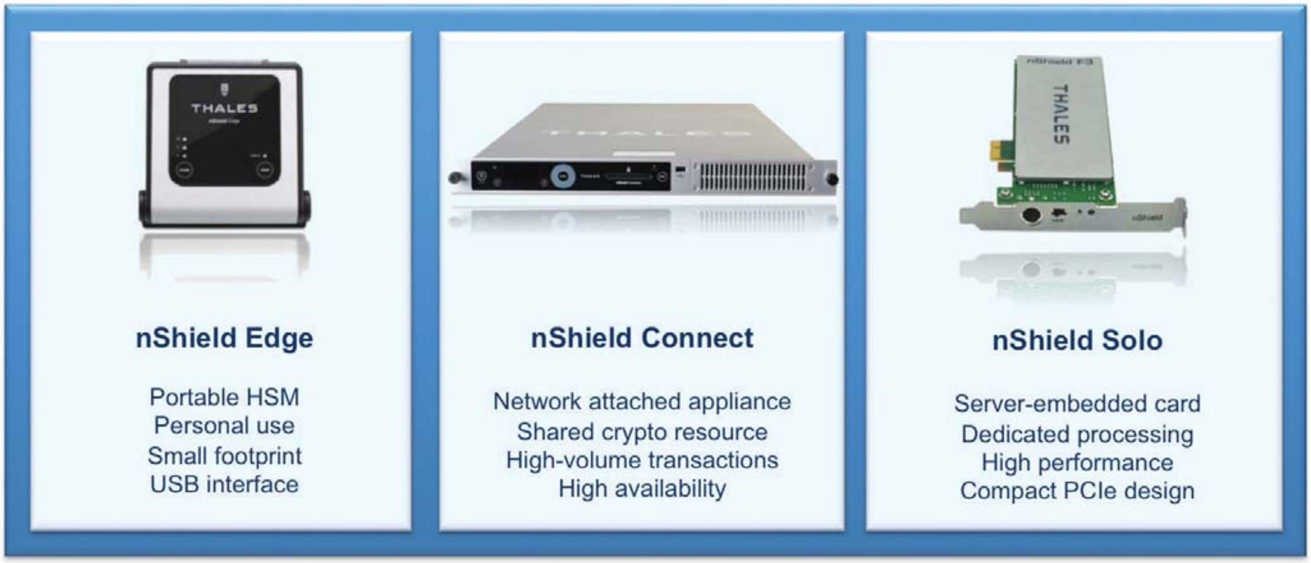


Figure 2 – Thales nShield family of multi-purpose HSMs.

Most nShield HSMs support the Thales' CodeSafe technology that enables customer-specific application code to execute within the HSM. This enables trusted applications running within the protected FIPS-certified boundary of the HSM – each having exclusive access to the relevant Application Keys.

nShield HSMs can be grouped together in any combination to deliver higher levels of processing capacity or resilience. During HSM initialization, the logical associations and trust relationships between each HSM in the group are established. This involves the creation and sharing of a series of Infrastructure Keys across these HSMs. Infrastructure Keys are used to define groupings of nShields. HSMs within the same group are referred to as being in the same Security World. Infrastructure Keys are critical to system availability and recovery, and are securely backed up, shared and managed using a common set of smart cards known as the Administrator Card Set (ACS).

To achieve high levels of security and operational convenience this set of cards supports the use of a quorum technique for authorizing operations including adding new HSMs to the group and for the recovery of HSMs or keys. Achieving a quorum requires a defined chosen number of smart cards (k) from a total set (n) be brought together for an operation to be authorized. Each administrator card can be created with an optional pass phrase, which must be entered when the card is used.
In addition to a pre-defined universal quorum, sub-quorums can also be defined for specific administrative tasks. Customers can enforce their policy by choosing values for k and n depending on the level of mutual supervision that is required. Once initialized, all HSMs in the group are able to perform operations on behalf of the same applications using the same Application Keys and to enforce the same security policies.

3.2. Managing Application Keys

The secure management of Application Keys is a core value proposition of any HSM. The Thales Security World architecture supports a number of powerful capabilities to manage these keys throughout their lifecycle including secure generation, storage, backup and recovery.

Application Keys are those keys that are associated with and used by specific business applications. In a pure software deployment these keys would exist in "wallets" and other key repositories within the operating system or application itself.
In an HSM deployment the Application Keys are protected within the HSM, used within the HSM and referenced by the application using key "handles" as these operations are required. Examples of Application Keys are private signing keys associated with an e-invoicing application, private keys used as part of an SSL handshake process, a secret symmetric key used to encrypt credit card numbers or a master key used to protect subordinate column-level encryption keys in a database encryption deployment.

In order to establish trust in Application Keys, they are typically generated within the HSM using a security-certified random key generation process (in some cases it may be necessary to import legacy keys into the HSM). New Application Keys are required when new applications are introduced or as keys are rotated or refreshed. As a result, in many deployments, the population of Application Keys can grow quickly. As a result of the dynamic nature of encryption keys, it is often necessary to retain historic Application Keys even after they are no longer in use, once again increasing the number of keys under management.

With the need to share Application Keys with an increasing number of HSMs in near real time, it becomes highly advantageous to abstract the Application Keys from the individual HSM devices. To enable this level of abstraction to be performed safely, the Security World architecture incorporates a secure tokenization technique. This tokenization process creates encrypted Application Key Tokens that protect the Application Keys and tightly bind them to the policies that define their use through an associated access control list (ACL). The ACL enables very fine-grained control by defining the operations that can be performed by individual Application Keys, for example to distinguish between different types of operations (encryption, decryption, signing etc.). In addition, various threshold techniques are also available to increase flexibility. For example, the ACL can be used to specify how many times a particular operation can be performed, or to specify a particular time period during which approved operations can be performed.

By abstracting Application Keys as Application Key Tokens, the security boundary of the HSM is effectively expanded from the physical boundary of the HSM to a logical security boundary that encompasses all HSMs in the Security World group.

It is important to understand that Application Key Tokens while outside the physical boundary of an HSM are still protected by the expanded logical boundary and represent no weakening of the certified protection of the HSM. The fundamental premise of any HSM is that keys should never be stored or exposed, even transiently, in the clear – outside the secure confines of the HSM. The nShield HSM with the Security World architecture absolutely adheres to that principle – clear text keys never leave the HSM. Application Key Tokens can only be used to perform cryptographic operations after the Application Keys are reconstituted inside the HSM's physical boundary via the de-tokenization process which is itself controlled and protected by the HSM and the Infrastructure Keys that underpin its trust model.

Security World establishes an expanded logical security boundary to protect Application Key Tokens.

As shown in Figure 3, abstracting the Application Keys from a specific nShield device as Application Key Tokens creates the ability to safely manage and provision the Application Keys within the Application Layer rather than the relatively expensive HSM Layer. This capability represents an attractive proposition and is frequently described by customers as one of the core strengths of the Security World architecture. The capability to abstract Application Keys from the constraints of the individual HSM delivers many of same benefits that accrue from the use of virtualization in the application space – higher scalability, greater flexibility, simplified resilience and reduced operating costs.

For a more in depth look at the process and the steps taken to securely tokenize application keys see Appendix A.

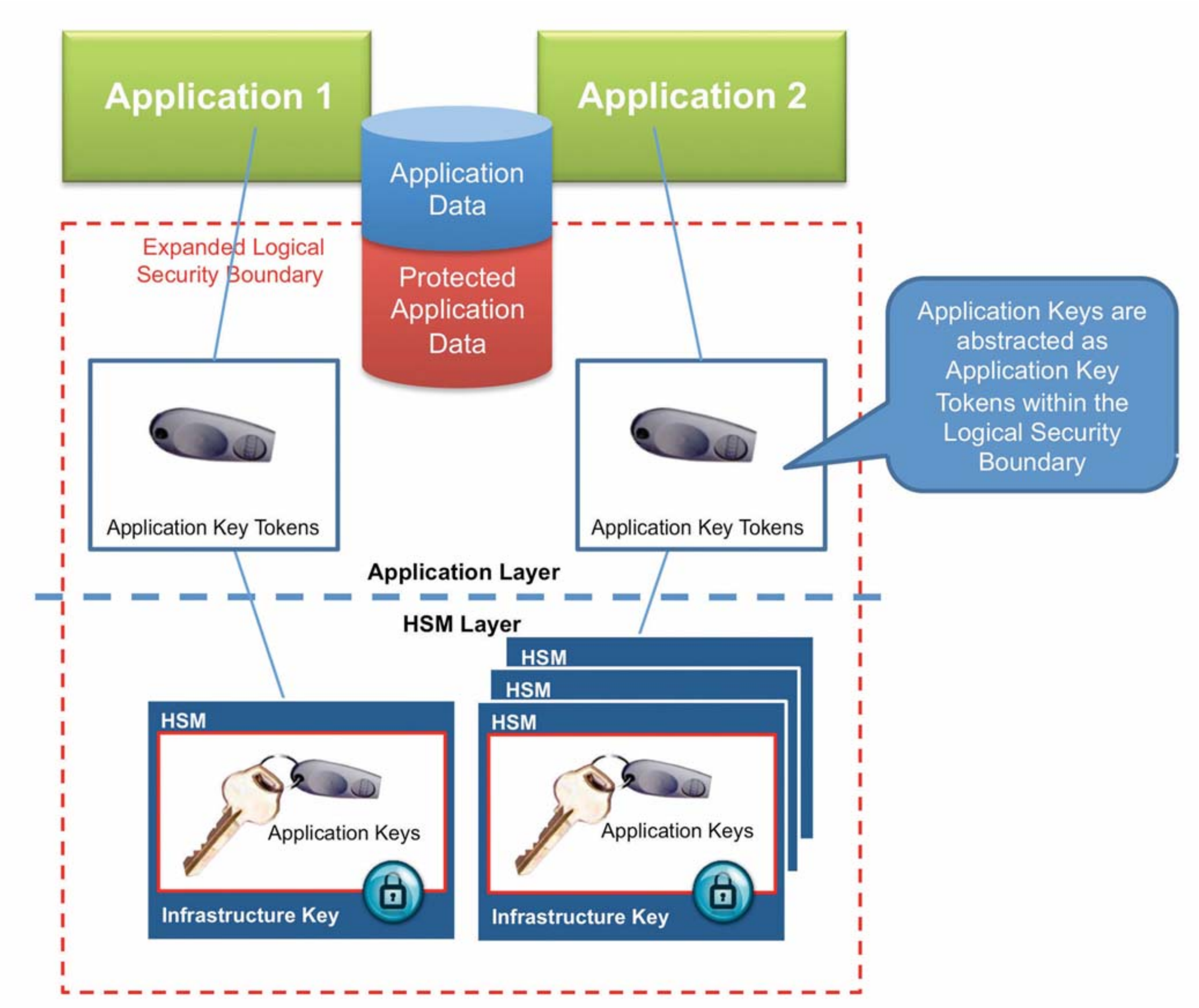


Figure 3 – Abstraction of Application Keys into Application Key Tokens.

3.3. Controlling the Use of Application Keys

One of the primary benefits of using an HSM is the addition of layers of strongly-enforced authorization to use specific Application Keys. The Security World architecture supports a number of mechanisms to achieve this. Cryptographically, this is enforced by a class of keys called Control Keys which can be considered to be authorization credentials.

The appropriate method for controlling the use of each Application Key is selected at the time the key is created and is applied every time that specific key is used. Control Keys are fragmented and provided to one or more security officers as a set of smart cards, known as Operator Card Sets (OCS). These card sets can be combined to enforce quorum-based authorization policies in the same way that a quorum of ACS cards is used to bring a high level of assurance to the management of Infrastructure Keys. There can be one or more sets of OCS in the Security World, each of which can be used to control the use of one or many Application Keys. OCS cards can be presented directly at the HSM front panel or remotely using a capability known as Remote Administration. When a quorum of OCS cards has been presented, the Control Key is reconstituted inside the HSM. For a more in depth look at the secure remote administration of nShield HSMs see Appendix B.

The process of fragmenting Control Keys across populations of OCS cards creates natural redundancy and resiliency. While an OCS card recovery process is available, the loss of an OCS card simply narrows the set of cards from which a quorum can be formed. Should sufficient OCS cards be lost that a quorum can no longer be formed, control over the Application Keys can be transferred, by authorization with the ACS, to a new Control Key on a new OCS.

Control Keys can be used to enforce authorization scenarios from manual authorization of individual cryptographic operations to fully automatic, on-demand authorization. When choosing an authorization policy it is necessary to balance the need for strong authorization with the operational resilience and performance requirements. For example, use of a key to issue a certificate may require per use manual authorization, while the use of a key to initiate an SSL handshake would require automatic authorization. When the key is created the authorization policy must be defined. Security World enforces policies through the use of different authorization modes as described below:

1. **Multi-factor per-transaction authorization:** In-person authorization by one or more authorized personnel requires presenting a quorum of Operator Cards belonging to an OCS. The quorum for each card set (which could consist of a single card) is uniquely assigned when the card set is initially created by the customer to suit the level of multi-party or dual control that is required. Application keys protected in this way are described as "card set protected keys". In order to achieve multi-factor authorization, each card in an OCS can optionally be associated with a unique pass phrase to ensure the card can only be used by the legitimate cardholder.

2. **Pre-authorized operation:** In order to support unattended and automated operations in environments with high transaction volumes, pre-authorization is necessary. A number of pre-authorized modes are available as described below:

   a. **Basic pre-authorization:** Designed to provide essential key security while fully automating key usage authorization processes. Application Keys controlled in this way are described as "module protected keys".

      While this policy is a popular choice for many deployments where HSMs are installed in a physically secure environment or for demonstration and test purposes, customers often consider one of the following three modes to offer enhanced security without compromising operational availability or ease of use.

   b. **Revocable pre-authorization:** Suited for unattended operation where a single card must be left perpetually in the HSM. This card authorizes all subsequent operations until the card is physically removed, essentially creating an instantaneous "kill switch" for physically disabling all HSM operations in the event of system maintenance or suspected breach. The use of the HSM cannot resume until the card is re-inserted.

      This approach is particularly useful when contractors are present in a data center or where HSMs must be shipped between facilities via an untrusted courier, since an attacker who is able to intercept the HSM cannot use any keys unless the smart card is present.

   c. **Start-up pre-authorization:** Designed for unattended operation where a quorum of OCS must be presented every time an application is started but where all cards can be removed after start-up.

      This approach can be used in an office environment where applications are shut down each evening and where authorization needs to be provided each morning. In the event that an HSM is physically stolen, an attacker cannot misuse any key without the quorum of smart cards.

   d. **Segregated pre-authorization:** Builds on the security provided by module protected keys by requiring a pass phrase to be provided each time a key is used but without requiring a physical smart card to be manually inserted.

      This approach is particularly useful where multiple hosts, applications or users share a single HSM concurrently (so called multi-tenancy or partitioning), since it allows individual users or applications to be strongly authenticated and then segregated by each application utilizing its related Control Keys.

3.4. Segregating Application Keys

Building on the segregated pre-authorization mode described above, groups of Application Keys can be restricted to work with selected applications that use them. This concept is often referred to as partitioning. This form of multi-tenancy enables client applications to be granted exclusive access to sensitive resources, thereby allowing multiple classes of business applications to share a single HSM or group of HSMs.

Security World supports this form of segregation of HSM resources through the use of pass phrase-protected Control Keys. These enable administrators to access keys at an organizational level, allowing multiple user applications to share network-connected HSMs. A schematic representation of the partitioned resources of the nShield HSM in this type of environment is shown in Figure 4.

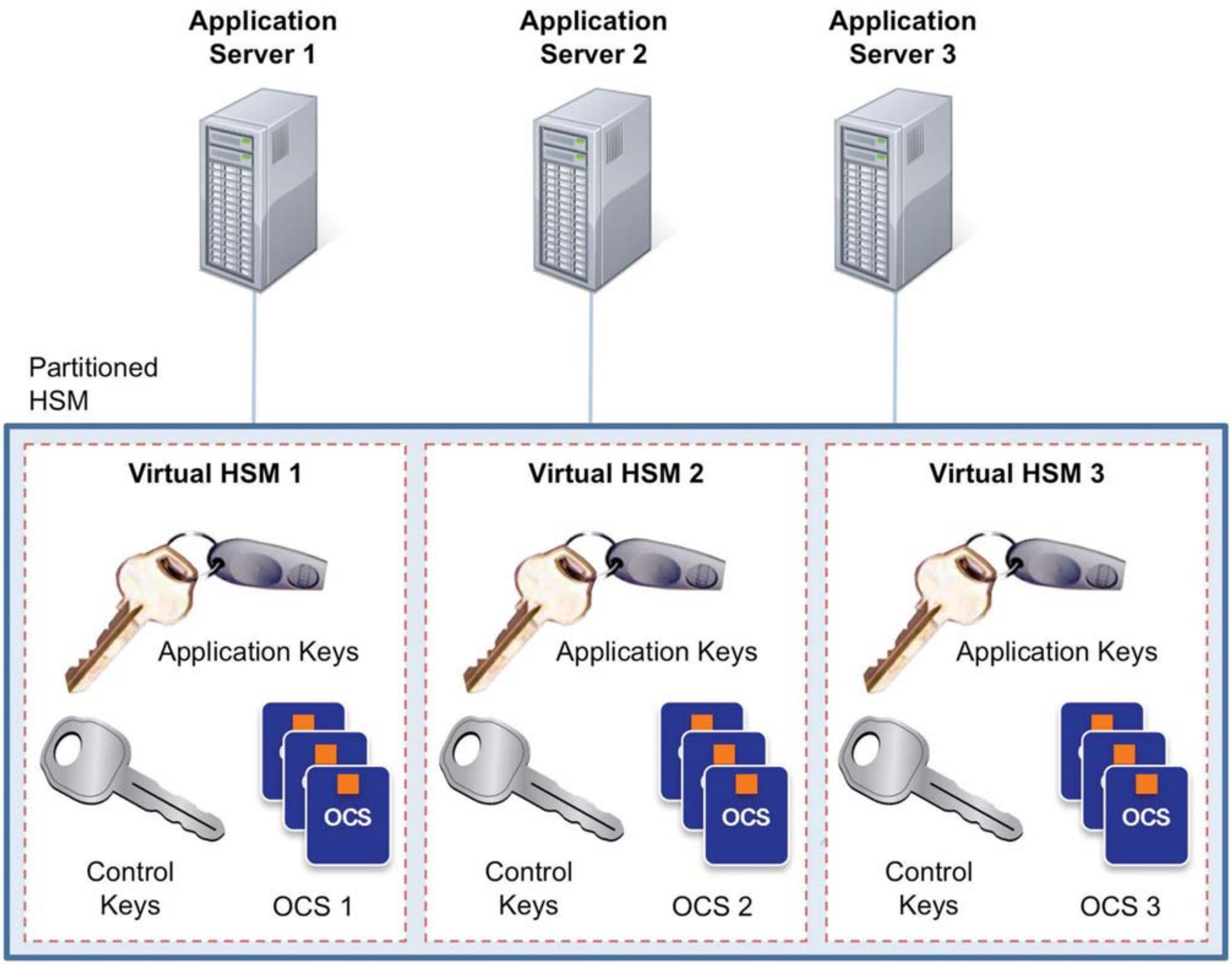


Figure 4 – Segregated HSM resources and associated Application Keys.

3.5. Deploying Security World

As described earlier in this paper, nShield HSMs all define the physical FIPS-certified security boundary or HSM Layer within which Application Keys, Control Keys and Infrastructure Keys are protected. Using quorums of ACS cards, Infrastructure Keys can be securely backed up and shared across multiple HSMs. When this is performed, HSMs that share the same Infrastructure Keys develop a common Security World that provides an expanded logical security boundary that extends beyond the physical HSM Layer and overlaps into the enterprise IT environment or Application Layer. The abstraction of Application Keys into Application Key Tokens enables these tokens to be stored outside the physical HSM and within the corporate IT environment within the Expanded Logical Security Boundary. A typical scenario depicting how the Thales nShield HSMs and Security World are deployed within an enterprise environment is shown Figure 5.

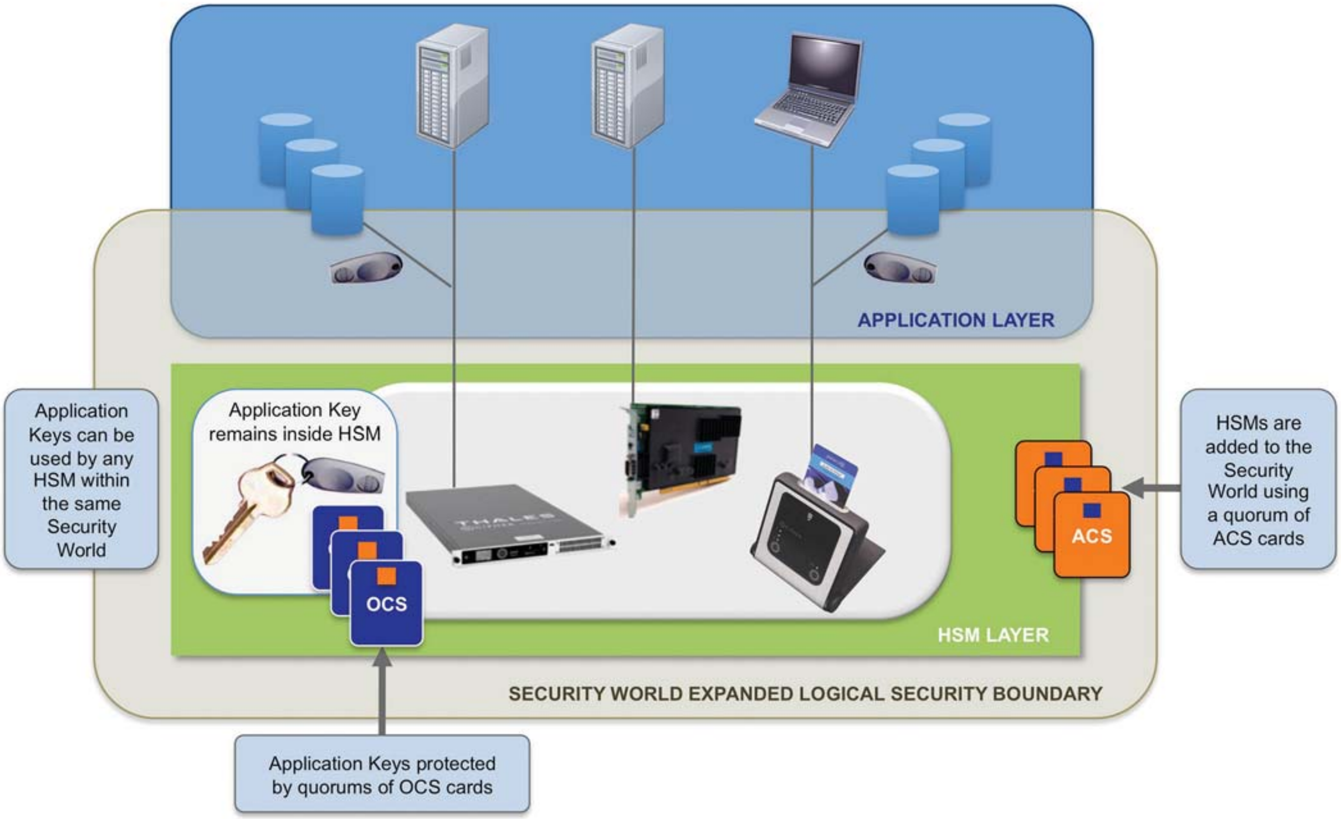

APPLICATION LAYER
Application Keys can be used by any HSM within the same Security World
Application Key remains inside HSM
OCS
ACS
HSMs are added to the Security World using a quorum of ACS cards
HSM LAYER
SECURITY WORLD EXPANDED LOGICAL SECURITY BOUNDARY
Application Keys protected by quorums of OCS cards

Figure 5 – Thales nShield and Security World deployment.

4. The Power of the Security World Architecture

Thales Security World provides a series of unique capabilities and benefits under the general headings of enhanced security, reduced operational costs, increased resiliency and superior flexibility and scalability. This section describes these benefits in greater detail.

4.1. Enhanced Security

- **High assurance controls for HSM administration:** All administrative functions associated with the HSM require the use of strong smart card-based credentials which can be combined to create quorums to establish mutual supervision and dual control, for example requiring 3 out of 5 security professionals to come together to instantiate a sixth HSM into a Security World group of 5 HSMs.
- **Strong authorization to control use of Applications Keys:** Unlike competitive solutions, Security World provides mechanisms to enforce event-by-event authorization of Application Key use, either by single individuals or based on a quorum of security professionals working collectively. This is particularly important in cases of high-value digital signing, credential issuance or key recovery. Alternatively, policies can be set to allow keys to be used automatically, for example in high performance encryption or transactional systems.
- **Powerful separation of duties:** Application Keys are abstracted from the HSM and separated from the Control Keys dictate their specific use. Since Control Keys are tied to the HSM devices and managed within the HSM Layer, this creates a powerful separation of duties between application and HSM administration staff, avoiding the creation of "super-users" and often dramatically simplifying compliance reporting.
- **Segregation and partitioning of applications:** By carefully associating Control Keys with individual application instances and groups of Application Keys, it is possible to achieve powerful, cryptographically-enforced segregation between these entities – creating isolated application security domains. This cryptographic approach provides a means to logically partition HSM resources in an auditable fashion. This increases the range of deployment options and in some cases can reduce the number of physical devices that need to be deployed.
- **Distributed backups reduce the risk of attack:** Control Keys and Infrastructure Keys are backed up as multiple key fragments, exported to smart cards and dispersed to offsite locations where physical security cannot be guaranteed. To launch an attack purely from backup information, the attacker would need access to a quorum of cards and their associated pass phrases for Control and Infrastructure Keys, a functioning HSM and the various Application Key Tokens – all of which might be managed by different individuals.

Security World avoids the need for expensive backup tokens and manual key cloning.

This is in contrast to other less secure approaches in the market that place all key backups onto single physical tokens – effectively surrogate HSMs. These represent a single point of attack particularly while in transit to offsite storage. Even if a backup HSM is suspected of being attacked, all keys are lost.

4.2. Reduced Operational Costs

- **Automated backup:** By abstracting Applications Keys as tokens, they can be safely managed by existing and automated data management and backup processes – quite literally by the same file management processes that already back up and recover the application software and application data. Compared to other approaches in the market that require manual cloning of all keys between interconnected HSMs, the adoption of the Security World architecture dramatically reduces the cost of key management for the majority of keys associated with the HSM.
- **Low cost backup tokens:** Backups of Control Keys and Infrastructure Keys benefit from the security provided by geographically dispersing key fragments, thereby enabling the use of low cost devices such as smart cards. Competing solutions utilize cloned HSM devices for backup, therefore requiring this single backup token to be effectively as secure as the operational HSM being backed up, driving up capital costs and operational costs on a per-backup basis.
- **Simplified provisioning of Application Keys:** Managing Application Key Tokens within the Application Layer means that day to day provisioning tasks such as making keys available to new instances of applications becomes trivial and can be performed safely by non-security oriented IT staff - resources that are generally more available and less costly than security staff.
- **Application independent key management:** All Security World functionality is independent of the application that the HSMs serves. Therefore administrators supporting multiple HSM deployments or business applications (data encryption, PKI, code signing, etc.) have a consistent experience across all devices, reducing complexity and training requirements.

By abstracting Application Keys, the size of the pool of available HSMs can be tuned dynamically to satisfy changing performance requirements without the need to clone application keys between HSMs.

4.3. Increased Resilience

- **Frequent and timely backups:** It is well known that manual HSM cloning techniques result in fewer backups being made and therefore reliance on backups that risk being out of date. A significant advantage of the Security World architecture is that existing data and application backup processes are used to back up Application Key Tokens. This means that backups automatically fall in line with corporate backup policies – typically a nightly backup regime. Not only does this mean that Application Keys are backed up far more often, but also that Application Key backups are naturally synchronized with application and data backups - making recovery a less complex task.
- **High availability and rapid response to failure:** By abstracting Application Keys they are immune from HSM hardware failure or any other factor that could make the HSM unavailable. Replacement HSMs can be introduced at a moment's notice without the need to manually import or clone potentially thousands of Application Keys. Once a new nShield HSM is enrolled into the prevailing Security World, it can access the same Application Key Tokens as other HSMs in the same logical world and be up and running immediately. By sharing access to the set of Application Key Tokens, creating a high-availability pair of HSMs becomes trivial.
- **Backups that can be relied on:** By backing up Control Keys and Infrastructure Keys using quorums of smart cards there is no single point of failure in the backup chain. Lost or broken smart cards merely narrow the card set, for example a policy that required any 5 cards from a set 9 to recover a backup would now require 5 cards from the remaining 8. HSM backup mechanisms that focus on a single device for backups increase the potential need to back up the backups to avoid disaster; particularly if those devices cannot be easily tested to ensure that they are still functional.

4.4. Flexibility and Scalability

- **HSM platform independence:** All Security World functionality is independent of the form factor of the particular nShield HSM used – USB-attached device, embedded PCI/PCIe card or Ethernet appliance. Platform independence provides the ability for customers to mix and match form factors within the same Security World deployment. For example, a large data encryption deployment would use multiple servers and nShield Connect appliances in its primary data center, but a more limited disaster recovery deployment could use fewer servers and an nShield Solo board in each server. Both environments utilize the same operational procedures and have secure access to the same encryption keys because the nShield hardware is programmed in the same Security World. This same independence enables customers to migrate between HSM form factors over time without forcing re-keying headaches – for instance, customers migrating from embedded nShield Solo cards to nShield Connect appliances as they expand their use of virtualized IT environments.

- **Rapid scalability:** By abstracting Application Keys as tokens, it becomes easy to add modules to an existing group of nShield HSMs, increasing overall processing capacity of the group. By configuring additional modules with the same Security World Infrastructure Keys each new module is able to work as part of the original group without the need to import or clone potentially large volumes of Application Keys into the HSM and without the need to modify the application infrastructure. Common scenarios where multiple HSMs are deployed include:
  - Two or more nShield Connect HSMs configured to support load balancing, appearing as a single high capacity network resource to one or more application instances (for example, an online certificate validation server).
  - Multiple nShield Solo PCIe card HSMs embedded within a single physical server or appliance to provide higher processing capacity than a single HSM.
  - Multiple servers, themselves acting a distributed group to support a single business application each with an embedded nShield Solo HSM.
- **Unlimited capacity for key storage:** Because Application Keys are abstracted from the HSM the number of keys that can be accommodated within a Security World is virtually unlimited, constrained only by the size of the external file system. This is in contrast to legacy HSMs designs where keys are stored within the module and capacity is therefore limited by the physical memory of the HSM.

Platform independence provides the ability for customers to mix and match HSM form factors within the same Security World deployment.

5. Independent Certification of the Security World Architecture

Not only has the Security World architecture been successfully proven in thousands of real world deployments, the nShield family of HSMs has one of the most extensive track records of independent security certifications on the market. By far the most widely recognized security certification for HSMs is FIPS 140-2. This Federal Information Processing Standard defines the security requirements for cryptographic modules used to protect sensitive data within government and enterprise information systems. This standard and associated Cryptographic Module Validation Program (CMVP)[1] is maintained in the US and Canada and has been in place since the mid-1990s.

In addition to FIPS validation the nShield HSM and Security World architecture are fully compliant with the National Institute of Standards and Technology (NIST) Special Publication (SP) 800-131*A Recommendation for the Transition of Algorithms and Key Sizes* issued in January 2011 and Rev 1 draft version dated July 2015. This establishes a transition plan for algorithms, key sizes and other mechanisms commonly used by cryptographic products. With the issuance of this recommendation, nShield software now presents users with an option to create a Security World that uses the latest key length requirements for digital signatures. With this transition in mind and to overcome the potential impact on performance of adopting longer keys within an HSM, the nShield hardware is performance optimized to support operations using the recommended key lengths of 2048 bits.

[1] Further information on this certification and the CMVP can be found at the National Institute of Standards and Technology (NIST) web site: http://csrc.nist.gov/groups/STM/cmvp/documents/140-1/1401vend.htm

6. Conclusion

With an ever-growing volume of information used to conduct operations, the management of keys used to secure these vital data resources has become increasingly complex. As a result, key management is now the most challenging task in the area of cryptography and this is particularly true when implementing cryptography to protect business applications in an increasingly cost-conscious environment. To address this, the role of HSMs has taken center stage – not only to simplify these key management tasks but also to do so in way that delivers incremental security to this predominantly software-oriented landscape.

As the use of HSMs becomes more mainstream, this places new burdens on their traditional concept and is forcing the industry as a whole to adopt more efficient methods for deploying and managing HSMs in the context of a broader IT environment. The Thales Security World architecture described in this document has been designed with this principle in mind from the outset. This widely proven architecture maximizes the flexibility and scalability of the nShield HSM family by using a robust and flexible key management framework to provide specific advantages over alternative, more traditional approaches to architecting HSMs and over software-only deployments. Together these capabilities dramatically simplify day to day management tasks such as backup and provisioning and also extraordinary events such as disaster recovery, ultimately reducing operational costs and increasing responsiveness – all without compromising security, which remains paramount.

Governments and enterprises looking to capitalize on this flexible and scalable HSM key management framework will find that the Thales Security World architecture offers an unmatched and forward thinking approach to maximizing the capabilities of the family of nShield HSMs.

Appendix A

Secure Tokenization of Application Keys

Many of the advantages of the Security World architecture stem from its ability to securely tokenize Application Keys so that they can safely exist outside the physical boundary of the nShield HSMs while not in active use. The creation of Application Key Tokens enables key material to be stored, provisioned, shared and backed-up in a convenient way without weakening security. This capability to free Application Keys from the constraints of the individual HSM delivers many of same benefits that accrue from the now routine use of server virtualization: higher scalability, greater flexibility, simplified resilience and reduced operating costs. This appendix provides a high-level description of the security characteristics that underpin this tokenization process.

Application Keys Tokens are derived from two main constituent parts: the Application Key itself, and its Access Control List (ACL). Other constituent parts can include application specific metadata but they are beyond the scope of this paper.

The Access Control List

The ACL associated with an Application Key defines the key policy in a form an nShield HSM can strongly enforce. In Security World, an ACL is expressed in a powerful policy language which defines not only what a key can do (for example, separating encryption and decryption permissions), but also how that operation is to be authorized and what constraints might apply (for example, any time limits or constraints).

Whether provisioning, backing up, or performing any other key management task, it is absolutely vital that a key's ACL (the key policy) is tightly bound to the key itself to ensure that the ACL is strictly obeyed in all circumstances and in any device where the key is loaded. If it were possible to change a key's ACL in transit this type of attack would be considered as serious as an attack on the key itself – imagine the impact of modifying an ACL to add permission to export key material or changing permission to allow an encrypt-only key to start decrypting data.

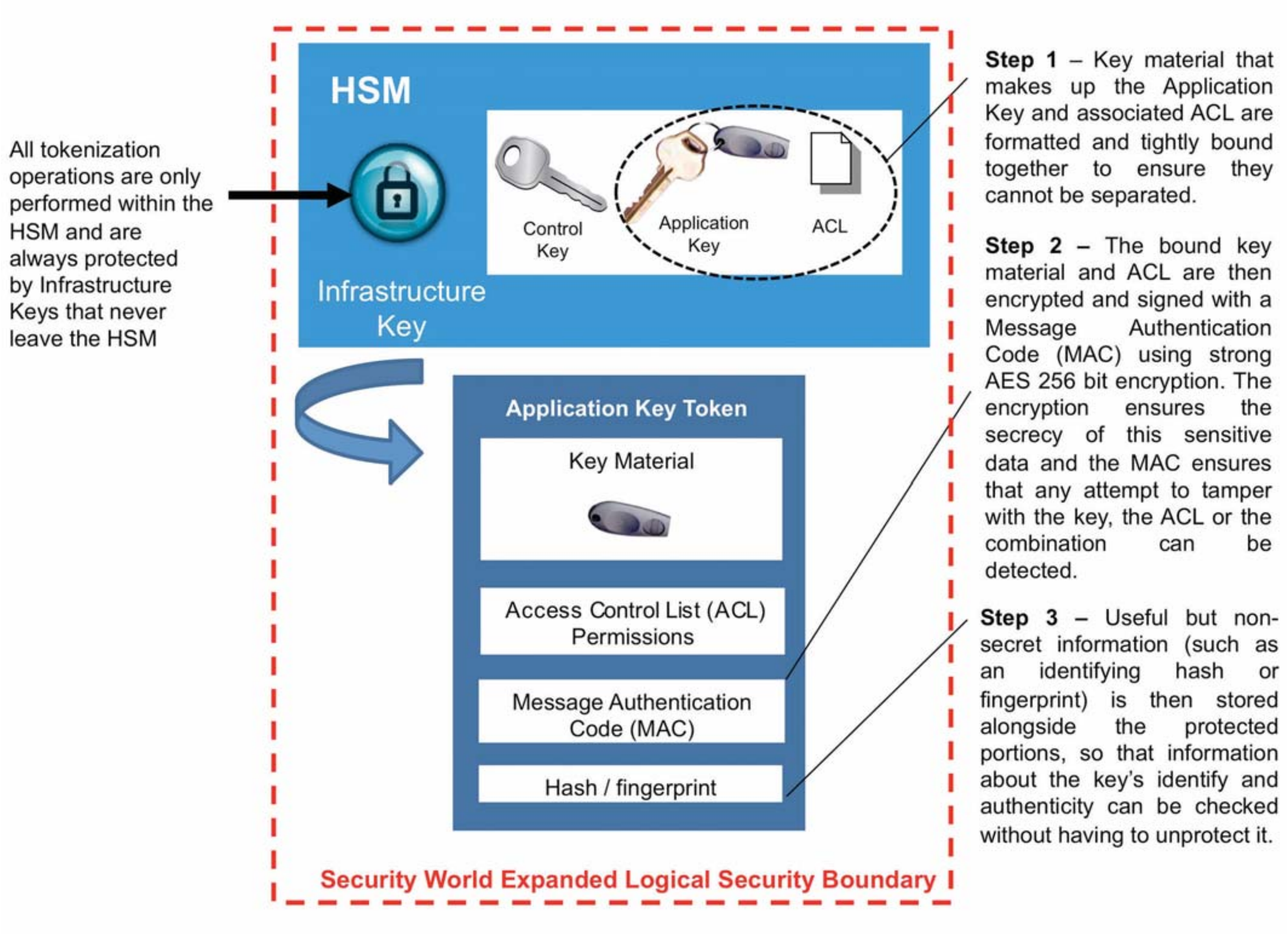
Many general security APIs fail to pay sufficient attention to the policy of a key but Security World Application Key Tokens combine flexibility and scalability with complete security of the key material by guaranteeing the strong enforcement of the security policy.

The Tokenization Process

The process of secure tokenization provides a powerful mechanism that protects Application Keys from attack while enabling them to be used in multiple approved locations. To enable this, a series of layers of cryptographic protection are applied to create and secure Application Key Tokens.

At the heart of this protection are the Infrastructure Keys which provide a consistent security underpinning to the whole HSM Security World model. With this approach, Application Keys never leave the safety of the nShield HSM without first being tokenized, and they cannot be used again until they are detokenized within the safe environment of a properly authorized nShield HSM.

Figure 6 illustrates the steps taken to securely tokenize an Application Key into an Application Key Token.


HSM
Control Key
Application Key
ACL
Infrastructure Key
Application Key Token
Key Material
Access Control List (ACL) Permissions
Message Authentication Code (MAC)
Hash / fingerprint
Security World Expanded Logical Security Boundary

All tokenization operations are only performed within the HSM and are always protected by Infrastructure Keys that never leave the HSM

Step 1 – Key material that makes up the Application Key and associated ACL are formatted and tightly bound together to ensure they cannot be separated.

Step 2 – The bound key material and ACL are then encrypted and signed with a Message Authentication Code (MAC) using strong AES 256 bit encryption. The encryption ensures the secrecy of this sensitive data and the MAC ensures that any attempt to tamper with the key, the ACL or the combination can be detected.

Step 3 – Useful but non-secret information (such as an identifying hash or fingerprint) is then stored alongside the protected portions, so that information about the key's identify and authenticity can be checked without having to unprotect it.

Figure 6 – Anatomy of an Application Key Token and the steps taken to tokenize a key.

Encrypted and protected Application Key Tokens can now be abstracted from the HSM and stored within existing data storage systems in the Application Layer with complete confidence, dramatically simplifying routine key management tasks.

To ensure complete security, the tokenization process implemented by Thales Security World meets or exceeds the guidance established by FIPS 140-2, ISO 15782, X9, TR31 and the OASIS Key Management Interoperability Protocol (KMIP), all of which include governing procedures for the safe handling and wrapping of cryptographic keys. In particular, by binding policy and keys with strong authenticity and integrity controls, Security World provides strong key typing and policy enforcement to ensure keys can only be used with appropriate algorithms and modes of operations. As a result, the Security World architecture actively blocks several classes of attack that have been applied to manual key transport schemes.

The process for presenting and redeeming Application Key Tokens to reconstitute the Application Key for secure key use can only be performed inside an nShield HSM that is pre-enrolled as a member of the same Security World as the HSM that originally created the token. Before any action is taken with a key token the authenticity and integrity of the token is checked by verifying the message authentication code (MAC). Once the application key is reconstituted within the HSM there are a number of potential methods to control and authorize its use. This is defined when the Application Key is originally created and forms part of the ACL embedded within the Application Key Token. The various methods and choices under which application keys can be controlled and used were further described in Section 3.3.

Appendix B

Remote Administration helps customers:

- **Cut travel costs**
- **Reduce downtime**
- **Eliminate the risk of carrying cards to remote locations**

nShield HSM Remote Administration

Companies are increasingly turning to remote, lights-out data centers to cut costs, but often they're still saddled with sending administrators to these distant facilities to manage their HSMs. nShield's Remote Administration feature enables administrators to manage their HSMs remotely once physically installed, removing the need to be physically present.

With Remote Administration, customers can perform the same functions remotely as they can in person. Useful functions include adding new HSMs and applications to Security World, creating new Security Worlds, and maintenance activities such as updating firmware and monitoring nShield HSM status.

How Remote Administration Works

Remote Administration is simple to use. The following components, included in Remote Administration Kits, enable the process:

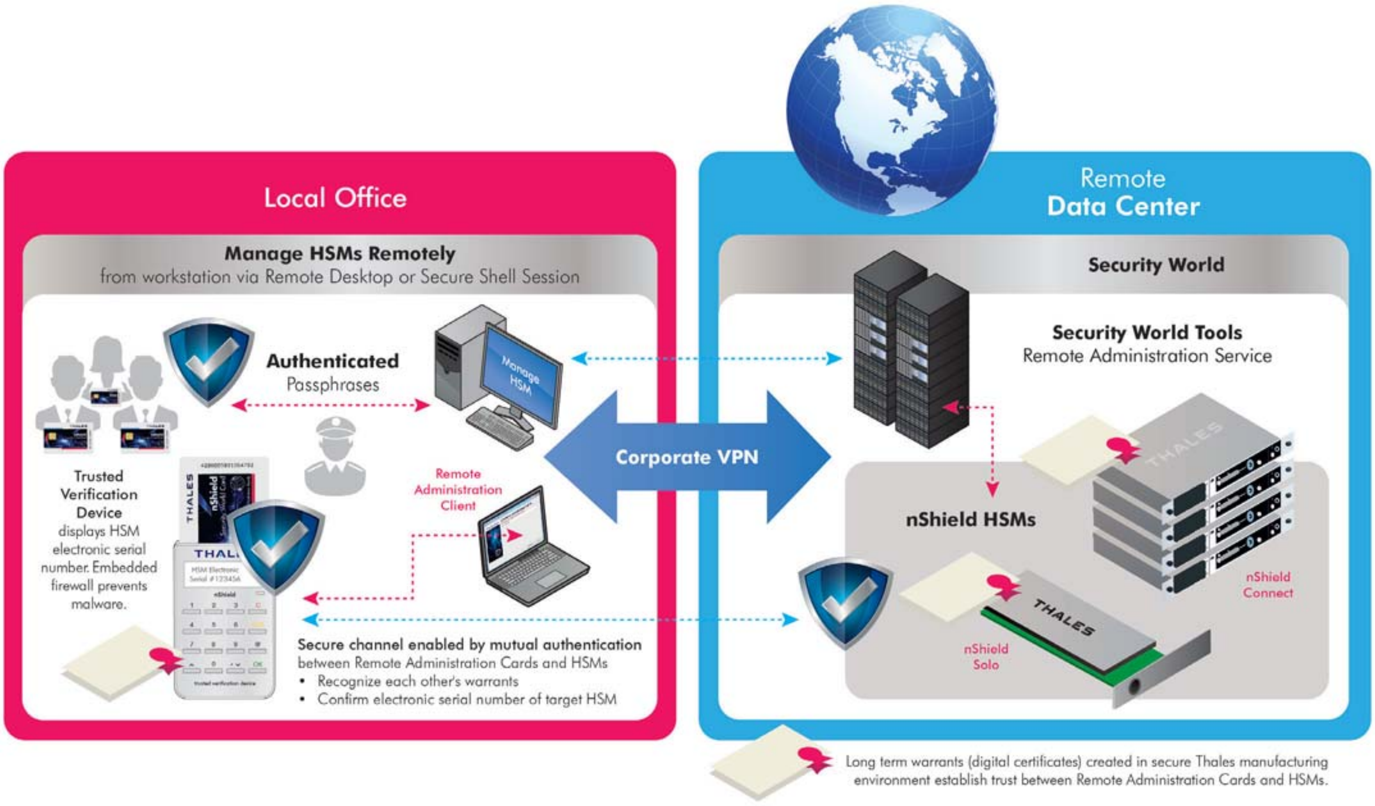
- **Remote Administration Cards**: Like ACS and OCS cards, these smart cards contain logical key tokens but are additionally equipped with applets which work with the Trusted Verification Device, defined below, to connect with the remote HSM.
- **Trusted Verification Devices (TVDs)**: TVDs authenticate the Remote Administration Cards, and, in conjunction with the Remote Administration Client software defined below, create a secure connection between the smart cards and the remote HSM.
- **Remote Administration Client (RAC) Software**: The RAC software mutually authenticates the Remote Administration Cards and the remote HSM based on the HSM electronic serial number.

Once the Remote Administration Cards are connected to the HSM, administrators can securely gain access to the Remote File Server within their Security World. Communicating over a VPN, the administrator controls the HSM from his or her workstation via a Remote Desktop (RDT) or Secure Shell (SSH) session.

Remote Administration safeguards the process:

- **Authenticates the Remote Administration Cards, the card holders, and the remote HSMs**
- **HSMs accessed via remote desktop (RDP) or Secure Shell Session (SSH) sessions over VPN connection**

The figure below illustrates how Remote Administration enables users in a local office to manage HSMs in remote locations while safeguarding their security.



Trusted Verification Device displays HSM electronic serial number. Embedded firewall prevents malware.

Secure channel enabled by mutual authentication between Remote Administration Cards and HSMs

- Recognize each other's warrants
- Confirm electronic serial number of target HSM

Long term warrants (digital certificates) created in secure Thales manufacturing environment establish trust between Remote Administration Cards and HSMs.

Figure 7 – Remote Administration overview

www.thales-esecurity.com

About Thales e-Security

Thales e-Security is a leading global provider of trusted cryptographic solutions with a 40-year track record of protecting the world's most sensitive applications and information. Thales solutions enhance privacy, trusted identities, and secure payments with certified, high performance encryption and digital signature technology for customers in a wide range of markets including financial services, high technology, manufacturing, and government. Thales e-Security has a worldwide support capability, with regional headquarters in the United States, United Kingdom, and Hong Kong. www.thales-esecurity.com

About Thales Group

Thales is a global technology leader for the Aerospace & Transportation and the Defence & Security markets. In 2013, the company generated revenues of €14.2 billion ($18.3 billion) with 65,000 employees in 56 countries. With its 25,000 engineers and researchers, Thales has a unique capability to design, develop and deploy equipment, systems and services that meet the most complex security requirements. Thales has an exceptional international footprint, with operations around the world working with customers and local partners.
For more information, visit www.thalesgroup.com

Follow us on:

Americas – Thales e-Security Inc. 900 South Pine Island Road, Suite 710, Plantation, FL 33324 USA • Tel: +1 888 744 4976 or +1 954 888 6200 • Fax: +1 954 888 6211 • E-mail: sales@thalesesec.com
Asia Pacific – Unit 4101 41/F 248, Queen's Road East, Wanchai, Hong Kong • Tel: +852 2815 8633 • Fax: +852 2815 8141 • E-mail: asia.sales@thales-esecurity.com
Europe, Middle East, Africa – Meadow View House, Long Crendon, Aylesbury, Buckinghamshire HP18 9EQ • Tel: +44 (0)1844 201800 • Fax: +44 (0)1844 208550 • E-mail: emea.sales@thales-esecurity.com