Operational Management Manual for the CEX6

Version 1.1

Authors:

Eric B. Smith, Todd Arnold, and Richard Kisley

IBM Corporation

New Orchard Road

Armonk, NY 10504

Produced in the United States of America

October 2021

1 About this information

This operational management manual provides instructions and information concerning the identification, use, repair, updates, and configuration of hardware and firmware components of the 'IBM 4768 Cryptographic Coprocessor Security Module'. The 'IBM 4768 Cryptographic Coprocessor Security Module' is marketed as the "Crypto Express6S", abbreviated as CEX6S, when used in an IBM Z server. Note that in some marketing materials the IBM HSM is referred to as the "Crypto Express6S with CCA", abbreviated as CEX6C. The rest of the document refers to the IBM 4768 PCIe Cryptographic Coprocessor as the "IBM HSM". This manual is in addition to user and application operation and configuration manuals.

The PCI (Payment Card Industry) Security Standards Council (https://www.pcisecuritystandards.org) develops standards to ensure security in the payment card industry. The Council defines its standards as “a set of security standards designed to ensure that ALL companies that accept, process, store or transmit credit card information maintain a secure environment.” One of their standards is the PCI PIN Transaction Security (PTS) Hardware Security Module (HSM) specification.

Compliance with the PCI PTS HSM standard has a great deal of value for customers, particularly those who are in the banking and finance industry. The two fundamental reasons why this certification is important to customers are (1) compliance is increasingly becoming mandatory, and (2) the requirements in PCI PTS HSM make your system more secure.

1. Industry requirements for PCI PTS HSM compliance

   The PCI Security Standards Council itself has no power to require compliance with its standards. Compliance with PCI standards is enforced by the payment card brands, namely Visa, MasterCard, American Express, JCB International, and Discover. If you are a merchant, processor, acquirer, issuer, service provider, or other participant in the payment card systems that wants to process their payment cards, or an entity that stores, processes, or transmits cardholder data (which includes Primary Account Number, cardholder name, expiration date, and service code), sensitive authentication data (which includes magnetic strip data or equivalent on a chip, CAV2/CVC2/CVV2/CID, and PINs/PIN blocks), or both, the card brands have the ability to impose requirements on you. One set of requirements that they have been increasingly enforcing are the PCI standards.

   The card brands work with the Council to develop the PCI standards. The standards that the card brands developed first were the ones that they considered to be most important. Particularly important to them is the PCI Data Security Standard (PCI DSS). Other standards have been added over time, and PCI PTS HSM is one of the latest to be developed. In addition, many of the standards have been changed over time to make them stronger and to mandate previously optional standards.

   In general, the trend is for the card brands to enforce more of the PCI standards and to enforce them more rigorously. The trend in the standards themselves is to impose more and stricter requirements in each successive version. The net result is that companies subject to these requirements can expect that they will eventually have to comply with all of them.

2. The requirements in PCI PTS HSM make your system more secure.

   The main reason for developing PCI PTS HSM was to improve security in payment card systems. It imposes requirements in key management, HSM API functions, device physical security, controls

during manufacturing and delivery, device administration, and a number of other areas. It prohibits many things that were in common use for many years but are no longer considered secure.

The result of these requirements is that applications and procedures often must be updated because they used some of the things that are now prohibited. While this is inconvenient and imposes some costs, it does truly increase the resistance of the systems to attacks of various kinds. Updating a system to use PCI PTS HSM compliant HSMs will reduce the risk of loss for both the institution and its customers.

2 Who should read this information

This information is intended for customers who deploy an IBM HSM that is involved in support of a variety of payment-processing and cardholder-authentication applications and processes for the financial payments industry, including these relevant processes:

- PIN processing
- 3-D Secure
- Card verification
- Card production and personalization
- Electronic funds transfer at point of sale (EFTPOS)
- ATM interchange
- Cash-card reloading
- Data integrity
- Chip-card transaction processing
- Key generation
- Key injection

3 IBM CCA 6.* features and enhancements

IBM provides a Common Cryptographic Architecture (CCA) for its hardware security modules (HSMs) that includes an application programming interface (API) which is intended for systems analysts, applications analysts, and application programmers to evaluate or create programs that employ the CCA API. In response to the PCI PTS HSM standards and adoption trends in the industry, version 6.0.*z and newer of CCA is designed to meet the June 2016 "Payment Card Industry PIN Transaction Security Hardware Security Module Version 3.0" standard when running in a PCI-HSM compliant state Confirm your CCA firmware version by reviewing IBM members of the PCI PTS HSM approved devices list on the PCI SSC website.

CCA features in support PCI-HSM compliant operation:

- The ability to simultaneously support PCI PTS HSM compliant applications and non-compliant applications
- Help in determining what parts of your current system need to be changed to be compliant
- Mandatory dual control for sensitive operations
- Separate logical key spaces to support both compliant and non-compliant workloads
- Secure auditing of sensitive operations
- Key usage restrictions for keys used in PCI PTS HSM compliant applications

- Cryptographically protected information about firmware versions in the IBM HSM, which can be viewed from a remote administration workstation

These features and this environment have been provided to support your needs when dealing with applications subject to PCI standards. For more information on PCI Council and PCI standards, visit their website at https://www.pcisecuritystandards.org.

4 IBM HSM life cycle

This manual provides instructions for the operational management of the IBM HSM. The following sections describe the recording of the entire life cycle of the device's security-related components and the manner in which these components are integrated into a single device:

- Data on production and personalization
- Physical and chronological whereabouts
- Repair and maintenance
- Removal from operations
- Loss or theft

The function of the IBM HSM is dependent on the firmware (embedded software) that is loaded into it. The following sections address the IBM HSM when it is loaded with the IBM Common Cryptographic Architecture (CCA) firmware.

4.1 Data on production and personalization

The IBM HSM is a Secure Cryptographic Device (SCD) that can contain a variety of Critical Security Parameters (CSPs). Some are associated with the IBM HSM itself and are installed by IBM at the time of manufacture or are generated by the IBM HSM itself. Others are created and owned by the end user of the IBM HSM, such as cryptographic keys used when the IBM HSM is in operation supporting business applications. No CSPs are ever exported from the IBM HSM in plaintext under any circumstances.

4.1.1 Manufacturing the IBM HSM

The manufacturing process for the IBM HSM is performed in secured facilities with trusted personnel under tightly controlled conditions. When an IBM HSM is fully assembled and tested but prior to shipment, two operations are performed that provide customers and other external entities a way to ensure that the IBM HSM came from IBM and it is secure and untampered:

1. The IBM HSM generates an Elliptic Curve *device key pair* unique to that device which can later be used to verify that the IBM HSM is from IBM and untampered. The device keys and their use are described further in Section 4.1.5 on page 8.
2. After the device keys are generated, the protection features are fully enabled.

4.1.2 IBM HSM tamper detection and response

The IBM HSM has self-contained tamper detection and response technology to protect customer data. It is designed to meet the most stringent requirements of Federal Information Processing Standard Publication 140-2 (FIPS PUB 140-2), Level 4. Check the NIST website for current/pending IBM HSM certificate status.

Once the protection features are enabled, if the tamper-responding secure module of the IBM HSM card detects any attempt to tamper or attack it (for example, the tamper-sensing mesh enclosure is

disturbed, there is an extreme in temperature or voltage, or external and battery power is lost), the IBM HSM quickly zeroizes itself (within 100 nS).

After an IBM HSM is zeroized, it is permanently disabled and the only functions that it can perform in that state are diagnostic functions. The diagnostics do not provide a way to recover any sensitive information. They only enable engineers to determine the cause of the tamper and to perform other forensic analysis.

4.1.3 Device-level Critical Security Parameters (CSPs)

The following CSPs are associated with the IBM HSM itself and installed by IBM at the time of manufacture or generated by the IBM HSM itself:

- IBM Class Root key pair – IBM creates an IBM Class Root key pair for each class of IBM HSM that it develops. The public key for each IBM Class Root key pair is in a certificate that is signed by the IBM Root private key. The private IBM Class Root key is securely held by IBM outside of the IBM HSM.
- The Device key pair – Each IBM HSM generates its own Device key pair as part of the secure manufacturing process. The Device private key never leaves the IBM HSM. The Device public key is inserted into a certificate that is signed by the IBM Class Root private key. That signed certificate is then stored in the IBM HSM and can be read out by external program.
- Operating System key pair – The IBM HSM generates a key pair called the Operating System key pair (O/S key pair), and the public key of this pair is stored in the IBM HSM in a certificate that is signed by the Device key pair private key.
- Seg3 Epoch key pair – A key pair generated by the CCA firmware, used to sign messages from the IBM HSM to the host in order to prove that the IBM HSM is valid and untampered. The Seg3 Epoch public key is in a certificate signed with a chain of keys that goes back to the published IBM Root public key.
- File system encryption key - A 256-bit AES key that is randomly generated when you establish ownership of Segment 2. This key, stored in non-volatile RAM and owned by the Security Service Processor (SSP), is invisible to the Module CPU (MCPU). The SSP is the microprocessor in the IBM HSM that performs and secures service operations and the MCPU is the microprocessor in the IBM HSM that processes the CCA API requests that come from the host computer system. A variant of the file system encryption key is used to encrypt one of the two device-resident file systems, and a different variant is used to encrypt the other one.
- Function control vector (FCV) – Allows IBM to limit the IBM HSM cryptographic functionality in order to enforce export regulations or other restrictions.
- Enable/disable flag – A flag that determines whether the IBM HSM is in “disabled” state such that is can be safely removed from the server without risk of sensitive keys being used.
- TKE CMSSN – Crypto module signature sequence number (CMSSN) used in a secure communications protocol between the IBM HSM and the Trusted Key Entry (TKE) administrative workstation.
- Deterministic random bit generator (DRBG) state – The IBM HSM maintains three independent random number streams. All three streams use the hardware-based NIST SP 800-90A DRBG, but each uses an independent seed and maintains the DRBG state so that each stream can restore that state to the hardware when a new set of random numbers must be generated. This data is stored in plaintext in DRAM and is zeroized on tamper or card reset.

4.1.4 Domain-level Critical Security Parameters (CSPs)

The z System server can be separated into multiple logically separate servers with hardware-enforced separation for *logical partitioning*. Therefore, the IBM HSM supports virtualization such that it appears as a set of multiple logically independent IBM HSMs, each with its own CSPs at the scope of the logical IBM HSM and its own execution context. The separate, virtualized IBM HSMs are called *domains*.

Each domain can have the following possible CSPs created and owned by the end user:

- TDES 168-bit DES master key – Protects DES and TDES keys that are stored outside of the IBM HSM.
- TDES 168-bit PKA master key – Protects RSA private keys that are stored outside of the IBM HSM when those keys use the legacy RSA key structure section IDs of X'05', X'06', or X'08'.
- AES 256-bit AES master key – Protects AES keys and HMAC keys that are stored outside of the IBM HSM.
- AES 256-bit APKA master key – Protects RSA and ECC private keys that are stored outside of the IBM HSM when those keys use RSA key structure section IDs of X'30' or X'31', or ECC key structure section ID X'20'.
- Access control roles – Roles used by the integrated access control system that define the access rights for classes of IBM HSM users.
- Access control profiles – Profiles used by the integrated access control system that define individual users of the IBM HSM and maps their access rights to one of the roles.
- KPIT key part table – Contains key parts or combined key parts for operational keys that are being manually entered.
- Retained private keys – CCA-generated RSA private keys that are generated and stored inside the IBM HSM but cannot be exported in any way.
- Public Key Infrastructure (PKI) root certificate keys. These public-key certificates are entered under dual control to the IBM HSM to form the base of a chain of trust for validating operational certificates that are used for public key operations on the IBM HSM. Only those certificates that are trusted are loaded to the IBM HSM, and so each chain of trust allows authenticated operation for a set of end-user operational certificates.
- Secure audit log. This immutable log of administrative state changes is managed on a per-domain basis, retrieved and managed using a TKE workstation.

Recommended user policies:

1. Users should have a policy for key rotation such that all symmetric keys or private keys are replaced with new keys before there is significant possibility of their compromise.
2. Users should manage certificates in the IBM HSM internal PKI for expiration and revocation and ensure a properly set clock value in the IBM HSM for internal expiration calculation.

Users will manage and configure the domain-level CSPs using a secured remote administration tool known as the Trusted Key Entry workstation, or TKE. Proper operation of the TKE to configure and manage a PCI PTS HSM compliant domain on the IBM HSM is detailed in the TKE publication, *Trusted Key Entry Workstation User's Guide*, available online via Resource Link. The publication for TKE version 9.2 is at this link:

https://www-01.ibm.com/servers/resourcelink/svc00100.nsf/pages/zOSV2R4sc147511/$file/csfb600_tke_9_2.pdf.

4.1.5 Validating the origin and integrity of the IBM HSM

Prior to the use of a new IBM HSM, it is important to verify that the device came from IBM and that its integrity has not been compromised. The TKE administrative workstation automatically does this for each IBM HSM that it administers. The IBM HSM is entirely self-protecting and there is no need for visual inspection to determine if tampering has occurred. As noted above, a tampered IBM HSM is not operational.

Each IBM HSM has a unique **serial number**, assigned at the factory. This serial number is printed on a label attached to the IBM HSM card, and it is also programmed into the internal memory of the IBM HSM. The serial number in card memory is immutable; it cannot be changed. When read from the IBM HSM, the programmed serial number can be digitally signed. The digital signature uses elliptic curve cryptography (ECC) keys that are part of a chain that can be verified back to a known IBM Class Root signing key. The TKE administrative workstation uses this digital chain to automatically verify the integrity and authenticity of each IBM HSM that it administers.

In addition to the serial number, the IBM HSM **Vital Product Data (VPD)** is programmed into the card. This data includes:

1. firmware versions currently loaded in all memory segments,
2. part number (for the non-encapsulated portion of the IBM HSM),
3. secure part number (for the encapsulated portion of the IBM HSM), and
4. hardware version information.

The VPD can be seen using the ISPF panels provided with the z/OS ISPF cryptographic support software, the Support Element (SE) server administration tool, and the “panel” utility provided with CCA support on z Systems Linux. TKE can also display the VPD, under “Other CM info.”

The following picture shows the IBM HSM with the labels showing the hardware part number and serial number:

Figure 1: IBM HSM with labels

The following picture is a close view of the label showing the serial number:

Figure 2: IBM HSM Serial Number Label

The following picture is a close view of the label showing the hardware part number:

Figure 3: IBM HSM Hardware Part Number Label

4.1.6 Logging of security-relevant events

The IBM HSM maintains an internal, non-volatile log of security-relevant events that have occurred. This log includes events in categories such as key management, changes to access control settings, enabling or disabling of IBM HSM functions, firmware updates, and other administrative operations. It is the customer's responsibility to examine the logs to note any suspicious activity and to perform appropriate investigations to determine the cause. In addition, the customer must monitor the logs in each IBM HSM to ensure that they are read out and cleared before they fill.

Each log record includes:

1. the date and time of the event,
2. the identity of the person who performed the operation,
3. the serial number of the IBM HSM, and
4. a sequence number.

The records are digitally signed when they are read from the IBM HSM to prove that they have not been modified and that they came from a valid IBM HSM. The signatures are computed using the Seg3 Epoch key. The corresponding public key certificate and the other certificates required to verify the signature are read automatically by the TKE workstation and do not have to be done by the end user. The signature is computed using SHA-512 hashing and the ECDSA digital signature algorithm.

Events are logged separately for events relevant to each separate IBM HSM domain. In addition, there is one log for events with scope that is global to the IBM HSM and not related to any specific domain.

The management of the logs differs slightly between domain-scoped logs and IBM HSM-scoped logs:

- **Domain-scoped logs.** Each domain-scoped log can hold up to 512 events. While in PCI PTS HSM compliant mode, each domain-scoped log must be read from the IBM HSM and cleared before it fills. In order to avoid losing any log records, the logs do not wrap. When the IBM HSM is not in PCI PTS HSM compliant mode, the log can be configured either to wrap or not to wrap.
- **HSM-scoped logs.** The HSM-scoped log can hold up 2,560 events. HSM-scoped logs operate in the same way as the domain-scoped logs for some commands, but a subset of HSM-scoped commands behave differently. The set of commands that behave differently are those that should be allowed to execute even if the log is full, such as the loading of new firmware. If one of these events occurs and the log is currently full, the event is stored in a special cache that can hold one event of each type. When the log has been cleared, the cached events are copied into the log. Since each command of this type has a cache that can hold just one log record, the event in the cache is overwritten if the same command occurs again before the log has been cleared. However, these commands are ones that should only occur infrequently.

Note that the TKE keeps a log of relevant operations that are performed by the TKE workstation itself. The logs maintained by the IBM HSM meet all the requirements of PCI PTS HSM, but the additional logs in TKE are also useful in tracking and understanding use of the system.

Proper operation of the TKE to manage the secure audit log for a PCI PTS HSM compliant domain on the IBM HSM is detailed in the TKE publications and videos.

4.2 Physical and chronological whereabouts

There are several ways to receive an IBM HSM for your z System server:

- Purchase of a new IBM HSM feature for a server that is in your data center.
- Purchase of a new z System server that is preconfigured with the IBM HSM.
- Replacement of an IBM HSM that has failed.

Upon receipt of a new IBM HSM, a record of its physical and chronological whereabouts, the **IBM HSM inventory**, needs to be securely maintained throughout its useful life. The record for each IBM HSM in the IBM HSM inventory should contain:

1. the name of the person who received the IBM HSM,
2. make and model number,
3. serial number,
4. card part number,
5. date and time received,

6. physical location of the IBM HSM,
7. a record of the state of the IBM HSM, such as whether it is loaded with one or more master keys or other sensitive data, and
8. the method of delivery from one point to the next should also be described.

Each IBM HSM should be kept in a securely protected location. An inventory of each IBM HSM should be conducted at least quarterly. Section 4.5 on page 13 describes what to do when a loss or theft is discovered.

4.3 Repair and maintenance

The only maintenance operation for the IBM HSM is the updating of its firmware. No repair is possible for the IBM HSM, and there is no physical maintenance that is required. Firmware updates are performed either by the customer or the IBM Customer Engineer (CE) who normally services the z System server that contains the IBM HSM.

Customers should ensure that the appropriate responsible parties in their organization approve of each action by the CE and agree with the result. The IBM HSM inventory must be updated following each action that changes the location or the life-cycle phase for an IBM HSM.

Instructions for obtaining updated firmware and installing it are described in the appropriate IBM Service Guide for the IBM Z in use. Service Guides for particular models are available on IBM Resource Link: https://www-01.ibm.com/servers/resourcelink/hom03010.nsf/pages/library?OpenDocument&login.

To find the Service Guide for your model,

1. click the IBM Z model, for example "z14"
2. from the list of 'Publications', select the appropriate Service Guide, for example "Service Guide", which is 'Order number' of "GC28-6966-01a". Note that there may be a modifying document, titled "Service Guide Delta Pages".
3. Inside the Service Guide, look for the section titled "Licensed Machine Code changes". For example, this is Chapter 5 in the z14 "Service Guide".

The CE is also responsible for installing or removing IBM HSM cards from the server.

4.4 Removal from operations (decommissioning)

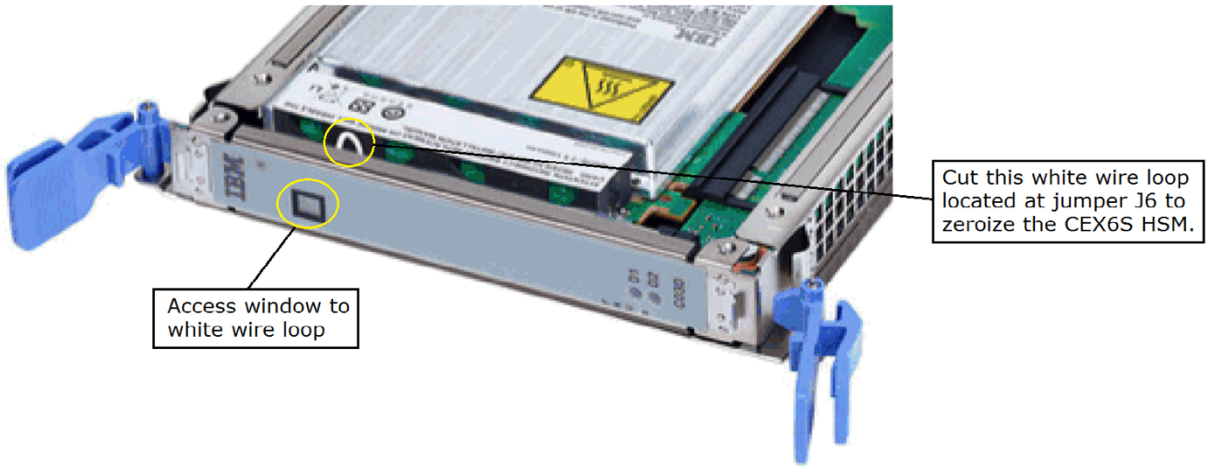
When an IBM HSM is removed from operations (decommissioned), either prior to its sale or discontinuance, all keys and other Critical Service Parameters (CSPs) installed by the customer must be deleted. There are several scenarios that must be considered, such as:

- The IBM HSM will no longer be used for the present application, but the customer will keep it for possible future use in other applications.
- The customer will never use the IBM HSM again and either (1) plans to sell it to another company, or (2) wants to render it permanently unusable.
- The IBM HSM has failed and is no longer working.

The reason for deletion of keys and other CSPs falls into two broad categories, one where the IBM HSM is operating normally, and one where the IBM HSM is not working properly.

1. **IBM HSM is operating normally.** Commands to the IBM HSM can be used to delete keys and other CSPs. CSPs can be deleted as follows:
   - CCA commands can be used to individually clear items such as master keys and access control information.
   - An entire feature domain can be zeroized with a single operation from the Trusted Key Entry Console (TKE) or the Support Element Console (SE), clearing all CCA keys and CSPs from the feature.
2. **IBM HSM is not working properly.** In this case commands generally cannot be issued, and other methods must be used to delete the keys and other CSPs. Choose between one of these methods to ensure all keys and other CSPs are deleted:
   - The IBM HSM is designed with an accessible battery disconnect wire for this purpose. It is a white wire loop located at jumper J6. Cutting this wire causes the hardware in the IBM HSM to delete all keys and CSPs. This renders the IBM HSM permanently inoperable. The battery disconnect wire is accessible through the access window, as shown in the figure below.

Figure 4: Wire location and access window on IBM HSM



   - Physically destroy the IBM HSM using a method such as mechanically shredding the device. The device must be physically destroyed, such that there is no possibility of the keys or other sensitive data being compromised.

4.5 Loss or theft

This section discusses what to do when a loss or theft of an IBM HSM, which is a Secure Cryptographic Device (SCD), is discovered. The loss or theft of an SCD can potentially cause a great deal of damage. It is very important to discover a loss or theft as soon as possible, and act quickly to minimize any damage. Keeping track of the physical and chronological whereabouts of each feature is discussed in Section 4.2 on page 11.

According to ANSI X9.24-1, *Retail Financial services Symmetric Key Management Part 1: Using Symmetric Techniques*, in the event of loss or theft of an SCD, all keys contained in that device should be considered compromised. The use of a key shall be discontinued if its compromise is known or suspected. The amount of time for which the compromised key remains active should be consistent with the risk to affected parties.

The notification of a key compromise shall be communicated to all parties that may be relying on the secrecy of the compromised key. In the case of a compromised key, key destruction and generating a new key shall be done in accordance with ANSI X9.24-1.

Once a key is compromised, or suspected to be compromised, all use of that key and related keys should be discontinued as soon as possible. Examples of related keys are keys that were encrypted by the compromised key or keys that were derived from the compromised key. If the compromised key is a derivation key, then use of any key derived from that key shall be discontinued. When generating a new derived key, no compromised key shall be used in the process to create the replacement key.