Designing a Secure System for Implementing Chip Cards in the Financial Services Industry

TERRY STANLEY
Mastercard

In order to design and implement effective security mechanisms for a chip card program in the financial services industry the designer must consider the current environment, understand carefully the constraints that he much work within, and have a clear view of the security objectives that he seeks to fulfill. Throughout this paper I will draw upon MasterCard International's implementation of EMV '96 Version 3.1.1, entitled MCPA^{TM} (MasterCard Chip Payment Application for Credit and Debit) in order to provide working examples.

The Credit Card Industry has experienced phenomenal success over the past thirty years, enjoying uninterrupted double-digit growth year after year. The logos of the major payment associations are among the most recognized in the world and consumers are able to pay for goods and services and obtain cash in virtually every part of the world. There are few, if any, international travelers who do not carry one or more cards from at least one of the major payment associations.

This unparalleled utility has also made credit cards a target for the criminal element. Traditionally, lost and stolen cards have been the largest category of fraud losses in the credit card business. Although a problem, the implementation of near 100% authorization of cardholder transaction at the point-of-purchase and sophisticated Issuer neural networks have contained losses well within a manageable level and year after year. As a result, this fraud category has declined as a percentage of overall credit card charge volume. However, increases in counterfeit and mail order/telephone order fraud have mitigated this success. For example, counterfeit fraud increased in actual dollar volume by 28% in 1998 as compared to 1997 and mail order

The original version of this chapter was revised: The copyright line was incorrect. This has been corrected. The Erratum to this chapter is available at DOI: 10.1007/978-0-387-35575-7_19

J. H. P. Eloff et al. (eds.), *Information Security Management & Small Systems Security*

telephone order fraud increased 35% during the same period. Keep in mind that this increase occurred while the other major fraud categories remained at a constant level or experienced a decrease.

The credit card business is based on risk management techniques that have been developed over the past thirty years. Fundamental to the security of credit cards is the magnetic stripe. In the early 90's there was a severe outbreak of counterfeit fraud that involved altering the contents of the magnetic stripe. From 1990 to 1994 counterfeit fraud climbed fourfold until MasterCard implemented a combination of preventive measures that included the following: Issuer and Acquirer awareness; implementation of CVC, which is a static algorithm that prevents that contents of the magnetic stripe from alteration; and an international crackdown on the organized criminal element

Nonetheless, counterfeit crime has begun to increase again as the organized criminal element is finding new vulnerabilities in the existing security system that is based on the magnetic stripe, the physical security features of the plastic card, and 100% authorization from the Issuer. The magnetic stripe is 35 year old technology. We may be nearing the limit of our capability to protect the magnetic stripe in a cost effective manner - changes to its security usually involve reissuing cards, upgrading terminals, and changing Acquirer networks and Issuer host systems. The physical features of the plastic card provide protection only in a face to face transaction environment. However, mail order/telephone order is the fastest growing charge volume category, and unattended points of transactions such as gasoline pumps are increasing. And while 100% authorization continues to provide an effective means of limiting exposure it also makes it difficult to move into off-line environments and increase the utility of the credit cards.

An additional goal of a new security system for credit cards is to leverage the investment to expand into new product lines that include electronic money, loyalty schemes and secure access mechanisms, among others. It is beyond the scope of this paper to discuss these possibilities but the security designer must incorporate enough flexibility to leverage the new infrastructure to service potential advances in these areas as well in order to share the cost of the investment.

An internal analysis at MasterCard revealed that chip is the best method to combat fraud and to provide a platform for expanding into new business opportunities. An example of the effectiveness of chip (combined with PIN) is the French implementation of chip on cards where there was a four fold drop in fraud as a percentage of charge volume between 1992 and 1996. Likewise business plans for expanding into new and emerging markets and providing new services to Cardholders are incorporating the expanded potential of chip cards into there designs.

With this background we can now begin to consider the important points for implementing a secure system for in the financial services industry. However, I will first point out some of the constraints that the designer must take into consideration.

- Card Costs – Today it costs less than one dollar to issue an embossed and personalized card in the US. With a chip, this cost may vary greatly, depending on the size of the chip, the volume of cards, maturity of the supporting infrastructure and whether a single or multiple application card is issued. Regardless of the scenario, there is a significant increase in costs that the designer must take into account.

- Terminal Costs – There is a large base of terminals deployed around the world that include traditional stand alone point of purchase terminals and integrated electronic cash registers. In addition, we need to consider the additional points of purchase that the payment industry would like to capture such as PC's, vending machines, toll collections, etc.

- Issuer and Acquirer Infrastructure Changes – Any significant deployment of chip cards will impact the existing Acquirer networks and Issuer hosts. These systems handle other functions and any changes are subject to resource availability and implementation prioritization. For example, the work necessary for Y2k and implementation of the Euro has taken priority.

- Commercial Infrastructure Changes – Most Issuers and Acquirers outsource at least some of their functions. This may include card production, personalization, terminal management systems, authorization and card management systems. The supplier of these systems and services must be given adequate specifications and a timetable for upgrading.

- Interoperability – Credit cards are the most universally accepted consumer payment mechanism. Any security system can not negatively impact the acceptance at the point of transaction either by creating problems in interoperability or making the transaction more difficult from the consumer or merchant's point of view.

- Knowledge Infrastructure – There is a large investment in terms of how to manage credit cards. For example, cardholders understand how to use their cards and pay their bills, merchants understand how to accept credit cards and settle the transactions and financial institutions know how to manage their risk and generate profits. The advent of chip cards will impact all of these players and functions and its implementation must be kept to the lowest common denominator.

Let us now move to our considerations for designing a secure system. First let's consider some of the basic premises that we must follow. Fundamental is the fact that any security system can be broken and the prudent designer must consider that it will be broken.

- The credit card business remains a risk management business. The fundamental design point is to make it unrewarding for the attacker. For example, the system should cause the attacker, who wants to make money, to expend more effort in breaking a card than he can gain by fraudulently using it. In the case of MCPA, this means that there are no global secrets in a card and exposure should be limited to the card that is compromised. It is a simple case of risk versus reward, and part of the risk should include criminal prosecution.

- There must be a migration strategy. It is impossible in a mature industry to implement a completely new security structure almost overnight for the reasons listed earlier. Accordingly, there needs to be a set of guidelines and implementation options. For example, MCPA permits an Issuer to implement static data authentication (SDA) or dynamic data authentication (DDA). SDA is an authentication mechanism performed by the terminal using a digital signature based on public key techniques to confirm the legitimacy of ICC-resident static data that is signed by the Issuer. This detects unauthorized alteration of data after personalization. Since the signed data is the same for each transaction it is subject to replay. DDA is also an authentication mechanism based on public key techniques to authenticate an ICC and confirm the legitimacy of critical ICC-resident data identified by ICC and terminal dynamic data. However, since the card signs dynamic data for each transaction it not subject to replay and therefore is the most effective security mechanism. However, DDA requires the use of a co-processor and the personalization of a unique RSA key pair per card that raises the costs per card. Accordingly, the SDA must be kept as a lower cost option.

- The system must be able to update cards deployed in the field. In order to justify the cost of the chip, many Issuers will find it necessary to extend the life of the card in the field. For example, if a magnetic stripe card has a two year expiration date, it may be necessary to reissue a chip card every four years instead of every two. As chip card security is a rapidly evolving field it is difficult, if not impossible, to foresee all the security threats over a period of five years. Accordingly, in order to protect the investment a mechanism must be in place to remotely update the operating system or the application when an unforeseen threat occurs.

- Likewise, the specifications must be published and updated on a regular basis. They should published, subject to peer review and periodically updated. Security practices should be based on industry standards.

- The security of the system should not depend on a single point of failure. For example, MCPA uses a protocol based on asymmetric cryptography for off-line transactions but employs symmetric cryptography when the card goes on-line to the Issuer's host where both the card and the Issuer can be authenticated. The card can execute off-line according to risk management parameters set by the Issuer and the payment association in the card and the terminal but must go on-line when the criteria is exceeded. This also permits the Issuer's host to monitor transaction counters that will detect if a card has been successfully counterfeited.

- The secure design must be end-to-end and the designer must understand all points of possible compromise. The attacker, on the other hand, does not need to understand the entire system. He only needs to learn enough to create and exploit a vulnerability.

The integrated circuit (IC) is the first component we will consider. The first stage is the design and development of the IC and the operating system that is masked in the Read-Only-Memory or ROM. There needs to be assurance that security critical parts of the design, such as the location of sensors that detect intrusion or secret keys in the EEPROM are not easily disclosed. Next we need assurance that the manufacturing process is secure and that the chip inventory is accounted for, that the design has not been altered and chips are safely shipped to the card manufacturer. Finally, we need evidence that the chip has an acceptable degree of tamper resistance to prevent disclosing security sensitive information. The attacker may attempt to modify the chip, manipulate it during a transaction or probe it internally. It has been known for some time that chips are vulnerable to invasive techniques used in traditional failure analysis such as Probing Stations or Focused Ion Beams, and recent attacks such as Differential Power Analysis (DPA) permit an attacker to obtain secret keys without opening the chip, simply by monitoring the external power consumption of the chip. The areas that a designer needs to protect are:

- Operating system code, data and keys
- Application code that an attacker may alter
- Application data that includes keys and PINS
- Test code that will permit the attacker to read out the memory contents
- Life cycle status that could cause a chip that has been blocked to revert to an active status

There is no single measure to provide acceptable defense but a combination of techniques such as masking chip operations by using either random logic or fixed branching, sensors and software that detect alternation and designing a layout that renders it difficult to determine or access security critical areas will raise the attack barrier. Nonetheless, the security of the system should not depend upon the tamper resistance of the chip.

The Operating System is a critical component of the overall security solution and must interface with the chip. Several important concepts are:

- Application data, static and dynamic, must be segregated. In a multi-application environment the designer needs assurance that his data will not be corrupted by another application.

- Writes to EEPROM must complete or else the OS reverts back to a previous state. This concept of atomic writes is critical to preventing attacks where a transaction is interrupted and the application is left in an unfinished state.

- The OS needs to guarantee the integrity of the code and data and to ensure that it executes code only form known and authorized memory locations

- The OS must react to intrusions to the chip and shut down. For example, if it detects that an environmental sensor has been disabled, it must not continue to execute

- The designer needs assurance that the OS conforms to the technical specifications, does not execute invalid commands and recovers from deviations in behavior such as terminating power.

Secure application design is difficult but the developer must be able to provide certain documentation. The steps that he should follow are:

- Develop a specification of the application's functionality

- Describe the protocol used by the application that details the state of the card and card reader after each message exchange.

- Develop a technical specification of the application that details:
  - How the protocol operates
  - Segmentation each module for both the card and the card reader
  - Description of the functions of each module
  - The main loop between the card and the card reader
  - The relationship between the application and the OS

- Develop a threat analysis of the application that includes:
  - Identifying the security-critical parts of the protocol
  - The actions necessary to execute an attack
  - An explanation as to why the attacks are or are not possible

- Provide a functional test report.

- Describe the security of the development environment that should include:
  - Configuration list
  - Version control system
  - Programming languages and compilers
  - Development process and tools used

- Explain the key management

- List additional security measures that might include host security and card lifecycle management

- Provide the risk management strategy that explains how the application will prevent, detect and recover from an attack.

Key Management is a crucial part. If an attacker is able to compromise the key management system of the Issuer, then he does not need to attack other parts of the system. He has what he needs – the key to create counterfeit cards. The security system must consider all stages of key management that include: generation; loading; usage; storage; withdrawal, archival; and destruction. There are several important principals that should be followed:

- Procedures should be documented, in accordance with industry standards, and subject to audit.
- Split knowledge and control must be used. No single individual should have access to a secret key.
- Keys should never be stored in a human readable form.
- Keys should only be stored in secure crypto devices that can pass tests such as FIPS 140-1. In addition, there should be adequate physical and procedural security around the site.
- There should also be a protected audit trail for any access or use of the key.