

Chapter 25

Gemplus: Smart Cards and Wireless Cards

Christophe Mourtel¹

Introduction

Smart cards have the tremendous advantage, over their magnetic stripe ancestors, of being able to execute cryptographic algorithms locally in their internal circuitry. This means that the user's secrets (be these PIN codes or keys) never have to leave the boundaries of the tamper-resistant silicon chip, thus bringing tremendous security to the overall system in which the cards participate.

Smart cards contain special-purpose microcontrollers with built-in self-programmable memory and tamper-resistant features intended to make the cost of a malevolent attack far greater than the benefits. This chapter is both a survey of the existing components and their applications and a description of some of their possible evolutions.

What Is a Smart Card?

In the time scale of the silicon industry, the idea of inserting a chip into a plastic card is rather old: The first patents are now 20 years old! But practical applications emerged only eight years ago because of limitations

¹ Christophe Mourtel is a hardware security expert in the security technology department at Gemplus, one of the world's smart card leaders.

in storage and processing capacities of past circuit technology. New silicon geometries and processing refinements lead the industry to new generations of cards and more ambitious applications such as wireless communications (GSM), pay TV, loyalty, and physical access control.

Over the last four years, there has been an increasing demand for smart cards from national administrations and large companies such as telephone service providers, banks, and insurance corporations. More recently, another market opened up with the increasing popularity of home networking and the Internet.

The physical support of a conventional smart card is a plastic rectangle on which can be printed information about the application or the issuer (even advertising) as well as readable information about the card holder (for instance, a validity date or a photograph) see figure 25.1. This support can also carry a magnetic stripe or a bar code label. An array of eight contacts is located on the micromodule in accordance with the ISO 7816 standard, but only six of these contacts are actually connected to the chip, which is (usually) not visible. The contacts are assigned to *power supplies* (V_{cc} , V_{pp}), *ground*, *clock*, *reset*, and a *serial data* communication link (commonly called I/O). Their specification part in the standard is being reconsidered because of requests from various parties (suppression of the two useless contacts, creation of a second I/O port, I2C bridging, etc.).

For the time being, card CPUs are still 8-bit microcontrollers, and the most common cores are Motorola's 68HC05 and Intel's 80C51, but new 32-bit devices will soon appear. RAM capacities (typically ranging from 76 to 512 bytes) are very limited by the physical constraints of the card. The program executed by the card's microprocessor is written in ROM at the mask-producing stage and cannot be modified in any way. This guarantees that the code is strictly controlled by the manufacturer. For storing user-specific data, individual to each card, the first generation of nonvolatile memories used EPROMs,² which required an extra "high" voltage power supply (typically from 15V to 25V). Recent components contain only EEPROM, which requires a single 5V power supply (frequently the same voltage used by card's microprocessor) and can be written and erased thousands of times. Sometimes, it is possible to import executable programs into the card's EEPROM according to the needs of the card holder. Finally, a communication port (serial via an asynchronous link) is available for exchanging data and control information between the card and the external world. A common bit rate is 9600bits/s, but much faster interfaces are commonly being used (from 19,200 up to 115,200bits/s) in full accordance with ISO 7816.

EEPROM size is critical issue in the design of public-key applications, as cryptographic keys must be large in order to be secure. Consequently, smart card programmers frequently adopt typical optimization techniques such as regenerating the public-keys from the secret-keys when needed, regenerating the secret-keys from shorter seeds, avoiding large-key schemes (for instance, Fiat-Shamir), or implementing compression algorithms for redundant data (text, user data, etc.). Some further employ EEPROM garbage collection mechanisms, with several manufacturers have developed real and complete operating systems for this purpose.

The security of smart cards starts with the fact that all of the card's various functions are gathered into a single chip. If this were not done, the external wires linking one chip to another could represent a possible penetration route for unauthorized access (or use) of the card. Complicating this design is the fact that ISO standards specify that the card be able to withstand a specific set of mechanical stresses such as bending and flexing. The size of the chip is consequently limited; many of the constraints on smart card functions—especially their limited memory and cryptographic capabilities—follow from this limitation.

Smart card chips are very reliable; most manufacturers guarantee the electrical properties of their chips for ten years or more. ISO standards specify how a card must be protected against mechanical, electrical, or chemical aggressions, but for most existing applications, a card is obsolete long before it becomes damaged. For example, the French phone card's failure rate is less than three per 10,000 pieces.

Figure 25.1

Smart Card Manufacturing

In general, smart cards can help whenever secure portable objects are needed and, in particular, whenever the "external world" needs to work with data without knowing its actual value. The card's tamper-proofness, combined with public-key cryptography (secretless terminals), generally provides adequate solutions to many everyday security problems.

² Electrically Programmable Memory (EPROM) and Electrically Erasable Programmable Memory (EEPROM or E²PROM); nonvolatile memory (NVM) means usually provided for data storage.

Smart Card Communication and Command Format

Communication with smart cards is ruled by the (previously mentioned) ISO standard 7816/3. Only two protocols are currently defined in this standard (byte-oriented T = 0 and block-oriented T = 1), although up to 14 are available for future expansion. These standards specify the electrical levels, frequency, and details of the protocol used to communicate between the smart card and the “external side” — that is, the rest of the world.

The minimal hardware needed to operate a card consists of:

- A mechanical interface: the *connector*
- An electronic interface: the *coupler*
- A box containing the above two elements: the *smart card reader* (or simply “*reader*”)

The simplest readers are similar to modems and manage only the ISO communication protocol without interacting “intelligently” with the operating system of the card. They are called “transparent readers” and should, at least in theory, operate with any smart card from any vendor that complies with the ISO standard.

The most sophisticated readers can be programmed with parts of the application logic and contain data (for instance, RSA or DSA public keys), files, and programs. They can execute cryptographic functions; completely replace a PC; have keyboards, PIN pads, or displays; and generally use a specific programming language. They do not support all types of smart cards, even if the cards comply with the ISO standard, because these sophisticated readers often use particular commands that are unique to given card designs.

To operate a card, the reader needs to implement the following four functions:

- Power the card on and off
- Reset the card
- Read data from the card (*get commands*)
- Write data to the card (*put commands*)

Get and put commands contain a header (actually a function code consisting of 5 bytes designated by CLA, INS, P1, P2, and LEN) according to which the card processes the incoming data. An acknowledge byte and a couple of status bytes (SW1 and SW2) are sent during (and after) the execution of each command.

Card Life Cycle

Although the card life cycle and manufacturing are described in many different sources, we particularly recommend Fuchsberger et al.'s excellent overview.³

Smart card manufacturing starts with the design of the card operating system and the application software, following the principles applying to any software for use in security applications. This is in itself a nontrivial task, but at least the memory available in smart card chips is relatively small, which limits the eventual size of the software. There have to be checks that the operating system meets its specification and also that no unintended features have been included.

The ROM mask of the operating system is then given to the chip manufacturer, which will return an implementation of the code for cross-checking before manufacturing the batch of chips. This is in itself a useful integrity check, but clearly, one normally requires this code to be kept confidential, so its distribution should be carefully controlled. Furthermore, the manufacturer has to be accountable for all chips made, some of which, because of yield failures, will need to be destroyed. Otherwise, an attacker may obtain raw chips to mount any form of counterfeit operation.

The batch of chips is distributed to the fabricator (smart card manufacturer) whose task it is to embed the chips in the plastic cards. The role of the fabricator varies considerably between customers and their services. At a very minimum, the fabricator must test the complete IC card to ensure its operational state. In some cases, the fabricator completely personalizes the card to the requirements of the issuer.

Smart Card Applications

New generations of smart cards are becoming available for applications that include pay TV, mobile communications, electronic cash, and a variety of other uses.

³ *Public-key cryptography on smart cards*. Proceedings of the International Conference on Cryptography, Policy and Algorithms, LNCS 1029, pp. 250-269.

Pay TV

One of the first applications of microcontroller-based smart cards was pay TV. The card appears to be both an ideal identification token (associated with the subscriber) and an efficient loyalty support.

In most pay-TV applications (the two best-known European examples are Eurocrypt and Videocrypt), the program provider broadcasts periodically (typically each 100 to 500ms) an encrypted *control-word* (temporary key) under which the image is encrypted. Only valid cards can extract this control-word from the data stream, decrypt it, and send it to the decoder.

A second interesting marketing model is “pay-per-view.” In this setting, the viewer buys a preloaded card and progressively spends the loaded amount (in general, a time coefficient is associated with each program).

Finally, cards also appear to be efficient sponsoring tools. In general, the sponsor issues free cards, valid only during a given event (for instance, a baseball game), and distributes them. The broadcast will then be accessible free only by the card-owners.

Mobile Communications

GSM security is essentially based on the tamper-proofness of the smart card (renamed Subscriber Identity Module, or SIM, in the standard). Each SIM is associated with a set of secret and public parameters (IMSI, Ki, PIN, PUK, etc.) that allow the operator to locate the mobile phone and route calls and digital messages (Short Message Services).

Smart cards present major security and cost advantages over passive code identification. Fraud losses due mainly to card loss or password disclosure are far less frequent than fraud losses due to password eavesdropping in non-tamper-resistant systems.

Electronic Cash

Microcontroller or authenticated-memory cards can be used for storing or representing legal tender in several ways:

- The smart card can be a bidirectional link between the cardholder and his bank account. In such a case, terminals may check the genuineness of the card by diverse cryptographic protocols (depending mainly on the terminal and card's computational powers).

- The card can act as an electronic purse and store in EEPROM a balance that can be converted or transferred from the card to the point-of-sale terminal and the bank.
- Payment protocols can be anonymous (preserve the user's privacy exactly as paper money does) or auditable (as with a regular checking account). The smart card industry considers that the banking sector will be one of the main card application fields during the coming decade.

Other Applications

Smart cards are also used in loyalty applications, electronic copyright (typically software protection), gaming, physical access control, Internet security, and many other areas.

Gemplus considers that, in the long term, the best way of fitting to clients' needs will probably consist of providing end users with a *blank card* whose ROM mask contains a general-purpose high-level operating system. On the top of that, users will add either their own (homemade) applications or ready-to-use programs bought and downloaded from a software editor. In this scenario, the card's natural tamper-resistance, combined with public-key cryptography capabilities, appears to be a natural approach to provide both passive and active software protection.

"Contactless" Cards

The term "contactless product" comprises a broad range of technologies such as infrared, optical radio, or high-frequency products. Some of those products are battery powered, while others derive their energy from the magnetic field they bathe in. In this section, we focus on radio-frequency products powered by a magnetic field.

Up to now, contactless products were used mainly in proprietary applications. This practice led to a severe lack of harmony among the communication protocols. Moreover, the products used were generally memory cards, which further limited the interest in making any studies.

The International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC) developed a standard for contactless products. This standard, named ISO/IEC 14443, defines the protocol communication between a contactless product and a reader through a magnetic field emitted by a reader. This magnetic field provides power to the contactless products and also carries the data

exchanged between the reader and the card through amplitude modulation. This standard permits the development of a range of readers and cards that will be able to work together.

At the beginning, most of the contactless products were memory based—that is, they provided storage but no processing capability. Increasingly contactless cards will be able to support processing as well. This will allow cards to be used in more security-critical applications, such as banking.

Dual-interface cards are cards that can be communicated with either through a contact or contactless interface. Moreover, such dual-interface products have expanded the scope of the contactless technology to applications like banking or e-purse, where for which the security constraints are more severe.

It is very important to understand the potential weaknesses and the intrinsic strengths of both the contactless and the dual-interface technologies. Some new attack paths or security flaws might have been introduced by the contactless technology itself or by the interoperability between the two modes.

In this scope, it is interesting to conduct a systematic and complete analysis of contactless and dual-interface products. For this reason, the following items are investigated:

- Specificities of contactless standards and protocols
- Constraints linked to the contactless interface
- Comparison of the contactless interface with the contact one
- Cryptography used in contactless products

Protocols and Secure Communication Schemes

For the past eight to ten years, contact smart cards have been thoroughly studied, and experts in physical security, cryptography, embedded O/S, Java security, protocol security, and other areas have spent a huge amount of energy assessing the security of existing applications and products. One important point related to the security of contact applications is the physical connection between the reader and the card. In many applications, physical constraints force and guarantee a one-to-one communication; that is, the reader can initiate communication with only one card at a time and a given card can respond to only one reader.

That basic property is no longer valid for the contactless and dual-interface applications, and the risk analysis has to be conducted with this point in mind. In contactless and dual products, there is no physical

link between the reader and the card, and the reader is designed to communicate with several cards at a time.

Since there is no physical link between the reader and the cards, it is simple to introduce a malicious probe, a malicious card, or a malicious reader without anybody being aware of it. One could imagine substituting a card for another; deleting some reader commands in order to shunt parts of the application, or forcing cards to perform malevolent transactions. All these configurations and scenarios have to be carefully analyzed when one is trying to build secure and trustworthy communication schemes.

Contactless applications are designed to allow data exchange between one reader and several cards at a time (the reader in fact communicates with only one card at a time, but the presence of several cards is managed through time-multiplexing). From the reader's point of view, it is therefore important to manage in a secure manner the context switch from one card to the other.

With the latest generation of dual-interface cards, the contact and the contactless applications share a core, RAM, and nonvolatile memory. Applications' sharing is something already tackled in the contact technology, but it is worth checking whether some additional protections have to be introduced when dual-interface products are involved.

Another specificity of the contactless technology relates to the high probability of having cuts in communication and power. Of course, antitearing mechanisms are implemented, even for contact cards, but the high occurrence of such events might have some impact on the design of the supported applications.

Constraints of Contactless Products

Contactless applications often have requirements that differ from those of the contact applications. The contactless interface is used mainly for its rapidity and user friendliness. Some parameters, like working distance, speed, environmental sensitivity, or interoperability, are very restrictive for contactless applications. In the following sections, we focus on those different items and analyze their impact on the secure design of contactless and dual products.

Speed and Working Distance

The feature governing speed and working distance is a key element to the design of the contactless products. In fact, the contactless power supply is assumed by the magnetic field provided by the reader. The effective power available on the card's side is linked to the distance between the card and the reader. Increasing the distance between card and reader decreases the available power; roughly 1 centimeter corresponds to a power of 1 mA. The power required for a cryptographic execution, for example—which is the highest-consuming function—is available between 0 and 7cm. When a card's chip has less power, it could not run at the maximum working frequency. The frequency's adaptation has must be finely tuned to find a compromise between speed and working distance. No limitation, except time or working distance requirements, exists for the use of cryptography on contactless products. Today, contactless products allow symmetric cryptography as well as asymmetric cryptography.

Transport applications are good examples of time-constrained applications for which the transaction time is upper-bounded to 150ms. Depending on the application, the cryptographic part is variable; it could be necessary to limit the working distance to increase speed for reaching this constraint.

Other constraints are defined in the standard ISO/IEC 14443. For example, the smart card has to work between 0 and 10cm. In that particular example, the standard does not specify if a working distance of 10cm has to be maintained even during cryptographic operations. (It is commonly admitted that 10cm cannot not be maintained on these kinds of executions using today's technology.) Nevertheless, an application designer will have to choose between speed and working distance when developing the card's operating system.

Interoperability

This notion of interoperability became relevant only after pilot applications were deployed in mass volume. Problems arose because of the analog interface between the reader and the card. This interface is very sensitive to the physical parameters of the reader and the card (reader's demodulator, card's modulator, card's resonance frequency, quality factor of the card, and so on). Depending on the variation of these parameters, a card could operate or not with a specific reader. When an application is widely deployed, there are often many manufacturers of cards and readers involved; this large number of possible configurations implies that some of them are not functional. Finding the test-table's parameters that guarantee that a card or a reader will be "interoperable" is a big issue for the deployment of contactless applications.

Contactless Products and the Contact Interface

Contactless applications have some specificity that could be summarized under three headings: communication, physical security, and software security.

Communication

A contactless product communicates with the reader without any physical link. This communication channel could be a potential weakness in terms of security; it will be interesting to identify the adequate countermeasures.

Lack of Physical Link

The absence of physical link between the card and the reader allows a “user-friendly” utilization of the smart card; cardholders can keep their cards in their wallets, for example, when they use them. This characteristic also requires a lower cost of maintenance for the reader, but it has its own weaknesses. The absence of any physical link forces the communication between the card and the reader to be secure and trustworthy.

Some contactless applications are designed for a possible communication between a reader and several cards at the same time. This means that reader can alternatively talk with one of the cards in front of it—cards cannot talk to each other, even through the reader. The parts must be able to trust each other and be sure that no part has been perverted. Well-known countermeasures, like mutual authentication, have to be applied. A mutual authentication permits the card and the readers to prove that they are authorized to speak to one another. The exchange of a shared secret, for instance, shows that the card and the reader are what they claim to be.

In contactless communications, the card could be (in)voluntarily removed from the magnetic field. For example, the “moving” card in front of a reader antenna could generate this natural “instability” of chip power supply. From the card operating system’s point of view, this characteristic has to be included during the development. For example, some backup mechanisms could be implemented to restore the context after tearing.

The lack of any physical link also allows a malicious system to hide some commands sent by the reader to the card, with the aim of blocking an action in the card; for example, the malicious system could try block a decrement command sent to the card with intent to undecrementing an electronic purse. Blocking this command would prevent the amount of money in the card from being decremented after a purchase! From

the reader's point of view, the command to decrement the counter would have been sent, but from the card's point of view, no money would be removed. Some systems try to detect this kind of attack by querying the card to make sure that the decrement command has actually been executed—for example, by reading the card's value a second time to make sure that the card's value has actually been decremented.

Spying

Because contactless communication is done through the magnetic field, it is easy to catch and spy on the exchange of data between the card and the reader. No countermeasure can block this spying. On the other end, some countermeasures like data ciphering can block the capacity to interpret the data in a way to use it in some command reply scenarios.

Use of a Smart Card Without the Cardholder's Consent

Contactless communication allows a card to initiate communication without a clear engagement of the cardholder. We could envision that a malicious reader could be put in front of a smart card without the knowledge of the cardholder and collect some information from the card. This privacy issue is the main threat that inhibits the acceptance of the contactless communication. Fortunately, some simple countermeasures, as already explained (mutual authentication between the card and the reader, PIN code, etc.), are efficient against this risk. Some other solutions exist, like a push button that stops the card from functioning until the cardholder takes the card in hand or some specialized encasings that shield the card from the magnetic field of a malicious reader.

Card Holder's Privacy

The ability to read a contactless card's content without the explicit agreement of the cardholder can be a serious threat to the owner's privacy. This point deserves to be mentioned even if its social and legal impacts may vary from one country to the other. The countermeasures already mentioned limit this possibility, but it is important to say that the low-level commands in the ISO 14443 standard require that a card always replies to a command from the reader. The first response sent back by the card to a reader command is a serial number. This information could present, with some cross-checking method, sufficient information to outline a person's movements.

ISO standards limit that threat, allowing the card to send back a random serial number in a countermeasure that blocks the link between a card's serial number and a "named" cardholder. We could also imagine that the card is put in a shield case.

Denial of Service

We saw that it is possible for a malicious organization to collect some information without the cardholder's consent. Another effect of a malicious use could be a card malfunction. Imagine that a malicious reader is designed not to collect information but to block the card by physically destroying it through an electromagnetic spike or presenting the wrong PIN code many times, blocking the correct PIN code presentation attempt. This attack could have a big effect in an airport or on a subway, for example. A denial of service could damage all mass transit systems, forcing them to switch to a backup system with less security or transit flow speed.

Physical Security

We saw previously that almost all attacks could be blocked with well-known countermeasures if the threat is correctly identified. From an application point of view, the contactless communication is as secure as a contact one. We now have to look at security from a physical point of view. For contact products, some attacks using the card's leakage are well known (such as Simple Power Analysis and Differential Power Analysis). The literature mentions a lot of possible attacks that use power analysis or fault analysis. Many software countermeasures were developed against these threats, and their efficiency is independent from the communication channel. On the other hand, some chip manufacturers developed hardware countermeasures to blur the chip's current consumption. These countermeasures have to be correctly studied and evaluated to determine their efficiency in the contactless mode.

For dual-interface products in particular, we need to study the security of the contactless interface to ensure that this interface does not bring in new paths of attacks. It may be that additional design countermeasures have to be implemented by chip manufacturers to guarantee a high level of resistance against fault attack.

Software Security

We already spoke about the possible compromise in performance with respect to working distance and speed. From the cryptographic point of view, it is the same stake. All known cryptographic algorithms could be performed on contactless products; technically there is no limit on contactless products. Chip manufacturers choose to clock the chip with low-frequency values. This choice allows the cryptographic algorithms' computation with an acceptable working distance by lowering the amount of power required, but

it simultaneously increases the length of time required for computation. Other technical choices could reduce the time of computation. This time is often the cause of the assumption saying that cryptography is not possible in contactless applications.

When considering the range of available security mechanisms, the primary factor determining whether or not they can be adapted to contactless cards is the amount of electrical power that is required to implement them.

Conclusions

The number of contactless applications is rapidly growing. However, some customers are hesitant to use them because of concerns about security and the limited knowledge regarding the specifics of contactless technology. The good news is that adequate security measures do exist; concerns over security should not be an impediment to deploying this promising technology.

Of course the recent introduction of dual- interface smart cards, and the utilisation use of this kind of products in high-level applications, increase the level of confidence needed in this technology. A good identification of particularities details permits to shows that contactless is as secure as the contact technology. We think that dual-interface products represent opportunity to combine the strengths of both contact and contactless cards with technology that is available today.