S M A R T C A R D

A HIGHLY RELIABLE AND PORTABLE SECURITY DEVICE

LOUIS C. GUILLOU CCETT/ASP Rue du Clos Courtel BP 59
35530 CESSON SEVIGNE - FRANCE

MICHEL UGON BULL CP8 Rue Eugène Henaff BP 45
78193 TRAPPES CEDEX - FRANCE

ABSTRACTS :

At first glance, the smart card looks like an improvement of the traditional credit card.

But the smart card is a multi-purpose and tamper-free security device. And behind a standardized interface, the built-in electronics may evolve, in memory size and in processing power. This evolution, while resulting from economic considerations, is in tune with an enhancement of both physical and logical security.

Some mechanisms in key-carrier cards are described, thus giving a taste of the state of the art in card operating systems. The underlying reality is an invasion of our lifetime by cryptology and computers. This invasion will have a large influence on security in various fields of applications, not only banking operations, but also data processing, information systems, and communication networks.

INTRODUCTION :

Including magnetic stripes and embossed areas, a banking IC card looks like a gadget plastic credit card. But, embedded in the thickness of its plastics, there are one or more integrated circuits designed to perform both **processing and memory functions** .

Magnetic stripes and embossed areas, traditional technologies of credit cards, are deprived of any processing power, with only memory functions ; the relevant standards specify every details without any degree of freedom for further evolutions.

An IC card is not defined by IC number, position, or performance. On actual cards, the interchanges with the outside are conducted through electrical contacts ensuring a galvanic continuity between built-in electronics and an external interface device. Other types of IC cards may be developed in order to avoid contacts. But an IC card is and will be defined through a standardized interface taking into account its **processing power** .

Trade-offs between costs and performances are evidently related to a current state of the art and to current needs of the applications. Technological evolutions deal with power consumption and integration scales, resulting in a tremendous increasement of both memory size and processing power. Behind a standardized interface, **built-in electronics may evolve** while terminals remain unchanged.

Reductions in both engraving size and power consumption will complicate physical investigation of processors dedicated to smart cards, the famous Self-Programming One-chip Microprocessors, abbreviated as SPOMs : existing SPOMs are today tamper-free, and we don't know any successful violation of their transaction memory.

Additions in processing power (CPU and RAM) and in operating systems (ROM) will complexify the logical security, and allow the use of more complex and more various cryptographic algorithms. While unpublished and proprietary, the algorithm named Telepass2, and used in French bank cards, has been successfully evaluated by a notoriously specialized agency. Current SPOMs are already able to implement very secure algorithms.

In semi-conductor industry, technological trends are in tune with security. This exciting situation is very new! And at ISO level, the general approach of this interface takes into account these potential evolutions.

CONTACT LOCATION AND ASSIGNMENT :

In existing ISO standards, credit card surfaces had been partially reserved. It seems impossible to locate contacts on magnetic stripes or in embossed zones. Marketing considerations have been expressed in the US so as to share the front part of a credit card between the issuing bank and the credit card company. Moreover, Japan has chosen a national magnetic stripe position on card front side, in disagreement with ISO standards. The sum of these constraints explains the difficulties met by ISO in its international quest for an agreement on contact location.

ISO has reached a first basic agreement on contacts : number, minimum dimensions, relative position and assignment. On existing smart cards, the outside accesses the built-in electronics through six electrical contacts. With respect to GND (ground) as a reference voltage, the outside must provide VCC (supply voltage), VPP (programming voltage), CLK (clocking signal), and RST (reset signal) with suitable signals in order to exchange information on I/O (input/output). Two spare contacts, named RFU, are reserved for future use (see figure 1). This agreement protects the existing dedicated chips and the existing ways to package electronics in the cards.

Nevertheless, two 8-contact positions are yet described in the DIS (Draft International Standard, now under ballot). One beneath the other, the set of sixteen contacts forms a regular pattern located relatively to a corner, in such a way that mixt contactors are easy to design and to produce. In France, mixt contactors are being inserted in public telephone booths as well as in interface devices to be connected to Minitels. Extensively used in France, the upper position is mechanically more reliable : a card is more resistant to

bending with a chip closer to a corner. But the lower position is more in tune with Japanese constraints, and allows to locate the contacts on the back side. While being the traditional one, the upper position is said to be "transitional" in the ISO documents.

The standardization aims at an international agreement on a complete and unambiguous specification of this interface, not only : physical characteristics, contact location and assignment, but also : electrical signals, answer to reset, exchange protocols, (now also at a DIS stage) and interindustry requirements (now under study). ISO assigned in 1981 these general tasks to subcommittee SC17 "Identification Cards" in technical committee TC 97 "Information Systems".

In 1985, ISO entrusted TC68 "Banking" with two new specific work items on security and data contents related to banking operations. The adoption of these new work items was felted so important that it produced a rearrangement of TC68 with creation of a new subcommittee SC6 dealing with "Financial transaction cards, related media and operations".

SOME TECHNOLOGICAL ASPECTS :

As a result of a transaction, the card delivers information (stored data, computation results), and/or modifies its content (data storage, event memorization). The built-in electronics always include an electrically **Programmable Read Only Memory** (PROM). Each PROM cell originally in state "1" may be turned to state "0" by an electrical process under control of the built-in electronics. This PROM contains the **transaction memory**, the content of which evolves during card life.

Two major technologies are currently producing PROM components :
bipolar and **MOS** (metal oxyde semiconductor).

Though quicker, bipolar logic is more power consuming and less easy to integrate than MOS logic. But more important : the bipolar writing process destroys a part of the memory cell, in such a way that bipolar PROMs are optically readable (see figure 2)!!

Nethertheless, at the outbreak of IC cards, bipolar technology has been fairly considered : the reason was the irreversibility of the writing process.

No physical method is known to investigate MOS PROMs so as to directly visualize a cell content without using the internal buses : the MOS writing process is reversible ! And the existing MOS PROMs can be erased either by ionizing radiations, such as UV-light or X-rays, or by another electrical process. All existing cards are now using MOS components which are cheaper, more secure, and allow to match microprocessor technology with a PROM memory.

The technological controversy lies now in the comparison between UV-erasable PROM, named EPROM, used as write only memory, and electrically erasable PROM, named EEPROM, and so rewritable. When widely available, EEPROMs are felted to be more flexible than EPROMs, but with much less capacity for a given die size. Actually, EEPROM technology is less mature and more expensive.

Other parallel technological evolutions are in progress :

- scales of integration move from a range from 4 to 3 μ to a range from 2 to 1 μ ;
- uprising of CMOS and HC MOS considerably reduces power consumption in the cards.

These reductions in engraving scale and in power consumption will increase considerably the difficulties in investigating the PROM contents. And to close the general considerations, let us remember a generally agreed limitation : to obtain a reliable card, the size of the chip must not exceed 20 mm².

A FAMILY OF IC CARDS :

Under control of the built-in electronics, the content of the transaction memory evolves during card life. Depending on the increasing complexity of this electronics, three types of IC cards are currently in use (see figure 3) :

- **Memory cards** containing only a transaction memory with very simple writing protections.
- **Logical cards** containing a transaction memory and a logic array of gates : this logic array of gates tests a confidential code before giving access.
- **Smart cards** containing a programmed microprocessor controlling itself all the accesses to the transaction memory.

In France, these three types of IC cards are illustrated by :

- Memory cards : 40-unit or 120-unit prepaid "Télécartes" anonymously used in public telephone booths ;
- Logic cards : "Télécommunications" identifying their bearer in order to charge phone calls on a number ;
- Smart cards : bank cards and key-carrier cards which may also be used in public telephone booths !

It is inefficient to search for relations between applications and types of cards : the three types are in use in the telephone system, while bank cards are designed such as to easily extend their use in various services.

- The simpler cards are specific to only one purpose, and it is rather difficult to share a chip production between several applications.
- On the other hand, the smart card is essentially a multi-purpose device. And the chips are programmed by mask during the manufacturing process. There is no difficulty to share a chip production. The development of a new mask is easy, and the same line produces chips, whatever the mask be !

This paper is devoted to smart cards including a specific integrated circuit, named SPOM, a microcomputer which merges on the same chip a PROM memory and a microprocessor controlling itself all the accesses to the PROM memory.

A NEW CHIP : THE SPOM

Including two chips (Fairchild 3870 microprocessor + Intel 2716 EPROM memory), the first smart cards were produced in 1979 after a strong international cooperation between MOTOROLA Inc. and CII HONEYWELL BULL.

This stage of a two-chip card is essential in order to prove the feasibility and to convince potential users to start experiments. These devices played also a prominent part in the development of the various other elements of the systems using smart cards in order to initiate applications.

Economical considerations led to merge PROM and microprocessor on the same ship. And the cooperation between BULL and MOTOROLA continued by the studies of a new microprocessor dedicated for smart cards. Such a microprocessor must be able to execute an internal routine which writes in its transaction memory. A new architecture has been invented to manage registers on the internal buses in such a way that the processor may continue its control while holding the right address and the right content on the buses towards the PROM. Such microprocessors are named : Self Programming One-Chip Microprocessors, abbreviated as "SPOMs" (see figure 4).

Since 1981, SPOM01 is produced by MOTOROLA Inc. in East Kilbride (Scotland, United Kingdom) ; since 1985, SPOM02 is produced by THOMSON EUROTECHNIQUE in le Rousset (Provence, France). Trade-offs between costs and performances are largely indebted to the know-how gathered from the first two-chip cards. Both in nMOS technology, they are about 17 mm² in size. BULL CP8 and PHILIPS are currently manufacturing cards with such SPOMs. Very recently, a prototype (30 mm²) comes from HITACHI, referred as 65901, while SPOM03 and 04 were announced by MOTOROLA and THOMSON.

MEMORY SIZES (in bites)

	C.P.U.	RAM	ROM	EPROM
SPOM 01 MOTOROLA	6805	36	1600	1024
SPOM 02 EUROTECHNIQUE	8048	44	2048	1024
65901	RISC	128	3072	2048 (EEPROM)
SPOM 03	6805	52	2048	2048
SPOM 04	8048	64	3072	4096

In addition to PROM memory, SPOMs contain also ROM written by mask during chip manufacturing process and RAM to store temporary results. Let us notice on SPOMO2 that a cell of RAM is roughly ten times larger than one of EPROM, and a cell of EPROM twice larger than one of ROM (see figure 5).

Before cutting the wafers on SPOMO2 production lines, a 512-byte internal routine is activated through a seventh test-contact. Each validated component receives various information : locks, codes, and a chip serial number, while nothing is written in rejected components. The test-contact is then systematically destroyed, thus definitively disabling invalid components. As a matter of fact, only the self-testing routine may write the witness indicators tested by the card before execution of any command during a transaction with the card.

ADDITIONAL SECURITY FEATURES :

Absolute physical security does not exist, no more for smart cards than for any other computing device. System designers must consider potential consequences of violations. Secret keys in a system must be as as diversified as possible, and in a user card, tied to the chip serial number. A violation then results in an attack against only one user and does not endanger the whole system, thus reducing the potential benefits from fraud. These aspects of logical security are strongly related to cryptology.

In key-carrier cards as well as in bank cards using existing SPOMs, the transaction memory is organized in 32-bit words. During card issue, an issuing block is written in a dedicated 6-word zone. This block stores the secret distribution key on four 32-bit words. Each card issuer owns his secret cryptographic function. He uses it to compute or to recompute a unique secret distribution key from each chip serial number. This key, unique for each card, must be correctly used hereafter to control various operations, such as writing new secret words and delivering new authorizations. Very different from a confidential code, this cryptographic key is used as a parameter in a cryptographic computation prescribed in internal routines and executed by the card itself.

The card issuer may remotely and securely authenticate a card : a random value is sent to each calling card which must answer both its chip serial number and the result of a computation using the distribution key. From the chip serial number, the issuer reconstructs the distribution key and then the computation result, thus authenticating the card.

The card issuer may also identify the card user by including a user confidential code during the authentication process : the code given locally modifies the random value sent by the issuer. In a first solution, only the card knows the code : the card modifies the incoming value by an internal code, and the internal modification must cancel the external one. In a second solution, the issuer modifies the random value before sending it to the user : the external modification must then restore the initial value.

In addition to these functions, when the cryptographic computation in the card reverses an external cryptographic computation performed by the issuer, the card may identify its issuer before remotely executing its directives. At the end of a cryptographic computation, the card tests the result : for example, when the 64-bit result consists of two identical 32-bit fields, the card assumes that only its issuer might induce such a result. So the issuer is now the real master of the silicon, because each SPOM may securely and remotely identify its master.

KEYS, AUTHORIZATIONS, ENTITLEMENTS :

First developed in a television environment in order to access broadcast information such as picture, sound and data, the key-carrier cards are deprived of any banking functionality. The conditional access method may be general, while the scrambling method clearly depends upon the nature and the coding of the service components. The key-carrier cards are usable in access control to a large range of services and resources, including terminals, network gates, databases, computers, and even buildings.

Materializing entitlements, these cards are issued and managed by card issuers who are their real owners. The card issuer must be clearly distinguished from both the bearer and the service provider ; the bearer uses the card to recover control words, but cannot alter the card, nor get a copy of recorded keys ; the service provider checks entitlements by verifying their validity, by storing a debit, by consuming a credit ; the card issuer manages entitlement in his cards by delivering new entitlements, by clearing debits, by giving credits.

The key-carrier cards store blocks of authorization, each one consisting of three fields : an identifier, a status and a secret key. Each operation on an entitlement, as well checking as management, results from a transaction with the card. During this transaction, a command asks for a cryptographic computation with incoming data consisting of three fields : an identifier, a parameter, and a cryptogram.

During an entitlement checking transaction, the identifier in the incoming data must correspond to the identifier of an authorization in the card, the status of which must comply with indications given in the parameter ; then the card reconstructs the control word by a cryptographic computation using the secret key of the authorization. The outgoing data in the command asking for the result is a control word which is either sent back to the controller as a witness, or used locally to descramble subsequent service components.

During an entitlement management transaction, the card uses the distribution key, unique to each card, to execute a cryptographic computation. The card tests for a redundancy in the 64-bit result which must consist of two identical 32-bit fields. The card, having identified the voice of its master, executes the directive instructed by this field.

Depending on the way the card and the system manage the status of the authorizations, there are various entitlements : -fixed and renewable subscriptions, -prepayed special events, -pay-per-views either in a prepayed credit or with a limited debit, and -consumptions of tokens in activable blocks.

FOUR TYPES OF AUTHORIZATIONS :

The status of a basic authorization consists of an initial number, coding a starting date, and a gap, coding a duration. The parameter is an operation number, coding a current date. Before reconstructing a control word, the card verifies that the current date lies in the subscription period. The status of such an authorization is fixed and can only be verified. This first type of authorization is a fixed subscription.

The status of an authorization with controlled sessions consists of an initial number, a gap, a limit and the content of a zone reserved for sessions. Each session is defined by a shift from the initial number and a width. The sessions must lie in the main period and the sum of their widths must not exceed the limit. Before computing a control word, the card verifies that the operation number coded in the parameter lies in a session. Only a successful entitlement management transaction may modify the status of such an authorization by opening a new session (shift and width).

Depending on the way the system manages the numbers, this second type of authorization is used to access prepayed special events. It may also be used like a renewable subscription : the renewal is obtained by opening a new session. For example, depending on the value

of the session width, when numbers are coding weeks, the subscription is either on a monthly basis with width = four, or on a quarterly basis with width = thirteen.

The status of an authorization with impulse sessions consists also of an initial number, a gap, a limit, and the content of a zone reserved for sessions. Before reconstructing a control word, the card verifies that the operation number coded in the parameter lies in a session. In the absence of suitable session, the card, after an explicit agreement of the user, opens directly a new session (a shift and a width) according to indications given by the parameter as long as the sum of the widths does not exceed the limit.

This third type of authorization is a pay-per-view working either on an additive basis (when the limit is an authorized debit), or on a subtractive basis (when the limit is a prepaid credit), depending on the payment.

The status of an authorization for consumption consists of an initial number, a gap, and the content of a reserved zone divided into blocks of tokens. Only a successful entitlement management transaction may open new blocks. Before reconstructing a control word during an entitlement checking transaction, the card consumes the amount of tokens indicated by the parameter. This fourth type of authorization is based on a consumption of tokens in provisions selectively activable.

THE EXISTING KEY-CARRIER CARDS : KCO AND KC1

The first version of key-carrier cards, KCO was developed as a basic tool in pay-TV systems. The MOTOROLA SPOM with such a mask is available since 1983. And these specifications were used to test the THOMSON SPOM in 1985. KCO, now available on both SPOMs, includes the first three types of authorization : subscription and both sessions.

An unpublished cryptographic algorithm, named Twisted Double Fields and described in about 200 bytes in the mask, reverses another expanding external algorithm used by the card issuer to compute entitlement management messages, and by the service provider to compute entitlement checking messages. In the card, a 61-bit result (for example, a control word) is computed from a 23-bit incoming parameter, a 127-bit incoming cryptogram, and a 127-bit internal secret key.

The second version of key-carrier cards, KC1, developed on the THOMSON SPOM, extends KCO functionalities, taking thus advantage of the additional ROM in SPOM02. KC1 introduces the fourth type of authorization with consumption of tokens. KC1 includes also a cryptowriting mechanism : the redundant result of a cryptographic computation using the distribution key indicates the address and the content of the 32-bit word to be written.

An unpublished cryptographic permutation, named Vidéopass, is described on about 200 bytes in the mask of KC1 which may a-priori execute the algorithm in both directions but locks restrict user cards to one direction. A 64-bit result is computed from an incoming 64-bit cryptogram, an either incoming (a parameter) or internal (a non-secret word) 32-bit argument, and a 96-bit internal secret key.

But, more important yet, these chips are specialised by locks written during card issue. Most of the cards are restricted to one direction of the cryptographic computation, while some cards keep both directions. To reverse a computation performed by a user card, the master card must have stored the same secret key written in the reverse order.

Thus, the cryptographic algorithm is dissymetrised :

- the user cards compute certificates on incoming redundant data, and the master cards can only verify their genuiness by recovering the redundancy in the result, without being able to forge certificates.
- the master cards securely produce cryptograms of control words and entitlement management messages : the user cards reconstruct the control word or executes the directives given by the master, without being able to forge management messages.

CONCLUSION :

This approach to smart cards, and more specifically to key-carrier cards, has deliberately excluded the banking purposes, in order to avoid the credit card syndrom. But in the existing banking smart cards, many key-carrier card concepts have been introduced : distribution key, authentication, identification, cryptowriting,... summarized by the user/master mechanism.

The cryptographic computation uses rather proprietary algorithms. But a new version of key-carrier cards, under development by PHILIPS, is implementing the DES. And future chips, with a more powerful CPU and at least four times the actual amount of RAM, will implement public key algorithms. The best way to protect the secret parameters of a public key signature scheme seems to store them in a user-friend and tamper-free device containing a SPOM, with high security features.

FIGURE 1

FIGURE 2a

FIGURE 2b

Let us compare the destroyed fuses on the bipolar memory (fig 2a) with the regular pattern of the nMOS memory (fig 2b).

FIGURE 3

FIGURE 4

FIGURE 5