Data protection: multi-application smart cards: the use of global unique identifiers for cross-profiling purposes - Part II: towards a privacy enhancing smart card engineering

Jean-Marc Dinant and Ewout Keuleers, University of Namur (Belgium)

A multi-application smart card scheme has to involve a win/win approach for all parties involved

In the first part of this paper,[1] the underlying privacy concerns of multi-application smart card schemes were identified. In relation to the use of global unique identifiers, the regulatory framework for the smart card manufacturer was assessed. To demonstrate that multi-application smart card technology can be reconciled with the principles of data protection legislation, in particular with the requirements that personal data be processed fairly and lawfully,[2] three technical solutions will now be analysed and commented upon.[3] The first two solutions will relate to the role and functioning of the scheme's application providers. The third solution consists of the development of a less privacy-infringing smart card technology in relation to the use of unique identifiers.

A. Introduction:

Generally speaking, information technologies are not privacy-killing by substance. They become privacy killing by side-effects, precisely because a particularity of a technology invented or implemented for a particular and legitimate use appears to be re-routed for another purpose. In fact, from a data protection viewpoint, engineers may very often appear to be very naïve. From their side, advocates of privacy rights seem to be extremist and unable to understand technological requirements and the technical feasibility of certain solutions.

The subtle - and difficult – approach in the SmartCities project was a collaborative one by engineers and data protection specialists. In the framework of informal workshops, these two opposite sides of the spectrum were given time to explain their concerns. Although it may have taken time to reach a common understanding, all participants agreed on two fundamental principles:

- The general data protection Directive 95/46/EC is no longer negotiable and such workshops cannot be the forum to re-write or to amend legal requirements. Furthermore, roughly speaking, given the substantial financial impulse offered by the European Commission to the SmartCities project, the legal and mandatory European requirements could not be ignored;
- The protection of privacy and personal data has its price. At a first glance, we may consider that a legal constraint tends to reduce the benefits of the deployment of such a multi-application smart card. However, this is a short-term vision. A multi-application smart card scheme has to involve a win/win approach for all parties involved. Card holders and citizens will have little confidence in the smart card if they have to renounce to a substantial part of their individual freedom to get the benefits of the scheme concerned. In the long term a privacy-killing scheme gets a bad reputation and dramatically raises the level of Big Brother paranoia. As a concrete result, this may lead to a dual society.

The collaborative approach described above has been a success thanks to the constructive spirit of all the partners within the SmartCities consortium and they should be given the credit they deserve. As a result, the tangible output was a team of privacy-minded engineers and technology-minded lawyers working together to develop a privacy-friendly multi-application smart card scheme. The three scenarios described below constitute a concrete output of this techno-legal brainstorming.

B. Scenario one: avoid cross-profiling by the use of symmetric keys

1. Presentation of the solution

A smart card is very often identified by a serial number readable by software, *i.e.*, by all the applications embedded in the card and by the smart card terminals. At the first glance, a very simple way to prevent global cross-profiling

Computer Law & Security Report Vol. 20 no. 1 2004 ISSN 0267 3649/04

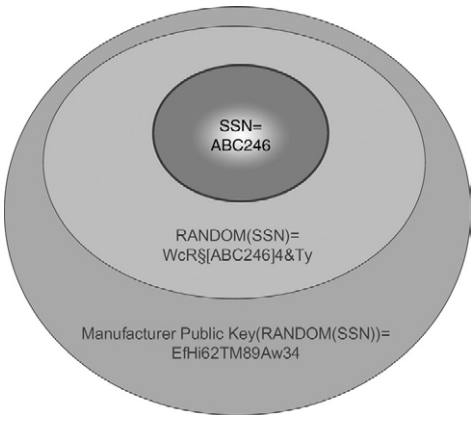
activities by the data warehouse consists in the encryption of the so-called Smart card Serial Number[4] by a secret key.[5] Each application provider will have its particular key to scramble and de-scramble the unique SSN.

In this scenario, a customer is no longer identified by the Smart card Serial Number but by an encryption of this number, proper to each application. As a concrete result, data issued from different application providers cannot any longer be merged on the simple basis of different encrypted numbers.

The main advantage of this approach is that each transaction stored in the data warehouse will be characterized by this particular SSN and the identity of a card holder will be diverged in multiple pseudonyms, depending on the number of application providers. Therefore, by default, the data warehouse will not able to link different pseudonymous SSNs to each other or to the original SSN of a particular data subject. In other words, it is not possible to discover that bus traveller BCD198 is the same person as library reader AEF435, nor the same as the swimmer ABC634 or the person holding the smart card with the unique Smart card Serial Number ABE403.

It has to be underlined that this method still allows each application provider to have a global overview of its own customers' transactions, but, as in a mono-application smart card, such an overview is limited to its own application(s).

Figure 2: Pseudonymous SSN by using a secret key proper to each application[6]



Nevertheless, it should be emphasized that, according to recital 26 of Directive 95/46/EC, scrambled personal data remain "personal data" and thus benefit from the protection offered by this Directive.[7] Although it is very unlikely that the data warehouse, *i.e.*, a data processor, will be able to identify the data subject via an encrypted SSN, an application provider can identify the data subject by using reasonable means. For instance, by applying the same secret key, initially used to scramble the unique SSN, the application provider can de-scramble the pseudonymous SNN and get back the original SSN.[8]

2. Permitting a local cross-profiling by using cryptography

Directive 95/46/EC states that personal data i) may only be processed fairly and lawfully and ii) that they may only be collected for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes.[9] In the light of these two principles, it will not be legitimate for any application provider to get access to the personal data concerning the use of a smart card from other applications. This implies, *inter alia*, that a library has no legitimate interest to know all the details of the use of public transport by library visitors. However, in some circumstances a limited cross-profile between specific applications may be considered legitimate. It could, for instance, be perfectly legitimate for a library to try to have an overview of the means of transport used by card holders visiting the library. Perhaps the library can then offer a price rebate for the use of public transport, for instance, to avoid traffic jams or to encourage people with a low revenue to go to the library. This purpose seems legitimate for both the library and the transport company, i.e., two distinct data controllers.

Unfortunately, the data warehouse in the configuration described above cannot technically achieve this global overview, because the application providers do not share the same ID of the same card holder. In fact, they do share the same ID, but in such an encrypted way that the data warehouse is unable to merge data relating to the library visitor with transport data on a one-to-one basis. This would have been possible if both the library and the transport company communicated their respective secret keys to the data warehouse. It is evident that this solution is difficult to advocate or to accept, to the degree that this key should remain secret and cannot be divulged to other persons than the application provider. After a few months the data warehouse will very likely know the secret keys of all the application providers. By linking secret keys to each

other, the data warehouse will be technically able to get a global overview of each citizen's behaviour.

In this regard, some recommendations concerning the cross-profiling of personal data can be brought forward, while respecting one's privacy. Furthermore, it should be emphasized that cross-profiling is not a general or systematic rule within a multi-application smart card scheme. In principle, and with regard to data protection issues, all data collected by the data warehouse from one application provider have to be considered as stored in a waterproof compartment, isolated from the data compartments of other applications.[10] In particular circumstances, however, the merger of secured compartments can be performed, provided i) adequate privacy safeguards are adopted and ii) to the extent that the purpose of this punctual cross-profiling matches with the initial processing purposes.

Cross-profiling is not a general or systematic rule within a multi-application smart card scheme

An *a priori* a cross-profiling procedure has to be set up between the data warehouse and the application providers embedded in smart card schemes. The following privacy-compliant procedure can be proposed:

- The application providers agree on the questions on which they want to get an aggregated answer, e.g., how many library customers are using the bus to go to the library? What is the average distance between the library and the library visitor's home? In this scenario, this is crucial. The library may be funded to know how many readers are taking the bus to go to the library, but certainly not to know all the details of bus transportation not linked to a visit to the library;
- Application provider A and application provider B define the transactions, e.g., by means of a transaction table on which they intend to achieve a cross-profile, *e.g.,* by time-slicing, the location of the card reader, the profile of the card holder, etc;
- The data warehouse extracts the relevant data for applications of A and B;[11]
- Similar to the session key of the SSL Protocol, the data warehouse randomly generates a session key;[12]
- The data warehouse communicates the relevant scrambled serial numbers to application providers A and B, by using the secret session key;
- Each application provider applies his own secret key to get back to the original SN. The SN is then encrypted by each application provider by using the common secret key;
- New scrambled SN are sent back to the data warehouse. Those SNs are now common to both application providers. The data warehouse merges the two transaction tables to achieve a one-to-one global cross-table;
- The data warehouse uses relevant data mining techniques to answer the questions asked by application providers;
- The results are communicated in an aggregate form to the application providers concerned.

3. Pro's and cons

By setting up such a kind of procedure, both the data controller and the data processor will be able to take the benefit from punctual and aggregate cross-profiling actions, while at the same time guaranteeing a high level of personal data protection. However, this solution does not offer the highest level of protection against systematic cross-profiling as far as the data warehouse remains technically able to link virtual identities of the same individual by, *e.g.*, performing the above mentioned matching transactions.

Although the strength of this solution lies within the secret key used to scramble the unique SSN, the same secret key remains the Achilles tendon. As indicated above, the scrambled SSN can be de-scrambled with the same convenience as it was scrambled in the first place. This not only demonstrates that collected data remain personal data, but also that the built-in security measures can be neutralized very easily. In this regard, reference can be made to the Article 29 Working Party's opinion on the Liberty Alliance Project. Although in the latter project the unique identifier is replaced by pair-wise identities, the Article 29 Working Party underscores that identities can still be shared among the participants.[13] For this reason, personal data protection concerns remain eminent and alternatives with a higher degree of security should be considered.

C. Scenario two: the use of fully anonymous data

1. Description of the scenario

Instead of encrypting the unique SSN, and thus creating a pseudonym for the card holder for each application provider concerned, a more privacy-friendly scenario consists in the transmission of pure anonymous data to the data warehouse. Furthermore, each application provider, for his own purpose, will use a proper ID without any link to the SSN.

Application providers will transfer all data generated by the use of a smart card to the central data warehouse, this irrespective of the information it contains. In this scenario, a distinction should be made between two types of data. On the one hand, the use of the smart card, e.g., a cash withdrawal from an ATM machine, will generate data concerning the person withdrawing money from his account, while on the other hand, data concerning the financial transaction will be generated and processed, e.g., the place of the ATM, time, amount, etc.

Figure 3: Personal data vs anonymous Transaction data

PERSONAL DATA	TRANSACTION DATA
All data relating to the identity of the card holder, e.g., identifying number, name, alias, account number, phone number, address, etc	All data relating to the transaction without including any reference, even indirect, whatsoever concerning the identity of the card holder

The legal regime for these two categories of data is very different, if not opposite.

According to recital 26 of Directive 95/46/EC, the principles of data protection will only apply to information concerning an identified or identifiable person and will not apply to data rendered anonymous in such a way that the data subject, i.e., the card holder, is no longer identifiable.

In this regard, application providers will only send transaction data to the central data warehouse, this by stripping all data referring, even indirectly, to an identifiable natural person. Therefore, the data warehouse will only contain anonymous transaction data, not subject to any legislation in the field of personal data protection. In consequence, application providers are entitled to cross-profile data stored and processed in the data warehouse without, e.g., considering the purposes or legitimate character thereof.

From a data protection viewpoint, this solution is quite satisfying as long as sufficient guarantees are given that the data processed in the data warehouse are and remain anonymous. In this particular case, the personal Data Protection Directive does not apply.

2. Socio-demographic granularity

Data warehouses may find it difficult to generate value-added and usable information from anonymous data. Therefore, before sending the anonymous data to the data warehouse, application providers may add general information to the transaction data, e.g., gender, age category and neighbourhood. Of course, particular attention should be paid to the *granularity* of such a kind of socio demographic data. If the crossing of the various criteria may lead to a very small group of individuals or to one single individual,[14] the data are no longer anonymous.

This granularity can be reached by sufficiently big slices of range. What is the added value of knowing that a woman is born 4 July 1950? Should it not be enough to know that she is in the age category of persons between 50 to 55 years? A convenient solution may be to require that every crossed category contains at least a hundred or more individuals.

3. Pro's and cons

This scenario appears to be well balanced. Application providers have their own customer ID and they are still able to analyse the behaviour of their own customers inside their own application. Furthermore, by transferring pure transaction data, i.e., anonymous data, to the data warehouse, they participate in - and benefit from - a global aggregate view of what is happening in the multi-application smart card scheme.

From a data protection viewpoint the scenario may appear as being highly idyllic. No more personal data are processed by the data warehouse and no individual cross-profiling is done. Nevertheless, this approach seems to us to be a little bit too naïve.

In fact, the Global Unique Identifier still remains accessible on the card by every application running on it and it would be very surprising that every application should ignore it. In the present scenario only organisational measures have been adopted to avoid an excessive cross profiling between applications. This has been done on the basis of article 17 of the general data protection Directive 95/46/EC. [15]

However, this is not sufficient. The same article of this Directive requires that in addition to the organisational measures, "technical" ones be adopted. This combination of measures is perfectly coherent with the concept of security. Security cannot be exclusively based on human goodwill. Although codes of good practice are useful and necessary, they cannot or may not replace the technical measures. In a certain way, it makes sense that application providers agree not to use such a convenient global unique identifier

but, in the end, who will monitor that they do so in practice? Moreover, if it appears that the SSN is used among various applications, which will take care of that? That is why technical measures have to be put into place: not to replace organisational security measures but just to enhance them.

D. Scenario three: replacing the static SSN by a dynamic cryptographic function

As indicated before,[16] smart card manufacturers often include a worldwide and unique serial number in their cards. From a data protection viewpoint, the use of these GUIs is only allowed if it is used for legitimate and particular purposes, e.g., to identify malfunctioning cards or to counter the fraudulent use of cards.[17] There is, however, one side-effect. As indicated before, the GUI can be used for other purposes, notably for cross-profiling purposes.[18] [19]

Smart card manufacturers often include a worldwide and unique serial number in their cards

If, however, the use of an identifier should extend beyond the particular purposes of that application, the unique identifier could be considered an identifier of general application and subject to a more stringent legal regime.[20]

In this regard, an appropriate technical solution may be the following:

- The static unique SSN remains on the card, but is not any longer accessible by applications running on it. Only one person, e.g., the smart card manufacturer, has access to the SSN;
- All applications have access to a random-asymmetric-cryptographic-identification function of the card.

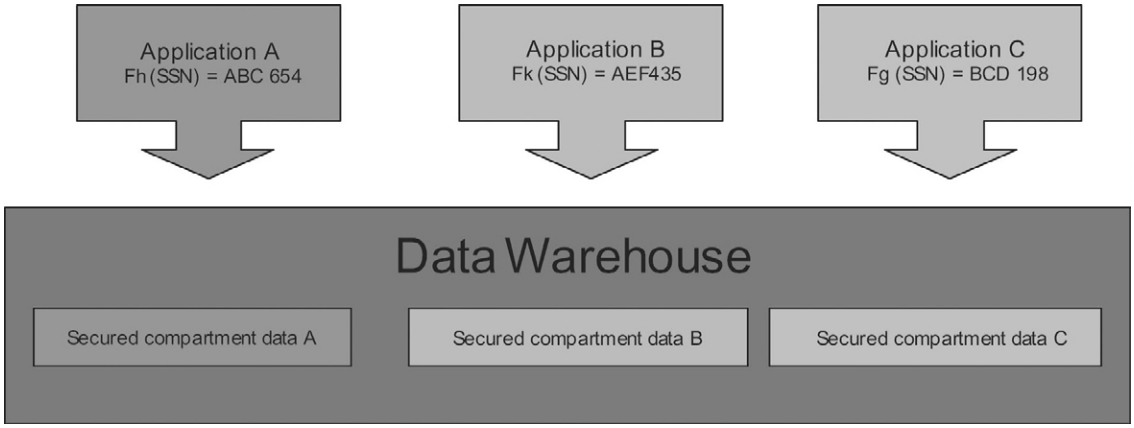
At a first glance, we saw in *Scenario one* that encrypting the SSN with a unique key proper to each application provider may help to guarantee non-abusive cross-profiling. This solution was not completely satisfactory, because the encryption was performed by each application provider and decryption was thus possible. The first enhancement proposed relates to an automatic "on the fly" encryption of the SSN by the smart card. Considering that this first encryption is done with a public key, this key may be securely stored on the smart card.

However, this solution is not the panacea. Of course, the SSN will only be known by the smart card manufacturer holding the corresponding private key, but the encrypted SSN remains a GUI that still may be used by every application to perform cross-profiling activities. For this reason, a second privacy enhancement can be proposed: before encrypting the SSN, random data will added to the SSN,[21] and this couple, *i.e.*, the SSN and the random data, will then be automatically encrypted by the smart card.

By designing such a privacy-enhancing smart card the original purpose of the SSN is kept: the card manufacturer still can identify a card by i) decrypting the SNN with its private key and ii) removing the added random data. At the same time, a cross-profile based on the SSN becomes impossible because the SSN remains secret and the identification function, even if available to each application running on the card, will always return a different ID.

Finally, it is important to make clear that the hardware and software used in the scheme[22] could be built in such a manner that one's identity is only verified, without revealing it. An example of such a less privacy-killing solution can be given by referring to the Processor Serial Number (PSN) of the Intel Pentium III processor.[23] A solution to

Figure 4: PKI encryption of the SSN with random noise



reconcile the advantages of the PSN[24] with the concerns in the field of privacy, notably the use of a GUI, could be to replace the serial number identification instruction by a serial number verification instruction. In the first scenario, an active component is able to read the PSN.[25] In the second scenario, the active component is only able to check if a serial number freely given by the data subject is the exact serial number of the installed processor. The authentication function of the PSN is preserved while the privacy-killing aspect is widely reduced.

Jean-Marc Dinant and Ewout Keuleers, University of Namur (Belgium)

Jean-Marc Dinant has a Master Degree in Computer Science and is currently working on his PhD on Personal Data Security on the Internet. He was CRID's coordinator for the SmartCities Project. He can be reached at jmdinant@fundp.ac.be

Ewout Keuleers is an attorney at the Bar of Brussels (ULYS law Firm) and a researcher at the Centre for Computer and Law (CRID). He can be reached at ewout.keuleers@fundp.ac.be or at ewout.keuleers@ulys.net

This article is derived from CRID's deliverable D12.7 on the legal aspects of the EC funded SmartCities Project. The SmartCities project aimed at and designed a dynamic smart card and multi-application management architecture to allow middle size cities to benefit from the numerous advantages of a smart card environment without being tied to a unique, proprietary applicative model. This IST Project involved Schlumberger Systèmes, Sema UK, Schlumberger Industries, Southampton City Council, University of Southampton, MasterCard Europe SA, Technolution, Crid, City of Goteborg, IT Innovation, and Black Sea Consulting. More information on this project can be found at http://www.smartcities.gov.uk

CRID stands for Centre de Recherche Informatique et Droit (Research Center for Computing and Law) of the University of Namur (Belgium). http://www.crid.be

FOOTNOTES

1 Ewout Keuleers and Jean-Marc Dinant, Data protection: multi-application smart cards: the use of global unique identifiers for cross-profiling purposes –[2003] 19 CLSR 480

2 Cf., article 6,1 (a) of Directive 95/46/EC.

3 Cf., infra point 5.

4 Hereafter referred to as SNN.

5 For efficiency reasons it can be proposed that for performing this encryption, single symmetric encryption functions are used.

6 F is a symmetric cryptographic function, h, k, g, are secret keys belonging to application provider A, B and C. *Figure 1* is displayed in Part I of this paper [2003] 19 CLSR 481

7 Recital 26 of Directive 95/46/EC states that the principles of protection must apply to any information concerning an identified or identifiable person.

"Whereas, to determine whether a person is identifiable, account should be taken of all the means likely reasonably to be used either by the controller or by any other person to identify the said person; whereas the principles of protection shall not apply to data rendered anonymous in such a way that the data subject is no longer identifiable".

8 This may seems strange to non-lawyers. However, in application of the recital 26 stated above, there is no doubt that personal data that are made pseudonymous by a third party remain personal data as long as the reverse identification remains technically feasible.

9 Cf. article 6, 1 (a) en (b) of Directive 95/46/EC, supra.

10 In this regard, reference can be made to article 17 of Directive 95/46/EC. In application of its paragraph 2, the controller must implement appropriate technical and organizational measures to protect personal data against accidental or unlawful destruction or accidental loss, alteration, unauthorized disclosure or access, in particular where the processing involves the transmission of data over a network, and against all other unlawful forms of processing.
A similar obligation is imposed on the data processor, e.g., a data warehouse.

11 Including scrambled Serial Number.

12 A marvellous solution should have been that A encrypts the scrambled SN of B by using his secret key.

13 Article 29 WP, Working Document on on-line authentication services, 29 January 2003, WP68.

14 Eg to a women born on July 1950, taking the bus every Sunday at 13PM and living in the 2nd Street.

15 Article 17 Security of processing: Member States shall provide that the controller must implement appropriate technical and organizational measures to protect personal data against accidental or unlawful destruction or accidental loss, alteration, unauthorized disclosure or access, in particular where the processing involves the transmission of data over a network, and against all other unlawful forms of processing.
Having regard to the state of the art and the cost of their implementation, such measures shall ensure a level of security appropriate to the risks represented by the processing and the nature of the data to be protected.

16 Cf., point 4.

17 E.g., it is not possible to have the same GUI, so identity, twice.

18 In this regard, reference can be made to the opinion of the Belgian Data Protection Authority concerning the use of GUI in relation to e-government and the electronic ID card. Commission de la protection de la vie prive, Avis n°19/2002, 10 juin 2002, p.9-11.

19 A well-known example of the re-definition of the initial purpose is the Social Security Number in the United States. Although the number displayed on a U.S. social security card was initially used for public health purposes, this number is nowadays used for other purposes, in particular for the identification of persons

20 Article 8 (7) of Directive 95/46/EC states that Member States shall determine the conditions under which a national identification number or any other identifier of general application may be processed. As an identifier of

general application could be considered every identifier that is a priori not designated for a specific purpose, or every identifier that was designated for a particular purpose, but is used for other purposes. An eloquent example is the Social Security Number in the United States.

21 By adding random noise to the unique SSN, the cryptographic function will each time generate a different identity.

22 Cf., recital 46 of Directive 2002/58/EC.

23 http://www.bigbrotherinside.org

24 E.g., one that will be able to trace its computer in case of theft.

25 Such as an Active-X control or a Java applet.