Schemes for secure management of digitally produced documents

D. Gligoroski[1], S. J. Knapskog[1] & S. Andova[2]
[1] *Centre for Quantifiable Quality of Service in Communication Systems, Norwegian University of Science and Technology, Norway*
[2] *Department for Telematics, Norwegian University of Science and Technology, Norway*

Abstract

By means of a recently proposed encryption and error correcting system we define schemes for building secure protocols for a number of different scenarios for management of digitally produced documents. The documents are both encrypted and robust against a certain amount of intentional or non-intentional errors. Some of the schemes are based on the property of the system that the introduced redundant information needed for error-correction is not algorithmically predetermined but can arbitrarily be chosen, i.e. it can be given a semantical meaning if necessary or if desired by the user.
Keywords: secure documents, error-correction, cryptography, quasigroup, quasigroup string transformations.

1 Introduction

Building methods that can deal with errors in presence of cryptographic operations is not new. Many of them combine encryption and encoding in various ways, depending for what problems they are intended to solve. For instance, there are several methods developed in biometrics, for use and protection of biometric data, as well as methods for a cryptographic protection of digitally produced documents. For example, Ray and Ellson [1] in 1994 registered a US patent, based on a combination of cryptographic and error-correction techniques; the method solves the problem of credit card authentication and is based on use of biometric data extracted from an image of the card holder. O'Gorman and Rabinovich in several works [2, 3, 4] from 1996-1998 used also a combination of cryptographic


WIT Transactions on Information and Communication Technologies, Vol 36, © 2006 WIT Press
www.witpress.com, ISSN 1743-3517 (on-line)
doi:10.2495/IS060401

algorithms, pattern recognition and error-correction techniques for defining systems for secure authentication. Juels and Wattenberg [5] in 1999 defined a fuzzy commitment scheme that is error tolerant. It is meant for biometric authentication and it is built on linear codes in conjunction with cryptographic primitives. Ruhl, Bern and Goldberg [6] in 2001 defined a scheme for digital signing of scanned paper documents by employing assisted channels for correcting errors produced in the scanning process. Dodis et al. [7] in 2004 and Dodis and Smith [8] in 2005, considered fuzzy extractors that provide error tolerance and extract randomness from biometric data. However, these methods do not have any cryptographic properties and as such have to be combined with cryptographic algorithms. Crescenzo et al. [9] in 2005 defined a model for a variant of Approximate Message Authentication Code which can be applied on biometric data. They also defined notions of approximate correctness and approximate security and argued that the method in [5] does not have these properties. On the other side, the method of Crescenzo et al. has been characterized as insecure by Chang and Li in [10].

In this paper we consider a so-called *Cryptographic Error-Correcting Algorithm* which functions at the same time as an encryption algorithm and as an error-correcting algorithm. We investigate the application of this method for secure management of digitally produced documents. We propose 14 schemes that guarantee security of cryptographically processed data and their robustness against possible errors that may occur. We consider this to be one of the important properties that some cryptographic scheme must posses in order to be widely accepted as a tool for security management.

Some of those schemes can be considered as cryptographic schemes for digital signatures, authentication, non-repudiation etc., while some of them require a further analysis in order to learn their potential and practical application on different real-life problems (such as possession of non-disclosed versions of paper or machine-readable documents that are error-resistant, keeping personal ID or biometric information in non-disclosed but error-resistant form, declaring commitments trough semantical encryption, etc.).

The paper is organized as follows. In Section 2 we discuss the traditional way of combining encryption and encoding functions. In Section 3 we discuss a new approach in which both functions are performed at once by a single function and we point out a recently defined algorithms which has this property. The focus of the paper, the schemes for secure document management, are defined in Section 4 where we also give examples of concrete realizations of two schemes. Section 5 concludes the paper.

2 The traditional encryption-then-coding approach

The traditional approach in which the input message is first encrypted and then encoded before sent via a noisy channel can be described by the diagram in Figure 1. As shown in the figure, after transmission the message is first decoded and then decrypted.

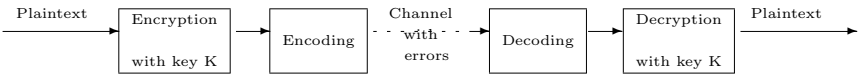

Figure 1: The traditional way of combining encryption and error-correction.

If Q denotes the alphabet over which the messages are built, the model in Figure 1 can be described by means of four functions: $Encr_K, Decr_K : Q^+ \to Q^+$ and $Enco, Deco : Q^+ \to Q^+$ where Q^+ as usual denotes the set of all finite non-empty strings over the alphabet Q. Note that the encryption and decryption functions are considered with respect to a fixed encryption key K. The processing of a message M, which is a subject of some information security procedure, before and after transmission is done in four steps (one step for each function): 1. Encryption: $Encr_K(M) = C_1$, 2. Encoding: $Enco(C_1) = C_2$, 3. Decoding: $Deco(C_2) = C_1$, 4. Decryption: $Decr_K(C_1) = M$

The cryptographic properties of functions $Encr$ and $Decr$ can be stated by the following property:

Property 1. *For every message M it is computationally infeasible to distinguish between $Encr_K(M)$ and a message from uniform random source, for an arbitrary encryption key K.*

The error-correcting properties of functions $Enco$ and $Deco$ are expressed as:

Property 2. *If $Enco(C_1) = C_2$, there exists a positive natural number $d > 0$ such that for every C_2' which is in a d-neighborhood of C_2 (i.e. for which $Hamming(C_2, C_2') \leq d$), $Deco(C_2') = C_1$. By $Hamming(x, y)$ we denote the Hamming distance between two strings x and y.*

3 Cryptographic error-correcting algorithms

We take a different approach to the problem of combining encryption/decryption and encoding/decoding. Namely, we consider a model where encryption and encoding are done by a single mathematical function (cryptographic primitive). Thus, instead of the model described in Figure 1 we propose a model of joint encryption and error-correction, as illustrated in Figure 2.

In this model we need only two functions, $\mathcal{E}_K : Q^N \times Q^m \to Q^{m+r}$ and $\mathcal{D}_K : Q^N \times Q^{m+r} \to Q^m$, where $N \in \mathbb{N}$ is the (fixed) length of the encryption key K, m is the length of the message M, r is the length of the redundant message M_R used for encoding, and $R = \frac{m}{m+r}$ is the rate of the error-correction. $\mathcal{E}_K$ is called an *encryption+encoding function* and $\mathcal{D}_K$ is called a *decryption+decoding function*. Again, we assume a fixed encryption key K.

Now the processing of the message M is done in two steps: 1. Encryption+Encoding: $\mathcal{E}_K(M, M_R) = C$; 2. Decryption+Decoding: $\mathcal{D}_K(C, M_R) = M$, where K is a key (either symmetric or public key) and M_R is some redundant message required for error-correction.

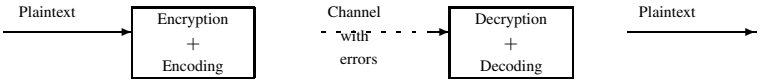

Figure 2: The cryptographic error-correction approach.

A pair of functions $(\mathcal{E}, \mathcal{D})$ that is considered as a *Cryptographic Error-Correcting Primitive* satisfies the following requirements:

1. (Invertibility of $\mathcal{E}$ and $\mathcal{D}$) For every key $K \in Q^N$ and every message $M \in Q^m$, $\mathcal{D}_K(\mathcal{E}_K(M, M_R), M_R) = M$.
2. (Cryptographic properties of $\mathcal{E}$ and $\mathcal{D}$) If the key K of length N is not known to the adversary, then under the adaptive chosen attack, the minimum number of computing operations needed for recovering the message M from the ciphertext $C = \mathcal{E}_K(M)$ is $O(|Q|^N)$ i.e. exponential on the length of the key.
3. (Error-correcting properties of $\mathcal{E}$ and $\mathcal{D}$) There exists a positive natural number $d > 0$, such that for every string C' that is within Hamming distance d from the string $C = \mathcal{E}_K(M, M_R)$ (i.e. $Hamming(C, C') \leq d$) $\mathcal{D}_K(C', M_R) = M$.

Recently an algorithm with the above mentioned properties has been proposed in [11, 12, 13] by Gligoroski et al. It is based on a technique called *quasigroup string transformation*. It uses a quasigroup $(Q, *)$, of order 16, as a parameter when generating the pair of functions $(\mathcal{E}_K^Q, \mathcal{D}_K^Q)$. We stress that this pair is chosen out of at least 2^{430} possibilities.

An advantage of this method is the possibility to choose the redundant information M_R arbitrarily. Note that in [12] the authors chose to use the redundant information which consists only of zeros, i.e. $M_R = 00\ldots0$. However, the method allows the redundant information to have an arbitrarily chosen content. In the next section we exploit this specific property and define various interesting security schemas.

4 Schemes for obtaining encrypted and recoverable documents

In this section we describe several security schemes which can be used in different real-life situations. All the defined schemes use a trusted third party (TTP) denoted by $\mathcal{N}$. In real life it can be a notary, lawyer, certificate authority, bank, a governmental organization, or any other institution. $\mathcal{N}$ guarantees that the content of any document certified by him, is authentic and identical to the originally signed document by clients Alice and Bob.

The general framework from which all schemes will be developed is described in the following way. $\mathcal{N}$ generates (or possesses) a quasigroup $(Q_{\mathcal{N}}, *)$ on the set Q. It can be ether public or secret. $\mathcal{N}$ uses a unique counter $Counter$ (equivalent

to a conventional archive number) for every processed document. The clients Alice and Bob both own keys K_A and K_B, respectively, which are strings of the alphabet Q. These keys can be either secret or public.

The TTP $\mathcal{N}$ combines the parameters $Counter$, $Q_{\mathcal{N}}$, K_A and K_B by applying an one-way operation

$$K = Hash(Q_{\mathcal{N}}, Counter, K_A, K_B),$$

and produces the encryption symmetric key K. In order to perform the error-correction function, the algorithm requires an additional parameter, the redundant information M_R. In opposite to other error-correcting algorithms where the redundant information is pre-determined and is derived from the input message M, here M_R can be any message. Thus it makes sense to talk about M_R being public or being known only to some of the involved parties. In the latter case we assume that this information will be given later to the rest of the involved parties when the verification procedure needs to be done.

We generate various scenarios by changing the initial knowledge of each involved party. Thus we assume that all parameters, $(Q_{\mathcal{N}}, *)$, K_A, K_B and M_R can be either public, secret or can be given publicly in a later phase. We suppose that the variable $Counter$ is always public. We assume that if Alice or Bob keep some information secret, then either they share this secret with the TTP, or they produce a cryptographic hash out of the secret and give this value to the TTP. For each scenario, in few words, we explain the specifics of that particular setting. As one can note, they strongly depend on the values of the parameters. But in general, all schemes rule out possibilities of changing the content of the document and they all guarantee recovering of the original document in case of (a certain degree) of intentional or non-intentional errors.

Quasigroup $(Q_{\mathcal{N}}, *)$ *is public* and

S1: K_A *is public,* K_B *is public,* M_R *is public.* The document is disclosed. Alice and Bob only have guarantees that its content can be recovered in case of intentional or non-intentional errors.

S2: K_A *is public,* K_B *is public,* M_R *is in possession of Alice to be given later to Bob.* Alice and Bob have guarantees that its content cannot be changed and that it can be recovered from intentional or non-intentional errors. M_R contains information, possibly commitments Alice is taking, which is known only to her. Later, when Bob will be given this information he can check the content of the document and check the commitments of Alice.

S3: K_A *is secret,* K_B *is public,* M_R *is public.* Only Alice can prove the authenticity of the document. Any change she may make in the document, will be discovered by Bob and $\mathcal{N}$ together, even if they only have the non-disclosed version of it. Bob cannot change the content of the document.

S4: K_A *is secret,* K_B *is public,* M_R *is in possession of Alice to be given later to Bob.* Only Alice can prove the authenticity of the

document. Any change she may make in the document, will be discovered by Bob and $\mathcal{N}$ together, even if they have only the non-disclosed version. Bob cannot change the content of the document. If Bob is given only the non-disclosed version of the document he will be able to produce the original document when he gets M_R.

S5: *K_A is secret, K_B is public, M_R is in possession of Bob to be given later to Alice.* Only Alice can prove the authenticity of the document. However she does not know the Bob's commitments hidden in M_R. Any change she may make in the document, will be discovered by Bob and $\mathcal{N}$ together, even if they have only the non-disclosed version. Bob cannot change the content of the document. As soon as Alice gets the M_R from Bob she can check his commitment.

S6: *K_A is public, K_B is public, M_R is a concatenation of two secrets M_{R_A} and M_{R_B} of Alice and Bob respectively.* The produced document contains the secret commitments of both, Alice and Bob. At some point of time they can publicly announce their commitments (or pass them to the trusted party).

S7: *K_A is secret, K_B is secret, M_R is public.* The document is non-disclosed. Alice and Bob have guarantees that its content can be recovered from intentional or non-intentional errors. They can publicly (at front of the trusted party) agree upon the semantics of the redundant information M_R. Then they can pass hashed values of their secret keys and the document will be produced.

Quasigroup $(Q_{\mathcal{N}}, *)$ is Secret. In all scenarios that follow the authority $\mathcal{N}$ has a guarantee that the content of the document cannot be changed due to the assumption of secret quasigroup known only to $\mathcal{N}$. In addition, if:

S8: *K_A is public, K_B is public, M_R is public.* All three parties, $\mathcal{N}$, Alice and Bob have guarantees that the content of the document cannot be changed and can be recovered from intentional or non-intentional errors. Since public data of Alice and Bob are used, any authority can produce this document.

S9: *K_A is public, K_B is public, M_R is in possession of Alice, to be given later to Bob.* Bob has guarantees that the document can be recovered from intentional or non-intentional errors. If Alice hides a commitment she is taking in M_R, Bob can learn and check her commitment once he gets M_R from Alice.

S10: *K_A is secret, K_B is public, M_R is public.* Alice can prove the authenticity of the document. Bob has only guarantees that its content can be recovered in case of intentional or non-intentional errors.

S11: *K_A is secret, K_B is public, M_R is in possession of Alice, to be given later to Bob.* Alice can prove the authenticity of the docu-

ment. Bob has guarantees that its content can be recovered from intentional or non-intentional errors. If Bob has disclosed version of the document he can verify the commitments of Alice once he gets M_R from her.

S12: *K_A is secret, K_B is public, M_R is in possession of Bob, to be given later to Alice.* Alice can prove the authenticity of the document without knowing what was Bob's semantical statement hidden in M_R. She cannot change the document. Alice can check Bob's commitments as soon as she gets M_R from him.

S13: *K_A is public, K_B is public, M_R is a concatenation of two secret commitments M_{R_A} and M_{R_B} of Alice and Bob respectively.* The produced document contains the secret commitments of both, Alice and Bob. Since $\mathcal{N}$ is the only one who participates in producing the document with his secret, the authentication of the document can be done only by $\mathcal{N}$. At some point Alice and Bob can publicly announce her/his part of M_R (or pass them to the trusted party) and check whether the other has met his/her commitments.

S14: *K_A is secret, K_B is secret, M_R is public.* If the document is stored in a non-disclosed version then Alice and Bob have guarantees that its content is recoverable from intentional or non-intentional errors. They can publicly (at front of the trusted party) agree on the content of M_R.

Note that in every scenario except in scenarios S1, S2 and S8, the document can be stored in its disclosed (original) or in its non-disclosed (encrypted) version. Alice and Bob together decide on this question.

4.1 Some concrete realizations

In this section we show two practical realizations of the scenarios S8 and S4. One may note that these scenarios can be modified into two-party scenarios, by making all data of one of the parties public. Thus, since in scenario S8 all data of Bob are public, it can be modified into scenario S8a in which the quasigroup $(Q_{\mathcal{N}}, *)$ is secret, K_A is public and M_R is public. Now, both Alice and the authority $\mathcal{N}$ have guarantees that the content of the document can be stored and can be recovered in case of errors.

In the first example the two involved parties are, Alice, who purchases a yearly parking ticket for a parking place in the downtown, and the authority $\mathcal{N}$ who issues the ticket. If the ticket is just a paper document, there is a trivial possibility for Alice to make a forgery. Namely, she can print another ticket with similar shape but with different (extended) date. The authorities can prevent the forgery by use of a communication network. A parking police officer can check the validity of every issued ticket by contacting the central database and comparing the the data in the database by those on the ticket. Of course, this solution is expensive in terms of communication costs, speed and time.

The scheme S8a gives the following solution. The authority $\mathcal{N}$ possesses a parking ticket key $Q_{\mathcal{N}}$ used for issuing parking tickets. Alice gets a ticket in which the information about its expiration date, together with the information about the licence plates of the car are printed as a paper document. This information appears in both forms, disclosed (that Alice is able to read) and encrypted (that a parking police officer can read and verify the information on the ticket). The encrypted information can be printed in a data-dense format such as a 2D barcode, for instance PDF417 [14]. In order to be able to read the encrypted data, the parking police officer needs a 2D barcode reader. He only needs to read the barcode (encrypted data) and compare the disclosed (original) information on the ticket with the information on the display of his barcode reader. There is no need to communicate the central database in order to check the validity of the ticket.

Obviously, in this scenario the method we discussed earlier is used for three purposes: 1. as an encryption algorithm, 2. as an error-correcting algorithm, 3. as a method for digital signing of paper documents. In the literature similar solutions can be found but they all use separate algorithms for all three steps. As our method can perform all three functions in a single step, we argue that it offers much better solution.

In the next example we use scenario S4. Suppose that Alice wants to write a testament to declare she passes all her fortune to her son Bob after her death. However, taking precautions, she does not want Bob to know the content of the testament at that moment. Thus she rather keeps a copy of the document in her safe, but only in an encrypted from. As such, nobody will be able to reproduce the document or even parts of the document since she is the only one who knows the secret information (and possibly the TTP). Moreover, some intentional or non-intentional changes of the encrypted document can be tolerated and the proper decryption can be done. Alice can produce a letter in which she takes some commitments by which the redundant information M_R is produced. However, the content of her testament cannot be discovered.

In the literature one can find a solution of this problem based on commitment schemes. Those solutions requires additional encryption algorithm, and possibly encoding algorithm. Again, our method provides a solution which uses only one single algorithm.

5 Conclusion

In this paper we used a newly developed cryptographic error-correcting algorithm to define 14 schemes which can be used for secure management of digitally produced documents. The concept of semantical encryption that can be applied by this new technique is very powerful and allows us to define solutions for some real-life situations which cannot be captured by any other method. In the paper the proposed schemas have been given informally. Currently, we are working on their formal definitions as well as on developing concrete security protocols based on them.

Acknowledgement

This work was carried out during the tenure of an ERCIM fellowships of Suzana Andova visiting Department for Telematics at Norwegian University of Science and Technology, Trondheim, Norway.

References

[1] Ray, L.A. & Ellson, R.N., Method and Apparatus for Credit Card Verification. *U.S. Patent 5321751*, June 1994.

[2] O'Gorman, L. & Rabinovich, I., Photo-image authentication by pattern recognition and cryptography. *Proc. of the* 13^{th} *Int. Conf. on Pattern Recognition (ICPR '96),* IEEE Computer Society: Washington, D.C., Vol. 3., pp. 949–953., 1996.

[3] O'Gorman, L. & Rabinovich, I., Photo-ID encryption and pattern recognition for counterfeit resistance. *CardTech/SecurTech '96 Conference*, Atlanta, pp. 253–261, 1996.

[4] O'Gorman, L. & Rabinovich, I., Secure Identification Documents Via Pattern Recognition and Public-Key Cryptography. *IEEE Trans. Pattern Anal. Mach. Intell.*, **20(10)**, pp. 1097–1102, 1998.

[5] Juels, A. & Wattenberg, M., A Fuzzy Commitment Scheme. *Proc. of the* 6^{th} *ACM Conference on Computer and Communications Security*, ed. G. Tsudik, ACM Press, pp 28–36, 1999.

[6] Ruhl, M., Bern, M. & Goldberg, D., Secure notarization of paper text documents. *Proc. of the* 12^{th} *Annual ACM-SIAM Symposium on Discrete Algorithms*, Washington, D.C., ACM Press, pp. 437-438, 2001.

[7] Dodis, Y., Reyzin, L. & Smith, A., Fuzzy Extractors: How to Generate Strong Keys from Biometrics and Other Noisy Data. *Proc. of International Conference on the Theory and Applications of Cryptographic Techniques, Eurocrypt 2004,* eds. C. Cachin & J. Camenisch, Springer-Verlag, LNCS 3027, pp. 523–540, 2004.

[8] Dodis, Y. & Smith, A., Correcting errors without leaking partial information. *Proc. of the* 37^{th} *ACM symposium on Theory of computing*, ACM Press, New York, pp. 654–663, 2005.

[9] Crescenzo, Di G., Graveman, R., Ge, R. & Arce, G., Approximate Message Authentication and Biometric Entity Authentication. *Proc. of the* 9^{th} *Int. Conf. Financial Cryptography and Data Security, FC2005,* eds. A.S. Patrick & M. Yung, Springer-Verlag, LNCS 3570, pp. 240–254, 2005.

[10] Chang,E.C & Li, Q., Small Secure Sketch for Point-Set Difference. Cryptology ePrint Archive, Report 2005/145, 2005. http://eprint.iacr.org/.

[11] Gligoroski, D., Markovski, S. & Kocarev, L., New Directions in Coding: From Statistical Physics to Quasigroup String Transformations. *Proc. of the International Symposium on Nonlinear Theory and its Applications, NOLTA2004*, Fukuoka, Japan, 2004.

[12] Gligoroski, D., Markovski, S. & Kocarev, L., Error-Correcting Codes Based on Quasigroups. Submitted to *IEEE Information Theory*, October 2005.
[13] Gligoroski, D., Markovski, S. & Kocarev, L., Cryptographic Primitives, Error Coding, and Pseudo-Random Number Improvement Methods Using Quasigroups. U.S. Provisional Patent Application Serial No. 60/618,173, filed October 13, 2004, under 35 U.S.C. §119. (now disclosed), 2004.
[14] WIKIPEDIA, http://en.wikipedia.org/wiki/PDF417.