Biometry, the safe key

María Fraile Hurtado, Miguel Herrero Langreo, Pilar Menéndez de Miguel ,Valerio Delgado Villanueva

College of Engineering and Architecture, Pontifical University of Salamanca

Abstract: — Biometry is the next step in authentication, why do not we take this stepforward in our communication security systems? Keys are the main disadvantage in the cryptography, what if we were our own key?

Key words: — biometry, cryptography, iris, IrisCode, key, safety, security.

I. Introduction

Cryptography has been used since the beginning of the classic times. Since the age of the great conquerors there has been a need for the capacity to communicate between commanding staff and ground troops avoiding the enemy compromising those orders.

In present times safe data transfers have become a primary issue way above the military field, reaching companies and families on a daily basis. New technologies have delivered us a wider number of communication channels, along with these technologies the need for new ways to protect data flow from hacking have emerged.

That's the reason because criptopgraphy has evolved from the Greek scitala used by Spartans during their campaigns to communicate, throughout the Ceasar's code transposition attributed to the Roman emperor and up to the famous Enigma machine, used by the German army during World War II.

Criptographic systems that have been developed through history are based on mathematical algorithms, over time they have become increasingly complex and, therefore, safer. Evolving from plain alphabetical transpositions, to the current systems of vast complexity, where the use of a key linked to the text made it possible for a safe data transfer, even if the algorithm is compromised.

Mathematics and cryptographic sciences have made it possible for ciphering systems to be almost unbreakable. Curiously, as cryptology grew in complexity, so did cryptoanalysis, the science responsible for breaking encrypting systems. The continuous struggle among these contenders has taken these sciences to such a level of complexity that they've become matters of study all by themselves.

As an example of how mathematics has advanced the ciphering systems, we can look into RSA cipher algorithm key generation process:

1. Choose two distinct prime numbers with 200 digits, p and q.
2. Computed $n = p \cdot q$.
3. Compute $\varphi(n) = (p-1)(q-1)$, φ is Euler's totient function .
4. Choose an integer e such $1 < e < \varphi(pq)$ and $\varphi(pq)$ are coprime.
5. Determine d (using modular arithmetic) wich satisfies the congruence relation.
   $de \equiv 1 \pmod{\varphi(n)}$.
   Finally the public key will be (n, e) and the private key will be (n, d).

II. State of the art

At present there are two mainstream approaches regarding coding systems, symmetric encryption and asymmetric encryption. Both systems have sending information through safe or unsafe channels, while still being illegible by any third party, as its main purpose.

Symmetric encryption has evolved from encryption systems that only used an algorithm to cipher data. A key was added to the use of the algorithms in order to cipher and decipher. This has become standard procedure; therefore the Scientific Community is grateful for their continuous effort and improvement of this technology. Nowadays the security of encryption systems relies on the key length as well as on the algorithm in use. This is where these systems have a certain handicap, increasing the key length has an increase in security, but it implies a burden on the user, because long keys are hard to remember and use on a daily basis.

Asymmetic encryption was born as an evolution of symmetric encryption systems. In these systems, the users have a pair of keys, one of them as public key and the other as private. Each key only could decipher a text that has been cipher with the other. These encryption cipher system allowed the digital sign system's born. The disadvantage of these encryption systems is the time used in the cipher and deciphers process, because the use of complementary key forces the algorithm to do too many operations. As result the algorithms use a sort keys to reduce this time, allowing so a cryptanalysis attack.

Despite such progress in the encryption field, we are still far from the point where the algorithmic encryption is invulnerable. Just as important as the key length is the care jealousy with which the key is managed and stored; this implies that the use of very long keys is not as safe as it could be expected, as the average user tends to write down long keys in order to avoid remembering them, a terrible security flaw regarding the fact of the existence of social engineering.

Logic says us to use bigger key, because short keys are subject to brute force attacks (the strength of these attacks depends on

the capacity of the computers from which they are launched). The bigger computer's capacity gets longer keys to be used, in order to be safe from these attacks.

As said before, storage is one of the main issues when dealing with security; in fact, current studies talk about a three leveled based security.

When we talk about security, in the cryptography field, we are trying to find how to achieve the three levels of security. Those three levels are:

a) Something that you know (password),
b) Something that you have (like an ID card or a security token),
c) Something that you are (biometry).

There are already a lot of well developed products regarding step b, however current development is focused on step c.

The main objective is to fully identify the user beyond reasonable doubt. New technologies offer the possibility to do this through biometric recognition. These systems are based on the mathematical study of some unique biological features.

The characteristics that all biometric features must meet are:

- **Universality:** any and all human beings must have it. Everyone, by the fact to be people, must have it.
- **Uniqueness:** it must be unique for every individual.
- **Continuance:** the biometric characteristic from the individual must stay unchanged over time or, at least, for a limited timestamp.
- **Perennial:** unaltered by external factors (like weather alters the outer appearance of hair).
- **Measurable:** it must be posible to quantify it.

In the biometric field, there are two main groups in biometry studies, depending on the feature under analysis: static biometric systems and dynamic biometric systems. The first are focused on physical features as fingerprints, iris, retina, hand geometry or face geometry. The second are focused on the behavioral features, like written signature, voice, gesture and body language.

From safety point of view biometry is a part of the security third level. With biometry we can take a step forward, because before the beginning of biometry we only could indetify people: Does this password, ID card or token belong to this person? Now with biometry we can raise the following question: Is this person really who says to be? This supposes a great increase in the safety.

The following table shows in comparative manner different biometric systems, their features, their acceptances and possible interferences in order to have a wider, more accurate, view on them.

	Voice	Writing	Fingerprint	Retina	Iris
Reliability	High	High	High	Very high	Very high
Usability	High	High	High	Low	High
Attack prevention	Medium	Medium	High	Very high	High
Acceptance	High	Very high	Medium	Medium	High
Stability	Medium	Medium	High	High	High
Identification	No	Yes	Yes	Yes	Yes
Authentication	Yes	Yes	Yes	Yes	Yes
Interference	Noise, cold	Easy signature	Injuries, dirt	Special cases	Special cases

Table 3 Comparative of biometry systems

In order to fully understand the advantages that iris recognition system has to offer, we proceed to explain the previous table, and its stated characteristics, in detail.

Reliability: Human iris acquires its signature pattern during prenatal development, being unaltered throughout its whole life, becoming an unerring identity print. Even more, iris's digitalization is a concise and accurate process, implying therefore optimum rates of false positives and false negatives.

User friendly: Due to the minimum requirements of the system, a superficial eye image, they are neither intrusive nor hurtful for the user.

Attack prevention: As with retina recognition systems, in the iris recognition systems the hardware is directly controlled, so this biometric pattern becomes difficult and complex to misappropriate, due to the fact that obtaining a mock iris to place before the system requires substituting the original one from the forger, what's more, this forgery must be an exact copy of the original from the supplanted owner. These techniques are, although possible, not cost effective by any chance, because they require an extensive medical and technical knowledge.

Acceptance: Thanks to the systems simplicity and non-intrusiveness, clients forced to use it are fully satisfied. This is one of the main reasons of the comparative advantage of the iris recognition systems in the biometric verification market.

Stability: From its very start, the iris recognition systems algorithms have had a great acceptance in the market, ensuring its standard use. Besides, the existence of numerous successful iris recognition systems guarantees a bright future for these techniques and technologies.

Identification: Iris is a univocal biometric pattern that doesn't require any specific action taken by the user, whose pattern is spotted and captured with great precision, therefore granting a positive identification of the system's user.

Authentication: Those systems capable of identifying a user, which also have information regarding the user's biometric patterns, will be able to authenticate.

Interferences: Being this a method that requires superficial pictures to be taken, there are several factors that could cause "noise" within the data needed to make a positive identification. Due to the great progress made concerning the filtering of such mentioned noise, these systems embody the biometric recognition option which would be the least affected by these external factors.

It's because of this that the use of iris biometric recognition systems offer a great benefit-usability rating. To fully understand the advantages of the iris recognition system, it is imperative to study deeper into the process. This process is integrated by the following stages:

1. **Segmentation**: the relevant information on the captured image is located (that which belongs to the iris), it is targeted and cropped, disregarding the excess of unrelated data. This stage is extremely important, for if the information isn't correctly located, the next stages won't be using the appropriate data.

Figure 2 Segmentation process

2. **Normalization:** this stage is where, basically, an image transformation takes place. Regardless of the size or bearing of the input image, we`ll obtain a rectangular image of fixed dimensions, according to the polar coordinates of the iris's original ring shape.

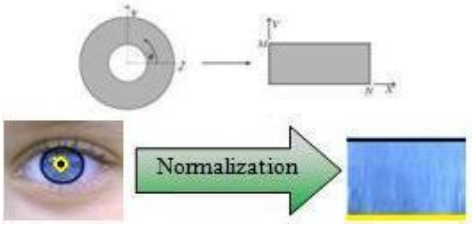

Figure 3 Normalization process

3. **Codification**: each pattern is processed in order to retrieve it`s data, the iris pattern is coded in a binary file (2048 bits) each iris pattern is treated to extract its information. The detailed patterns are codified/crypt in a 256 bytes (2048 bits) binary array, which represents all the texture's detail. It's what is known as IrisCode.

Figure 4 Codification process

One of the systems special features is that two IrisCode from the same user will hardly be identical. This happens due to the different external conditions that take place when the pictures are taken, not only that, also the eye angle may vary; all of this translates in a shift of position of the bits of the IrisCode. However, the unique patterns that identify the user are still located and checked univocally.

III. Our Research

The traditional key system, where the key storage is the user's responsibility, allows for these keys to be used for different objectives with only its length as a restriction, as most systems only take keys that have a certain length.

Biometric systems just as they are at the moment only allow authentication. The reason behind biometric systems no being used on modern cipher algorithms is its bits coding from the biometric pattern, like the one from recognizing the iris, which is never exactly the same, even there are similar IrisCodes, it's through the calculus of the distance between patterns how you can determine the property of the binary array to the user or not.

If we take the basic sketch of a symmetric encryption system its the one that appears in the figure and is based on applying a same key on a text, there may or not be different procedures to cipher and decipher, you will get the original text.

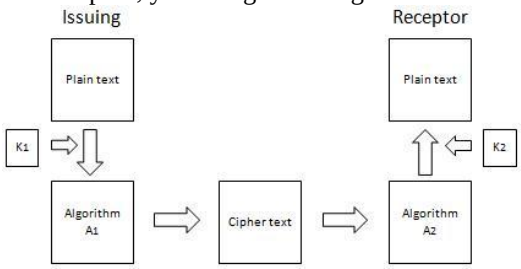

Figure 5 Basic sketch of the encryption

$$T_{Issuing} = T_{Receptor} \Leftrightarrow (K_1 = K_2 \wedge A_1 = A_2)$$

As explained, both, the algorithm and the key, have to be exactly the same, in order to use a biometric system for the encryption of the content we find ourselves against a rather simple problem.

$$(K_1 \cong K_2 \wedge A_1 = A_2) \Rightarrow T_{Issuing} \neq T_{Receptor}$$

As explained, both, the algorithm and the key, have to be exactly the same, in order to use a biometric system for the encryption of the content we find ourselves against a rather simple problem.

So the keys are practically identical but not exactly the same, therefore the recovered will never be the same one that was originally sent. Not a functional system.

The answer obviously must find a way for the procedure to receive biometric keys and still return a single key, as long as the biometric keys belong to the same subject; which means we're looking for a transformation.

$$f(K_n) = K_0 \; \forall \; K_n \in S$$

This way the system would be configured as depicted on the image.

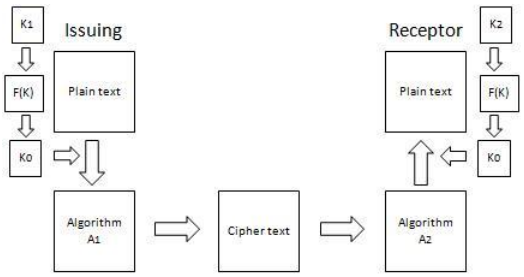

Figure 6 Expanded sketch of the encryption

It is thanks to this system that the problem is solved, as the encoding algorithm always gets the same key, as long as the function $f(K)$ receives the data from the same subject.

$$T_{Issuing} = T_{Receptor} \Leftrightarrow$$

$$\Leftrightarrow (f(K_1) = f(K_2) \wedge A_1 = A_2)$$

There's no need for an extensive knowledge on the authentication process through biometric values to figure out that the equalizer function's $f(k)$ process will be different depending on the pattern used, due to the inherent differences and the uncorrelated data it's impossible to implement a $f(k)$ that will allow us to extract an identical key from a person starting with fingerprints from different fingers, the comparison between the eye's iris or even the iris and the fingerprint; this is simply impossible, for starters the biometric element must be the same, and the treatment in order to obtain the single key anytime the biometric pattern is used will the extremely related with the special features of each parameter.

It's also trivial the fact that the k_0 key acquired by these functions $f(k)$ will always contain less information than the originally obtained key, for it's a certain kind of hash algorithm that process the entry to produce a common exit for a certain range of entries. The data that differ each of the keys k_n will get lost on having produced the key k_0.

Biometric patterns offer the advantage of having such an amount of information data that it allows us to reduce it in order to create a unique key per user, and still have a key length that makes attacks unfeasible through brute force.

In our case, we're going to concentrate on iris recognition systems due to their several advantages.

Keys generated from iris recognition are made from the patterns present on the user's iris, and the differences between IrisCodes captured from the same user consist on small pattern displacements. This is due to the inclination changes, lens proximity, environmental luminosity, etc.

So, even though the stripe patterns are basically the same, their position may vary and due to environmental elements its size can suffer small variations, but still we can identify, with enough precision, the number of patterns present and their dimension.

The suggested approach to create the mentioned transformation $f(K_n) = K_0 \; \forall \; K_n \in S$ is to make the most the stripe patterns present on the IrisCodes. In other words, as we can consider the number of patterns (N_p) as a constant from the same individual but a variable for different individuals, it's easy to set out a first equation f_1 capable of getting back of IrisCode key patterns as such:

$$f_1(K_n) = N_0 \; \forall \; K_n \in S_1$$
$$\wedge$$
$$\exists \; S_2 \mid f_1(K_n) \neq N_0 \; \forall \; K_n \in S_2$$

Likewise the dimensions of every pattern will be a very stable element, though not exactly the same one, always they will be in the same range for the same person, so if we do not measure the exact dimensions of a pattern, but his belonging or not to a certain range of size, we will obtain a new perfectly stable measure inside the same person. It is perfectly considerable $f_2(K_n, \mathrm{X})$ that such a pattern X in the IrisCode K_n is capable of returning always the same value.

Once explained these concepts it is necessary to do a consideration in the matter: the smaller the spectrum of ranges in use is, more possible it is that they give similar values for different individuals, whereas the bigger that the spectrum is, the least the loss of information will be, and therefore the capacity of distinction will increase.

In any case, the $\mathbf{f_2}$ equation shape shall be as follows:

$$f_2(K_n, X) = Q \; \forall \; K_n \in S_1 \wedge \mathrm{X} \in \mathbb{R}$$
$$\wedge$$
$$\exists \; S_2 \mid f_2(K_n, \mathrm{X}) \neq \mathrm{Q} \; \forall \; K_n \in \mathrm{S}_2 \wedge \mathrm{X} \in \mathbb{R}$$

Anyway, these obtained values will be treated according to the value of the patters of the bits to which they correspond, so that the values obtained by f will make the results more consistent, and therefore the whole process. We will call then $g(X)$ the function that will treat the values obtained of f to identify them according to the pattern to which they correspond.

Therefore realizing a linear combination of f_1 and f_2 we will obtain the already mentioned $F(\mathrm{K})$ one that will allow us to obtain always the same key for the same individual though we part from slightly different initial keys.

$$F(K) = [f_1(K_n) * g(X)] + \begin{cases} \sum_{x=0}^{f_1(K_n)} f_2(K_n, X)\, 10^{x+1} * g(X) & if \; f(k_n, x)/10 < 1 \\ \sum_{x=0}^{f_1(K_n)} f_2(K_n, X)\, 10^{x+2} * g(X) & if \; f(k_n, x)/10 \geq 1 \end{cases}$$

To better understand this process, we will explain it, with an image of real IrisCode, the functioning of the described algorithm.

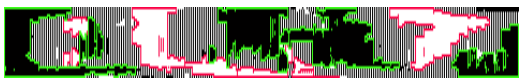

Figure 7 Image with the patterns marked from an IrisCode

In the present image an IrisCode can be observed, to which an algorithm has been applied to highlight those zones, that is to say patterns that we want to highlight to apply the mentioned algorithm. These patterns will have to be treated according to the spectrum of ranges raised, in order to unify these values so that the small variations of atmospheric noise do not alter the result. Let's bear in mind that the obtained value will be used as key in the communications, and it is necessary that this value is unique, both to the emitter as to the receiver.

In the previous IrisCode it is possible to observe the existence of patterns that due to their dimension are insubstantial, since they do not contribute relevant information to the key creation algorithm, since this one can lead to possible mistakes at the moment of the patterns process due to the fact that they can belong to small sheens or alterations of the original image (the picture of the user).

As an example, simplifying the procedure, we can suppose that from an IrisCode seven patterns have been obtained and each of them is classified according to its size, on having these treated with the type associated with the pattern, a series of values will be obtained, its final operation will give, as a result, the key with the one to work with the encrypting algorithm. The spectrum of sizes has been simplified by the following values: very small, small, medium, big and very big; in the same way, to illustrate better the example instead of speaking about zeros or some as pattern values of one, we tag them as white or black, since they are the colors that they appreciate in an image of IrisCode.

Following there are two tables to show how, for the same sequence patternsand their associated characteristics, in spite of the fact there are displacement due to the obtaining of the original image, it will produce anyway the same key value.

Pattern	Size	Type	Vlue
1	Medium	White	150
2	Small	White	100
3	Medium	Black	300
4	Big	White	200
5	Very Large	Black	500
6	Very Small	Black	100
7	Medium	White	150
		clave:	1500

Table 2 Normal pattern

Pattern	Size	Type	Vlue
6	Very Small	Black	100
7	Medium	Black	150
1	Medium	Black	150
2	Small	White	100
3	Medium	Black	300
4	Big	White	200
5	Very Large	Black	500
		clave:	1500

Table 3 Shifted pattern

IV. Bibliography - References

[1] White paper about mathematics in cryptography by Professor Javier Lobillo Borrero from Granada University of Spain. http://www.ugr.es/~matematicas_ugr/mathcrypto.pdf

[2] White paper about cryptography by Professor Manuel Pons Martorell from Mataró Polytechnic University of Spain. http://www.tierradelazaro.com/public/libros/cripto.pdf

[3] Study about security key management by the Segu-Info web page. http://www.segu-info.com.ar/proteccion/clavesseguras.htm

[4] John Daugman's web page. Jhon Daugman is who developed the iris recognition algorithm. http://www.cl.cam.ac.uk/~jgd1000/

[5] White paper about optical iris biometry. http://www.upc.edu.pe/html/0/0/carreras/ing-electronica/proyectos/Biometr%C3%ADa-%C3%B3ptica-de-iris.pdf