Towards Fraud-Proof ID documents using multiple data hiding technologies and biometrics

Justin Picard[a], Claus Vielhauer[b] and Niels Thorwirth[a]

[a]Mediasec Technologies, 6-8 Berliner Platz, Essen, Germany
[b]Otto-von-Guericke University, Universitaetsplatz 2, Magdeburg, Germany

ABSTRACT

Identity documents, such as ID cards, passports, and driver's licenses, contain textual information, a portrait of the legitimate holder, and eventually some other biometric characteristics such as a fingerprint or handwritten signature. As prices for digital imaging technologies fall, making them more widely available, we have seen an exponential increase in the ease and the number of counterfeiters that can effectively forge documents. Today, with only limited knowledge of technology and a small amount of money, a counterfeiter can effortlessly replace a photo or modify identity information on a legitimate document to the extent that it is very difficult to differentiate from the original.

This paper proposes a virtually fraud-proof ID document based on a combination of three different data hiding technologies: digital watermarking, 2-D bar codes, and Copy Detection Pattern, plus additional biometric protection. As will be shown, that combination of data hiding technologies protects the document against any forgery, in principle without any requirement for other security features. To prevent a genuine document to be used by an illegitimate user,biometric information is also covertly stored in the ID document, to be used for identification at the detector.

Keywords: ID cards, biometrics, watermarking, copy-detection pattern

1. INTRODUCTION

Identity documents, such as ID cards, passports, and driver's licenses, contain textual information, a portrait of the legitimate holder, and eventually some other biometric characteristics (fingerprint, handwritten signature). As prices for digital imaging technologies (software, printers and digital cameras) fall, making them more widely available, we have seen an exponential increase in the ease with which counterfeiters can effectively forge documents. Today, with only limited knowledge of technology and a small amount of money, a counterfeiter can effortlessly replace a photo or modify identity information on a legitimate document to the extent that it is very difficult to differentiate from the original.

This paper proposes a virtually fraud-proof ID document based on a combination of three different data hiding technologies: digital watermarking, 2-D bar codes, and Copy Detection Patterns (see attachment: ID card containing three different data hiding technologies), plus additional biometric protection. As will be shown, that combination of data hiding technologies protects the document against any forgery, in principle without any requirement for other security features. To prevent a genuine document to be used by an illegitimate user, in this paper we consider the additional use of biometric information, to enhance the security by several more degrees. The biometrics data are also covertly stored in the document, and as will be discussed, these data can be linked in various ways to the other security features.

The next section introduces the model of the ID document and its intended purpose. Section 3 will present the individual security components of the proposed ID card security system, and discusses their intended use and limitations. Section 4 describes the proposed system, which results from a compromise between technical and logistical feasibility and security requirements. Section 5 addresses different attack scenarios, and demonstrates the security offered by the ID card system against each of these attacks. In a real-world application, the error rate of the system will never be null: Section 6 analyzes the potential failure modes of the system, their consequences, and proposes ways to limit the impact of these failures. The last section will conclude on the proposed ID card security system.

Further author information: send correspondence to jpicard@mediasec.com and Claus.Vielhauer@iti.cs.uni-magdeburg.de

2. MODEL OF THE ID DOCUMENT

Figures 1 shows an abstract view of the ID document with its security components, and Figure 2 shows one possible example of such an ID document. It is assumed that the ID document always exists as a digital image, before print and for the rest of this paper, we assume that there is no external security feature, where "external" stands for any security feature that: (1) is not contained in the digital image of the ID document, or (2) requires a non-standard application process. For example, a hologram, laminate, OVD (optical variable device) require an additional process after the printing of the ID document, a special ink (e.g. UV ink) requires printing an additional layer of ink, and microprinting requires special, costly equipment, and can only be deployed for large-scale applications or for a high security environment that would justify the costs.

The kind of ID document we are referring to could in principle be printed on a regular sheet of paper with a standard inkjet printer, and still be -as we will see- extremely difficult to forge. Considering an abstract model of an ID document allows us to apply the principles of digital security to analog documents. This way, it is possible to measure more precisely the security of a given system, and to compare the security of two different system designs. Proving the security of a system might be a harder goal, though it is not out of reach.

Let us note that external security features are not to be rejected but can be used for additional security. However, this additional security comes at a cost, and most applications have cost constraints. Moreover, having to deal only with digital images brings the required flexibility to issue a document in a decentralized and timely manner

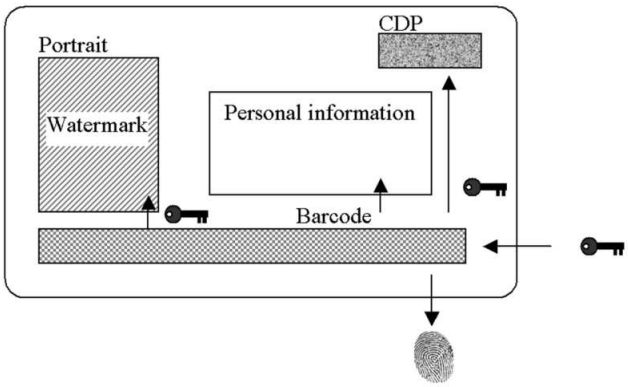

Figure 1. Abstract view of the ID document with the security features.

As can be expected, the two essential components of the ID document are the personal information and the portrait. These components allow a fast, human verification of the person's identity, however as they stand they can be easily forged or replaced. A non-encrypted biometric feature such as the handwritten signature or a fingerprint can also be found. As can be seen on Figure 2, the personal information is often found in an OCR-readable font, allowing for machine-verification of the ID document.

In addition to these basic components, in this paper we consider the use of four other security components: a digital watermark in the portrait, a high-capacity 2D barcode, a Copy-Detection Pattern, and encrypted biometric data. The next two sections presents each of these security components.

A fundamental property of the security components is that they are all encrypted, using the advantage of digital security mechanism for all parts of the verification. This makes reasonable security estimates of the system possible, without concern about restricted access to document components like ink and paper.

3. SECURITY COMPONENTS

3.1. Digital watermark

A secret key based digital watermark is embedded in the digital image of the ID document, or only in the portrait. The watermark, which can be spread all over the document or just embedded in specific areas such as

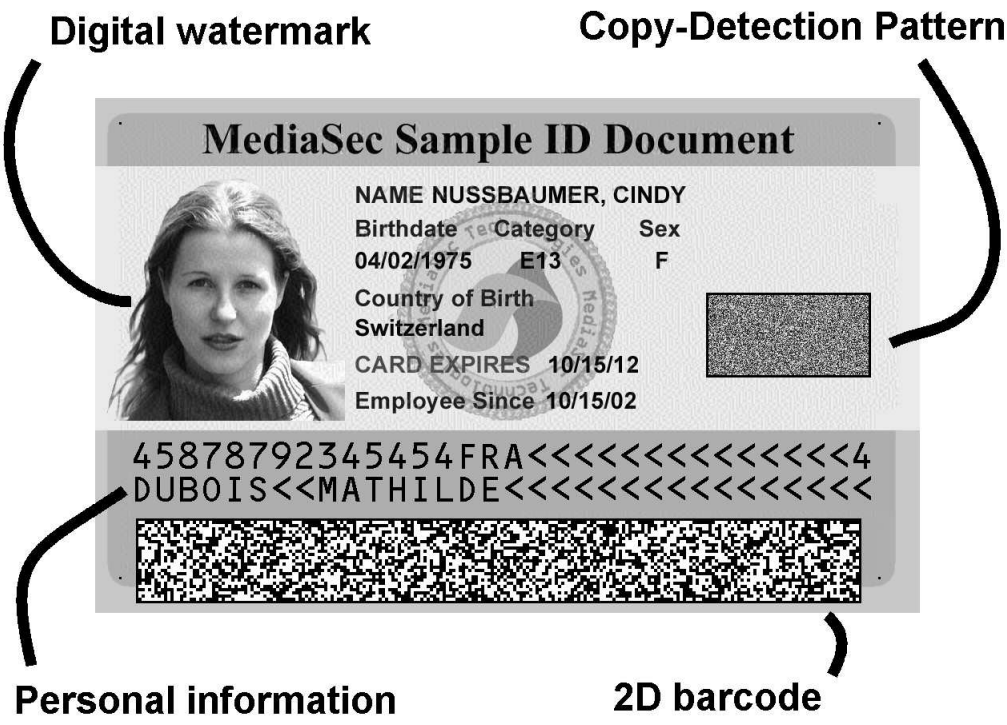

Figure 2. Possible example of the ID document.

the portrait, can later on be extracted from the physical ID, by scanning it and extracting the message with a watermark detector equipped with the decoding key (some specific processing of the document is required to cope with the geometrical transformation inherent to the print-scan process). The watermark on an ID document can serve three purposes: data hiding, authentication of the portrait, and cross-verification with other information. But, because the digital analog transformation strongly damages the watermark, the data hiding capacity for reliable detection is quite low -in the order of 10 to 20 bytes-, and a much higher capacity can be achieved with the 2-D barcode. However, the watermark is essential to protect the ID portrait against forgeries or: alterations to the portrait can be detected by locating the areas where the watermark is not present or significantly damaged.

3.2. 2-D barcode

A 2-D barcode can be used to store several hundred bytes of payload, and typically has a payload per surface of 100 times that of a printed digital watermark. That payload is encrypted, and is used to contain a copy of the personal information or its most significant parts -birth date, ID number-. The biometric data, which in case of a compressed dynamic signature based on a 2-dimensional pen position signal, can be in the order of 1kBytes, are also stored in an encrypted way in the barcode.

3.3. Copy-Detection Pattern

The digital watermark and the 2-D barcode do not offer protection against copies of the ID document (done by scanning and reprinting, photocopying, etc.). Though the energy level of the digital watermark can be affected by the additional digital-analog process required to produce a copy, this decrease of the energy level is often smaller than the natural variation of the energy level of the digital watermark in an original ID document.

To tackle this problem, we use a different technology called the Copy Detection Pattern (CDP). The CDP is a highly textured digital image that is inserted into the digital image of a document to be printed, and that is designed to be maximally sensitive to digital-analog transformations. Automatic analysis of the energy level of a scanned CDP can tell if the document is an original or a copy. The creation and detection of a CDP uses a secret key that can be made dependent on the information in the document, making the CDP unique to the ID document. The reader is referred to[6] for more information.

4. BIOMETRIC DATA

4.1. Capacity Considerations for Biometric Data

Using the three data hiding technologies above mentioned, the ID document is, as we will see, in principle secure against any forgery or illicit copy. However, there remains the possibility that the document is used by an illegitimate holder. Illegitimate use of a valid ID document can be restricted by the use of biometrics, for the automatic authentication of individuals. Indeed, biometric authentication technologies have reached a certain degree of maturity and a dramatic increase in the number of applications recently. A number of the behavioral or physical measures, which are taken into account for biometric authentication process, have already been used on ID documents for a long time. Examples for such biometric characteristics are the handwritten signature or fingerprint images, which are part of driver's licenses or passports in some countries. According to one possible classification of biometric systems, these biometric data can be referred to as overt biometric properties of the document holder. Additionally, an encrypted 2-D barcode allows for the inclusion of covert biometrics in ID documents for an automated ID holder authentication. In a number of applications, it can be desirable to authenticate the ID document and the holder in presence and by approval of the individual only. Cooperative biometric verification techniques such as signature verification include an explicit expression of intension and can be used for this purpose. Biometric authentication processes are based on the comparison of some stored reference with an actually presented trait. Both reference and actual test sets are digitalized representations of physical or behavioral phenomena presented by subjects. The process of a general biometric authentification consists of the steps of data acquisition, normalization, feature extraction and comparison. These steps are illustrated in Figure 4.1 .

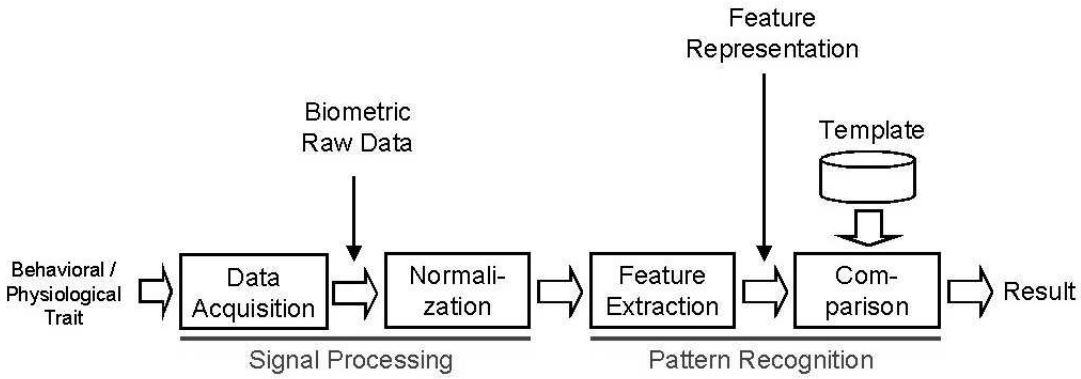

Figure 3. Process steps of a biometric system.

The biometric information representations are modeled as digital media data, and are typically based on the media types image, audio, video and movement. Examples for such biometric media are fingerprint or facial images, audio sequences of spoken language for speaker verification, video sequences of walking people for gait recognition or pen movement signals of digitizer tablets for movement-based biometric authentication.

Static biometric modalities like fingerprint, face, iris or retina recognition make use of discrete images, not taking into account any temporal information, whereas dynamic approaches consider the function of some physical measurement over time as the biometric signal. Examples for the later category are voice recognition (audio), handwriting and signature verification (e.g. pen-tip position and force signals over time) or gait recognition (video). Depending on the spatial and temporal resolution of the sensors applied for the biometric authentication, the raw data may become rather large.

It may vary from a few thousand bytes for small fingerprint images and low-frequency, low-dimensional online handwriting samples to several megabytes for video data. Table 1 presents some typical data sizes for non-compressed biometric raw data.

Biometric Modality	Typical Raw Data Size
Fingerprint Image	16 kB (128x128x1 Byte Gray Values) .. 900 kB 640x480x3 Bytes RGB
Facial Image	1,3 MB (780x570x3 Byte RGB)
Iris Image	1 MB (660x550x3 Byte RGB)
Retina Image	256 kB (512x512x3 Byte RGB)
Voice Print	214 kB (5sec Ω, 44kHz mono)
Handwriting signal	5 kB(x/y/p, 5sec, 200 Hz)

Table 1. Typical Sizes for Biometric Raw Data

Apparently, capacity requirements for embedding biometric raw data are rather demanding. Within the scope of our discussion on secure ID documents, it can be stated that embedding of full raw data is beyond the capabilities of today's technology both for digital watermarking and 2-D barcoding. Thus, in order to be able to embed biometric data in ID documents, data compression needs to be applied to the reference data. In this paper, we will discuss in more details possibilities of such compression for one specific biometric modality, the time signal of an online signature. We focus on this one modality, as it appears to be one of the most adequate modalities for integration in ID documents. Firstly, the offline signature already is one of the security features in many applications of ID documents. Quite often, the signature is used for user authentication, for example when writing a check or when making a purchase with a credit card. Integration of an online representation in addition to the offline (printed) representation of a signature allows an additional integrity check, as shown in.[1] Secondly, the process of subjects signing a document is socially well accepted and intrinsically implies a declaration of intension, which is not the case for passive biometric schemes. Furthermore, amongst the various different biometric modalities, online handwriting signals are rather compact in their raw data representation, as can be seen from Table 1.

Compression of biometric data can be performed either lossless or lossy. In a previous paper,[2] we have presented a lossless data coding technique optimized for online handwriting signals, and have shown that this method allows an increase in compression for this type of data of app. 30% additionally to the rate achievable by public packing algorithms such as Arhangel or ACB. However, even for samples with a low sampling rate and rather short writing sequences, the capacity requirements are still too high to allow an embedding in 2-D bar codes, not mentioning digital watermarks.

Therefore, in order to allow inclusion of biometric features into a document like an ID document, lossy data compression has to be applied. Indeed, for the purpose of comparison of reference data with an actual biometric sample, the full raw data contains a great degree of redundancy and feature extraction techniques can be applied to derive discriminatory characteristics from the original signal. A desirable goal here is to determine features, that show great intra-class stability and inter-class discriminatory power simultaneously.

If each specific feature vector in the value space represents one single user amongst all registered users of a biometric system, we denote functions having this property biometric hashes. In case the goal of such function is not to provide constant values for individual users, but rather codes with some intrinsic variability, those functions are called biometric signatures. Few discussions can be found in the literature on these biometric representations. Daugman for example presented the Iris code as a 2048 bit biometric signature,[4] Monrose

et. al. have presented a method to generate a biometric hash to be used as cryptographic key (hash) from spoken telephone numbers[3] and a biometric hash function for online handwriting has been presented in.[5] All these functions apply a lossy compression at high compression rate to the biometric raw data in such way that biometric authentication remains possible with a reasonable accuracy. As the achievable data sizes are within the capacity boundaries of 2-D barcodes in ID documents, we will briefly describe a possible scenario for biometric hashes based on online signature in the following subsection.

4.2. Biometric Hash

The biometric hash for online handwriting as presented in[5] generates a 24-dimensional hash vector from an actual sample signal and a stored personalized reference, called the Interval Matrix (IM). IM consists of two integer values for each of the 24 vector components, and can thus be represented as a 24x2 integer matrix.

The accuracy of all biometric systems is a trade-off problem between false-acceptances and false-rejections. Depending on the chosen operting point, a low false-acceptance-rate implies a high false-rejection-rate and vice versa. A common measure point for biometric systems is the so-called equal-error rate, the point in the operating curve where both the false acceptance and false rejection rates are identical. Our test results have shown that the biometric hash can be parameterized in such a way that a false acceptance rate of 7.05 % can be observed (7.05%of the forgers were able to generate an authentic biometric hash) at a false rejection rate of 7,05% (in 7.05% authentication attempts legitimate users where not verified). It is also possible to parameterize the system is such way, that the number of false-rejections is limited to some lower rate, at the cost of a non-zero false acceptance probability. In general, these test results justify the biometric hash to be reasonable for handwriting based authentication.

In respect to a general capacity requirement, each reference for a k-dimensional biometric hash requires exactly 2*2*k Bytes (i.e. two 16-bit Integer values per dimension), thus in case of k=24, 96 Bytes are necessary to store the Interval Matrix.

5. PROPOSED SYSTEM

5.1. System security design and key management

While a highly secure multi level security system could in theory be designed when the restrictions in an ID document application are ignored, for the actual deployment of security features it is important to consider the application environment. The following restricting factors need to be considered. (1) The provided security has to be in relation to the cost of implementation since the investment for any ID document is limited. (2) Another strong restriction is imposed by the available time for verification, in particular if documents are routinely verified. (3) The verification should be mostly free of possibility of human error, applicable without the need for significant qualification by the verifying authority. (4) Finally, the communication between the issuing authority and document holder and other verification authorities is limited, in particular for international travel documents. These limitations force a compromise between the achievable level of security on one side, and the usability and overhead for issuance and verification on the other side.

For the security of the identification document a combination of linked features is suggested. They vary in their purpose to either openly or covertly store data or detect fraudulent copies as described in Section 3. The paper is the carrier for all discussed features and a scanner and software are part of the verification device.

In particular for the digital watermark, the barcode and the Copy Detection Pattern, encryption is an essential component to increase the security of the system. Encryption limits access to the information to the groups in possession of the key but adds the necessity of key management. It would be a tempting approach to provide the ID document holder with the decryption key in order to enable her to grant rights over verification to any verification authority at her will. Furthermore, this solution would introduce an additional identification component, since this information would only be known to the legitimate ID holder. However, from a practical standpoint, this approach would in fact shift the burden of storing the key and keeping it secret to the user. This might not be a practical solution for IDs used in a large scale scenario, such as national identification documents.

We suggest encrypting the information of the decryption key for the watermark and the CDP in the barcode, using an asymmetric key to ensure that issuing of the barcode is restricted to the issuing entity while the

verification key can be distributed to several verification authorities. This approach minimizes the consequences of a compromised key -it could not be used to create a forged ID card- and keeps efforts for key handling to a minimum where the verification components could operate with a single key. Current digital watermarking schemes require a symmetric key, which is exposed to the risk of being compromised at the verification authority where it is required to read the watermark information. The suggested approach is creating an individual random key, used for the watermark and the CDP that is encrypted with a private key and stored in the barcode. The access and use of they key is limited to the time of verification and no storage is required. Every digital watermark and CDP is created with an individual key. The resulting advantage is an individual security feature for each document. Even if one document should be compromised the entire application is not at risk and the attacker cannot extend his knowledge to other documents. The concern with the proposed approach is that in the case the barcode could not be read, it would not be possible to proceed with the verification of the watermark and the CDP. This is a drawback that is inherently connected with the linkage of security features and the increased security that it offers

The payload of the digital watermark, the information that is hidden in an image, is used to connect the portrait of the document holder to her personal information contained on the document itself. The digital watermark thereby protects the image and the personal information. The personal information that is human readable on the document is protected against alterations by the duplicate of this information or its hash code stored in the encrypted watermark payload. The replacement of the image is protected since another image would not contain the correct watermark or no watermark at all. Even manipulation to the image can be detected. If the watermark is read despite the modifications, the altered areas can be located and highlighted for further inspection. Section 6 will discuss in more details the robustness of the system to different attacks .

An alternative approach would be to store a highly compressed image in the barcode. However, this storage requires additional space for the barcode on the ID. Furthermore, the watermarking solution supplies an automated way of authentification of the image, without the need of a human interaction.

5.2. Requirements and constraints

5.2.1. Personal information

To be read with a high reliability by the optical character recognition functionality (OCR), the personal information needs to be written in a standard OCR font such as OCR-A or OCR-B. The recognition of text from ID documents is standardized for international travel document and can seamlessly be integrated in this approach. The actual read from the document is optional and the trusted information is contained in the barcode. The information in the barcode can even increase the OCR performance since a proximity 1 to 1 matching approach can be applied.

5.2.2. 2-D barcode

Let us review the different types of data that have to be embedded in the barcode, and their corresponding bit size:

- 116 bytes (928 bits) to insert biometric hash
- 40 bytes (320 bits) to store the personal information that could be composed of:
  - name: 20 bytes
  - birth date: 2 bytes
  - ID number: 4 bytes
  - ID issue date: 2 bytes
  - ID expiration date: 2 bytes
  - additional information: 10 bytes
- 16 bytes (128 bits) to store the digital watermark and CDP key

The number of bits required is: $928 + 320 + 128 = 1376$ bits. To store the information, a widely accepted barcode format like PDF417 can be used that provides a storage density of 212 bytes / inch^2 or 22.3 bytes / cm^2, including error correction and detection coding. The physical space required is approximately: 0.81 inch^2 or 7.9 cm^2. Figure 2 shows an example barcode that can contain the 1376 bits information.

Let us note that the barcode is a sensitive element in the proposed application, since it must be read to authenticate the personal information, to enable CDP and watermark verification, and to confirm the legitimacy of the ID document holder with biometric verification. Barcode technology is mature and performance data exist for different substrates and printing technologies. To further reduce the risk of a deadlock situation when the barcode could not be read, it is suggested to apply a duplicate barcode printed on the back of the ID document to supply a verification backup.

5.2.3. Digital watermark

A compromise must be found between the visibility of the watermark and its readability. As the watermark does not contain information but is used only to detect alterations or modifications to the portrait, the actual watermark information that is embedded is known. However, due to the damaging print-scan process as well as wear and tear, several areas of the watermark will be damaged, and the detector will have to decide whether the damages are natural or result from an intentional manipulation. The reliability and granularity of this decision depends on the part of the watermark that is lost through the "print-scan-wear and tear" channel. To increase the amount that will survive, the watermark is embedded can be embedded with a higher robustness, which, on the downside causes visual artifacts in the image.

Also, three or four synchronization marks in the form of small printed dots are required to invert the rotation-scale-translation transformation that occurs during printing and scanning. The reliability of the watermark verification is very high, since the proposed scheme is not actually storing information but merely comparing the watermark information. This 1 to 1 match allows for a result that indicates the strength of the watermark rather than a binary decision about its existence, allowing to issue warning levels of severity that are combined with human judgment and the verification result of the other security features.

5.2.4. CDP

The CDP is based on measuring the quality of the printed image from a scanned version of it. The quality is lowered if the document has been through an additional print-scan process. However, there are other factors such as scanning resolution and device, paper quality, and wear and tear that influence the image's quality. The verification algorithm is designed to be able to discern different influences and is adjusted to the print and scan process. The outcome of the verification procedure is not a binary decision and, as for the watermark, can be used to issue warning levels of security.

5.3. ID document creation

Let us review the steps in the document creation application, where the input is: (1) personal information, (2) the digital image of the portrait, and (3) the biometric information. The output is a digital image that will be sent to the ID document printer.

- **Generate biometric hash** as described in Section 4: a 928-bit biometric hash is generated from the biometric data.
- **Generate watermark/CDP 128-bit key** using a pseudo-random number generator.
- **Create encrypted bitstream** using the private encryption key, and grouping the personal information, biometric hash, and CDP and watermark key in one bitstream.
- **Create 2D barcode** image from the encrypted bitstream
- **Embed digital watermark** into the portrait, using the watermark/CDP key.
- **Create CDP** using the watermark/CDP key.

- **Insert image elements in document template**: personal information, 2D barcode, watermarked portrait, and CDP.
- **Print ID document** from its digital image.

5.4. ID document detection

The detection process is basically the reverse of the embedding. As input, it requires (1) a scan of the ID document, (2) the decryption key, and (3) the ID card holder biometric data. We assume that any failure at any step will lead to a manual verification of the person's identity.

- **Read 2D barcode**, and extract the different pieces of data: personal information, biometric hash, CDP and watermark key
- **OCR** read personal information, and compare with the corresponding barcode information
- **Read watermark**, and detect altered areas (if any)
- **Read CDP**, and verify if it is a copy or original

To verify the biometric information, the user is prompted to provide biometric authentication data, by performing a dynamic handwritten signature on a digitizer tablet. The biometric authentication is then processed as follows:

- Extract the Interval Matrix IM B from the 2-D barcode
- Compute the IM B' from the actual writing sample
- The biometric authentication result is positive if the similarity between B and B' is over a threshold set for the application.

Again here, a level of confidence for the biometric verification, and different warning levels can be issued.

The time of the verification process is critical for the acceptance not only of the verifying authority but also the ID card holder. The times will vary depending on the system and are scalable by processing individual steps in parallel. A reasonable estimate for a common desktop system is as follow. The time to capture the image is approximately 1 second. The computationally intensive part that are the OCR, the barcode, CDP and watermark reading take about 0.5 seconds each. The matching and decision making process are not computationally intensive. The total time sums up to 3 seconds. During this time the biometric verification can take place.

The total reliability is composed of the reliability for the individual elements and, while the single components can be evaluated separately to flag a missing or crucial result, the reliability can be increased by comparing the result for different features and designing an optimized compromise between false detection and false rejection for ambiguous cases. Because the barcode is a crucial element required to verify the other features, it should be duplicated on the back side to enable a backup solution.

6. ATTACKS

The aim of ID security is to prevent the unauthorized use of ID documents. Unauthorized use can either be the usage of a legitimate ID document by an illegitimate user, or the usage of an illegitimate document. An illegitimate ID document can be created by either altering a valid existing document, or by re-creation of a document with reproduction of all security features. This section makes a review of these attacks, and attempts to estimate the cost and difficulty of executing them.

6.1. Usage by a non-authorized person

The biometric authentication process of the system presented in this paper makes use of the modality of handwriting dynamics, which may be imitated by a forger. Like for all other behavioral biometrics, the chances of being successful in this attack depend to a very high degree on knowledge and training. In,[1] three levels of forgery skills have been investigated for online handwriting verification systems. It has been shown that the False-Acceptance-Rate (FAR) increases roughly by one magnitude when an additional amount of knowledge and training is available to the forger. As a first estimation of false acceptances by the biometric authentication process, it can be expected that FAR will range between 2% (for zero-knowledge forgeries) up to approximately 50% for very skilled, well-trained forgers. However, in practice it must be estimated if there is a reasonable chance for a forger to get significant knowledge about the writing style of the original ID card holder. Due to the encryption and compression of the biometric reference on the ID card, reproduction of the original writing signals is virtually impossible. Thus, the only possibility to obtain information about the original holder's writing dynamics is a physical observation of the signature dynamics, or covert recording by some manipulated devices. These two attacks cannot be accomplished easily, particularly in observed environments.

6.2. Altering the personal information

This attack aims at altering the displayed personal information on the ID document. In the proposed solution, the personal information is replicated in the barcode, which is encrypted asymmetrically. A modification would not only require to change the print on the document but also to modify the barcode content. The strong encryption applied for the storage corresponds to a digital security level that is well defined. Although estimations about future possibilities to break encryption are not reliable, the encryption scheme and content of the barcode is not made public and poses an additional hurdle for this attack.

6.3. Replacing or altering the portrait

An important potential target for manipulation is a modification to the portrait, which is another biometric feature linking a person to the document. The digital watermark is a protection against this attack. If the image is replaced in its entirety, no watermark is present and the verification will flag it as suspicious. If the picture is taken from another document, the image will contain a watermark, but since each watermark is created with an individual key, the watermark will not be consistent with the key secured in the barcode and therefore will not be readable.Another possibility for manipulation is to change the semantic content of the picture without destruction of the watermark information in the picture. Various processes influence the image and the watermark stored therein. The watermark should still be readable after the image has been printed and if it is damaged or marked. Small alterations will be taken for regular loss within the procedure. Larger areas that significantly change the appearance of the image can be identified, since the structure of the watermark is known and regions with a high degree of alteration can be highlighted.

6.4. Creating a new ID document

To create a new ID document, the attacker has no other choice than to replicate the 2D barcode and personal information of a valid ID document. In the portrait to be displayed on this forged ID document, he must also insert a watermark matching the one with the replicated 2D barcode. For that, the attacker must extract the plain watermark from the original portrait, and insert it into the non-watermarked portrait. An estimate of the watermark can be made by scanning the ID document, estimating the underlying non-watermarked image, and substracting this estimate from the watermarked image.

The estimate will generally contain a fraction of the original watermark, the rest being mostly noise. To bring the watermark to an energy level that makes it readable by the verification device, the estimate must be amplified significantly before insertion in the non-watermarked image. From our internal tests, the estimate must have an energy level such that it results in a very noisy watermarked portrait. Furthermore, the CDP brings an additional level of security against this attack -see next subsection-.

6.5. Duplicating the ID document

Duplication using a valid document template is a common first step to create a fraudulent document and required for some of the attacks described above. The copy detection technology is for that reason an integral part of the proposed solution that allows identification of fraudulently copied documents and complements the solution for an attack that the barcode and watermark cannot identify. To attack the copy detection technology it could be attempted to either produce a very high quality copy or to re-create the original digital image and print it on similar printing equipment. While the printing and scanning technology is evolving, even very high quality and resolution duplication equipment was not successful in reproduction of a suitable CDP. The re-creation would require the understanding of the pattern, which is very complex and can not be adequately scanned or analyzed from paper. The reader is referred to [6] for a more detailed discussion on the security of the CDP.

6.6. System attacks: obtaining the key

The proposed scheme employs asymmetric cryptography for encryption and decryption of the 2D barcode. The encryption key is extremely sensitive, because if it is revealed the whole system would be compromised. Therefore, it should only be stored at the issuing authority, in a very secure way. The decryption key is stored in the ID document verification devices. To obtain the key the detection device has to be compromised and the key derived from there. This can be prevented or made extremely difficult with adequate security policies. The key can be sealed within the verification device physically or by cryptographic means.

In the worst case, if an attacker breaks into a verification device and gains knowledge of the decryption key, he is only able to read the watermark/CDP key in order to reproduce a valid ID document. Of course, the CDP creation and watermark embedding softwares are also required for that purpose, and can be made very difficult to access. Furthermore, the counterfeit document will contain the biometric information of the valid ID document. In all cases, the individualized key and biometric information ensures that no general solution can be enabled to create valid IDs.

7. CONCLUSION

The majority of ID cards, passports and other identity documents in use are poorly secured with archaic technologies, making them an easy target for counterfeiters. In the digital era, the "security by obscurity" principle cannot be seen as a persistent security strategy. However, by taking advantage of the huge progresses made in biometric authentication, digital data hiding and cryptographic techniques, we believe that virtually fraud-proof ID documents can be designed.

The suggested solution can be implemented with readily available technologies with hardware that is already largely available in the printing (issuing) side and the detection side with scanners that read passport pages. No special document treatment associated with additional cost is required. The only system modification is a software update that can be a single integration process to enable all features described above. The proposed solution offers a self authenticating document that does not require a database for evaluation and yet grants a highly secure document that is protected at several levels.

Acknowledgments

The research presented in this paper is partly based on an existing product, MediaSign Print, that took several man-years to develop, and the authors would like to thank MediaSec Technologies for this contribution.

REFERENCES

1. C. Vielhauer, L. Croce-Ferri: Applications of a hologram watermarking protocol: Aging-aware biometric signature verification and time validity check with personal documents, in: Delp, Edward J.; Wong, Ping Wah: Security and Watermarking of Multimedia Contents V, pp. 240 -248, SPIE 2003, 21. - 24. January, Santa Clara, CA USA, ISBN 0-8194-4820-6, 2003
2. L. Croce Ferri, M. Frank, C. Vielhauer, R. Steinmetz: Biometric authentication for ID cards with Hologram Watermarks, Security and Watermarking of Multimedia Contents, Proceedings of SPIE Vol. 4675, pp. 629-640, 2002

3. F. Monrose, M. K. Reiter, Q. Li and S. Wetzel: Using voice to generate cryptographic keys, Proceedings of Odyssey 2001, The Speaker Verification Workshop, June 2001.
4. J Daugman, High confidence visual recognition of persons by a test of statistical independence, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 15(11), pp. 1148-1161, 1993
5. C. Vielhauer, R. Steinmetz, A. Mayerhfer: Biometric Hash based on Statistical Features of Online Signature, Proceedings of the International Conference on Pattern Recognition (ICPR), 2002, pp. 1:123-126
6. J. Picard. Digital authentication with copy-detection patterns. in Optical Security and Counterfeit Deterence Techniques, SPIE 2004, San Jose, CA USA.