Print Signatures for Document Authentication

Baoshi Zhu
Institute for Infocomm Research
21 Heng Mui Keng Terrace
Singapore 119613
baoshi@i2r.a-star.edu.sg

Jiankang Wu
Institute for Infocomm Research
21 Heng Mui Keng Terrace
Singapore 119613
jiankang@i2r.a-star.edu.sg

Mohan S. Kankanhalli
Department of Computer Science
School of Computing
National University of Singapore
3 Science Drive 2
Singapore 117543
mohan@comp.nus.edu.sg

ABSTRACT

We present a novel solution for authenticating printed paper documents by utilizing the inherent non–repeatable randomness existing in the printing process. For a document printed by a laser-printer, we extract the unique features of the non–repeatable print content for each copy. The shape profiles of this content are used as the feature to represent the uniqueness of that particular printed copy. These features along with some important document content is then captured as the *print signature*. We present theoretical and experimental details on how to register as well as authenticate this *print signature*. The security analysis of this technique is also presented. We finally provide experimental results to demonstrate the feasibility of the proposed method.

Categories and Subject Descriptors

K.6.5 [**Management of Computing and Information Systems**]: Security and Protection—*Authentication*

General Terms

Design, Security

Keywords

Print signature, Laser printer, Authenticity, Originality

1. INTRODUCTION

Research in the authentication of printed paper documents has been growing because of its commercial potential. Although the problem has been tackled using cryptographic techniques in the digital domain [14], solutions for protecting physical documents, especially paper documents, is much less advanced. Paper documents still form the basis of today's business transactions and administrative processes, and "will continue to occupy an important place in office life, but will increasingly be used in conjunction with an array of electronic tools"[1]. For that reason, authenticating printed paper documents, which is the link between electronic tools and paper documents, becomes extremely important.

In the traditional paper-based world, when an authentic document is generated, it is usually signed/issued/approved by one or more authorized persons, with their signatures or seals to show the authenticity. The document with original signatures is considered to be original, authentic or legitimate. In the printed world, there are also requirements for such signatures to show the authenticity and originality of a document. Efforts towards this can be categorized into four classes:

- *Use of Special Material:* These solutions are based on either physical means or chemical means, such as special high-resolution (>4000dpi) printers not available in the open market, special papers/inks that are very sensitive to re-produce [21, 3, 13, 10, 11, 8, 26], and hologram labels [21, 5]. By controlling the availability of these materials, no forgery or duplication of the document is possible. However, due to the high cost of both the equipment and the efforts for controlling their use, these solutions are only used in applications which have strict security requirements, such as currency notes, checks, etc.

- *Fingerprints:* The idea of fingerprinting is to make each copy of a document unique so that illegal copies are identifiable, or the person who made illegal copies is traceable. This idea was first introduced by Wagner in [25], and then developed for various applications. In [17], nonuniformities in disk medium are utilized as fingerprint to discourage illegal copying of files. In [4], the width of each strip cut produced by a shredder is identified as the fingerprint, which in turn is used to trace the particular shredder that has been used. As for paper documents, Métois et al. [15] have proposed an identification system based on the naturally occurring inhomogeneities of the surface of paper. A special purpose imaging device is developed to capture the texture and fiber pattern of the paper. The pattern is then registered as a unique fingerprint for later retrieval and comparison. Physical fingerprints usually

CCS'03, October 27–30, 2003, Washington, DC, USA.

offer strong protection against duplication attempts. However, the medium is not content-related. Therefore, the integrity of the contents is not protected. Furthermore, the identification of typically invisible fingerprint often requires special devices. This inevitably increases the cost of the system. As a result, these methods are only used in applications which emphasize more on medium security than content integrity, such as checks, tickets, etc.

- *Digital Methods:* Originating from traditional cryptography, these approaches intend to transfer digital signature onto paper documents. Such approaches include bar codes [19, 20] and information hiding (notably digital watermarking) [6, 22]. These methods add some machine readable information onto the document to serve as a digital signature. Only authorized persons have access to the private key required to generate the digital signature, so the authenticity of the document is protected. However, since the information is machine readable, it can also be copied or scanned using photocopiers or scanners. The originality of the document is not protected effectively. Digital encoding methods have been widely used in applications which require machine based authentication, such as bills, ID cards, and so on.

- *Visual Cryptography and Optical Watermarks:* Visual cryptography utilizes secret sharing to split a graphical pattern into different pieces in a manner that the pattern becomes visible if and only if the shares are stacked together [16, 23]. By doing this, a paper document with one share printed can be validated visually using the remaining shares. Optical watermarks is an improvement over visual cryptography in terms of the ability to hide multiple layers of graphical information and enhanced visual quality with easy alignment [12]. Both visual cryptography and optical watermark have been designed for manual authentication of documents. They are most suitable in applications where the convenience of verification is important like in brand protection, ticketing, etc. However, both of these techniques cannot disprove the authenticity of a photocopy or scanned-copy of an original document.

The inherent shortcomings of existing authentication methods have limited their applications to niche areas. Developing a new technique suitable for business and administrative document processing is therefore imperative. In view of this, we present our novel *print signature* technique which has the following advantages:

- *Security:* The *print signature* is unique for each printed document. Any duplication attempt can be detected during the authentication phase. The content of the document is also used in the validation process. Thus, both authenticity and originality of printed paper documents are secured.

- *Convenience:* Our system can be implemented in a fully automated manner for high–speed batch processing. It can also be incorporated in handheld devices for manual operation.

- *Low Cost:* Our solution works on any ordinary laser printers. No special material or accessory is required. The cost of automatic verification devices is quite low as well.

The paper is organized as follows. In Section 2, we'll discuss the basis of our method. In Section 3, the detailed authentication process is analyzed. Experimental results are given in Section 4, followed by demonstrating some application areas of *print signature* in Section 5. The paper is concluded in Section 6.

2. BASIS OF THE METHOD

Authenticity and originality are two major requirements for printed document which need to be authenticated. It can be concluded from Section 1 that physical methods (special material and fingerprinting) prove more effective for establishing originality whereas cryptographic methods (digital encoding and optical watermarking) protect authenticity better. Our proposal combines the advantages of both these approaches. We first discuss new properties for protecting the originality of documents, then consider the integration issues with cryptographic techniques to lead to a complete solution.

2.1 Print Signatures

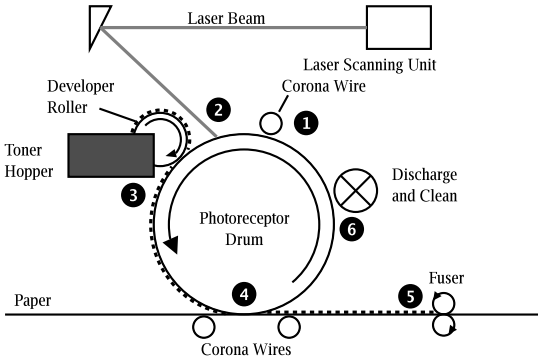
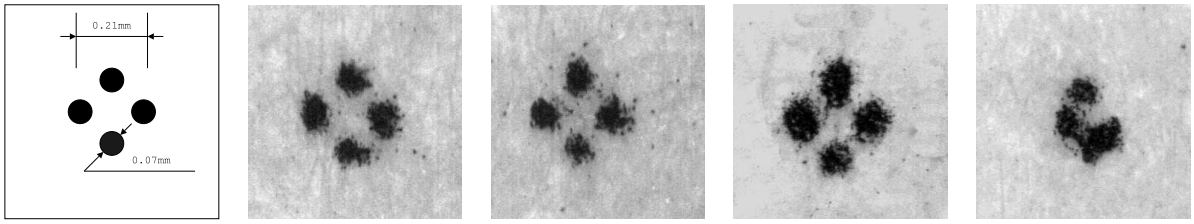


Figure 1: How laser printer works[1]

Figure 1 depicts the major components of a laser printer's imaging unit, which develops a piece of printed paper over six steps [2]: A photosensitive surface (photoreceptor) is uniformly charged with static electricity by a corona wire (1). Then the charged photoreceptor is exposed to an optical image through laser beam, which discharges it at desired positions to form a latent or invisible image (2). Development is done by spreading toner, a kind of fine powder, over the surface. The powder adheres only to the charged areas, thereby making the latent image visible (3). In the next step, an electrostatic field transfers the developed image from the photosensitive surface to a sheet of paper (4). The transferred image is then fixed permanently to the paper by fusing the toner using pressure and heat (5). The last step cleans off all excess toner and electrostatic charge from the photoreceptor to make it ready for the next cycle (6).

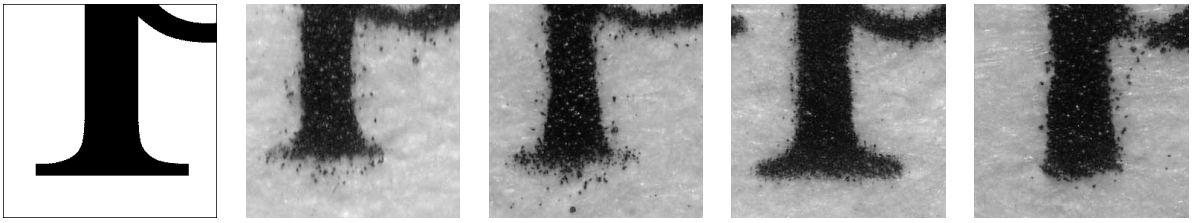
As no process repeats exactly, we expect to observe variations in each step. Such variations include the unevenness of the photosensitive surface and paper surface, the variable

[1]http://www.howstuffworks.com/laser-printer.htm



(a) (b) (c) (d) (e)

Figure 2: Printouts and Photocopies of the Testing Pattern



(a) (b) (c) (d) (e)

Figure 3: Printouts and Photocopies of Character "p"

granularity of the toner powder, unstable heat and pressure of the fuser, amount of excess toner remaining on the photoreceptor, and many other such factors. The net outcome of all these variabilities is that some toner powder gets randomly misplaced at undesired positions. Such misplacement is non–repeatable for each print run. This is because any repeatable defect can be detected during the quality control process and thus is fixed by improving the printer design. It is much harder to fix random phenomena hence they persist. Therefore, the pattern of misplaced toner powder on each paper is unique. We refer to this unique pattern by *print signature* as a metaphor for the manual signature on paper documents.

To study the characteristics of *print signatures*, we created a representative test pattern as shown in Figure 2(a). The pattern comprises of four rounded dots. The diameter of the dots is 1/360 inch and the horizontal and vertical distance between two adjacent dots is 1/240 inch. These two numbers are selected by taking both the physical limitations of the printer and experimental results into consideration. The size of the dots is larger than the theoretically smallest dots the printer can print (in this case 1/600 inch for a 600 dpi printer), so that the dots are clearly visible after printing. On the other hand, the dots are enough small for the random misplacement of toner powder to be significantly noticeable around their boundaries. The distance between two adjacent dots and the configuration of dots ensure that the printed dots will not merge together, which is very useful for our later segmentation process. The number of dots balances the authentication performance and required computational resources. We will provide more details on this topic in the next section.

Figure 2 shows some experimental printouts and photocopies examined under a 200× microscope. Image (b) and (c) are the test pattern printed using HP[2] LaserJet 8100 (600dpi) office printer. Image (d) is the same test pattern printed on a high resolution HP LaserJet 4050 (1200 dpi) printer. Apparently, the dissimilarity among these patterns is large. Even for the same printer, we obtain a large variance. Image (e) is a photocopy of image (b) using a 600×600 dpi digital photocopier Minolta Di152f[3]. It is quite obvious that the photocopied image is very different from the original one.

Besides the test pattern, occurrences of random toner powder misplacement can also be noticed at boundaries of printed characters, as shown in Figure 3, where images (a–e) are the source character, two test printouts on LaserJet 8100, one test printout on LaserJet 4050, and a photocopy of (b) on the Minolta photocopier respectively. We observe the same phenomenon noticed in the previous experiment that the *print signature* is random and non–repeatable for each print run.

We have performed many such experiments and have consistently observed this occurrence for several types of laser printers. The experiments demonstrate the uniqueness and randomness of our proposed *print signature*. Our method utilizes some features of this phenomenon to authenticate the originality of printer paper documents.

2.2 Overview of the Method

Without loss of generality, we describe our proposed method based on the type of *print signature* shown in Figure 2. We call this test pattern used in the experiment *secure pattern* as it enables certain security features. With some minor modification, our method can also apply to the *print signature* detected on printed characters as well as hand signatures. As illustrated in Figure 4, our method contains two procedures: registration and authentication.

- *Registration:* Given a document to be protected, we print the *secure pattern* onto some blank area of the paper. Several auxiliary landmarks are also printed around the pattern to facilitate alignment. The printed paper is then examined by a microscope. Features describing the *print signature* such as the shape of the

[2]http://www.hp.com

[3]http://www.minolta.com/flash-copier.html

dots are detected and extracted. The feature description, together with some critical information about the document (such as the seat number in a concert ticket), forms an unique identifier for this specific document and specific print run. A digital signature is then generated for the identifier. The digital signature and the identifier are printed onto the same document using digital encoding methods such as bar codes or OCR fonts. These printed information and the *secure pattern* are used for later authentication. Figure 5 is a sample concert e–ticket protected using our method.

- *Authentication:* In order to verify the authenticity and originality of a printed document, we first perform feature extraction like in the registration process to get the feature description of the *print signature*. Also, the encoded information is read from the document using either a bar code or an OCR scanner. The digital signature is verified first to ensure there has been no modifications to the document identifier. We then compare the extracted feature with the decrypted one, and the decrypted critical document information with the contents on the paper, through a decision process. If the results match, the document is considered to be authentic and original. Otherwise it considered to be a fake or to have been tampered.

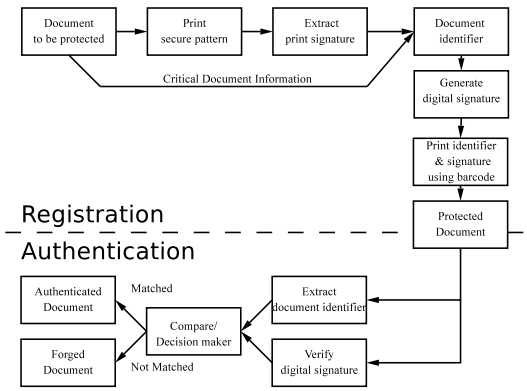


Figure 4: System diagram

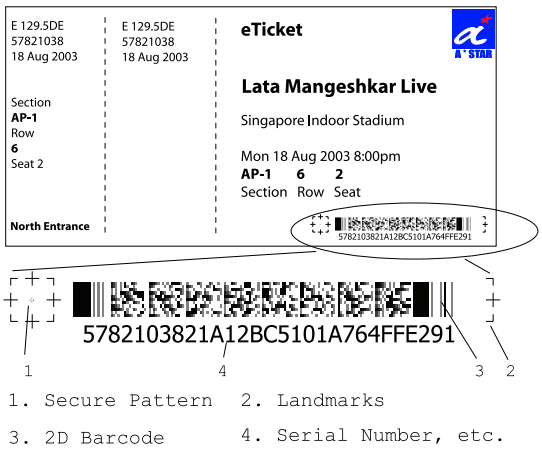


Figure 5: Protected e-Ticket

2.3 Feasibility Analysis

We formalize the registration and authentication procedures as follows:
The registration process can be described using:

$$S = (\{F(P), I\}, Sig(\{F(P), I\})) \tag{2.3.1}$$

where S is the printed information; $Sig(\cdot)$ is the digital signature scheme. P is the *print signature*; $F(\cdot)$ is the feature extraction function which is used to generate the description of *print signature*; and I is some critical information related to the document.
The authentication process can be described as:

$$\begin{aligned} V1 &= V_{sig}(\{F(P), (I)\}, Sig(\{F(P), I\})) \\ V2 &= DM(F'(P'), F(P), I, I') \end{aligned} \tag{2.3.2}$$

where $V1$ is the verification of digital signature, $V2$ is the verification of critical document information and *print signature*. $DM(\cdot)$ is the discriminative decision function; $F'(\cdot)$, P' are the feature extraction function and the *print signature* respectively. It should be noted that P and P', $F(\cdot)$ and $F'(\cdot)$ may not necessarily be the same. This is because $F(\cdot)$ and $F'(\cdot)$ are built into two different devices, and the similarity between P and P' depends on the condition of the paper (e.g. any salt and pepper noise) and the inspection environment (e.g. illumination, focusing of the microscope, etc.).

Suppose an attacker intends to forge a document either by recreating a new document or by modifying the contents of an authentic document. In this case, his major task is to create a valid digital signature which can pass the authentication procedure $V1$. This task is computationally infeasible unless the digital signature scheme used is compromised.

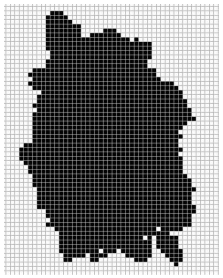


Figure 6: Quantized Dot Image

The attacker can also photocopy/scan–reprint an authentic document and claim it to be the original. The underlying task is to create a *print signature* P' which is the same as P, or satisfies $F'(P') = F(P)$ in order to pass $V2$. In [7], the authors have shown that completely recreating P through photocopying or scanning–reprinting with commercially available tools is impossible because of the nonlinear distortions and halftoning effects. Since it can be argued that P can be duplicated using professional equipment with higher resolution, let us refer to Figure 6. This is the leftmost dot of Figure 2(b) being examined under a 200× microscope with an CCD array of 320×288 pixels. After quantization, the dot covers an area of 40×59 pixels. Considering the physical size of the dot is about 1/360 inch, this specific shape needs at least $59/(1/360) = 21240$ dpi resolution printer to reproduce. The number is 17 times larger than today's highest–end laser printer which has a

resolution of 1200 dpi. This analysis shows that even if the attacker knows how an authentic *print signature* looks like, he does not have any method to create it. What he can do is exhaustively create and test various P', trying to find a collision wherein $F'(P') = F(P)$. In the following sections, we'll show that the probability of successfully creating such a P' is extremely low.

3. AUTHENTICATION PROCESS

The authentication function essentially compares the *print signatures* and assesses the degree to which a retrieved *print signature* matches the registered one. In what follows, we will first describe the feature extraction of *print signature*, then describe our matching algorithm and analyze its performance as well as security.

3.1 Feature Extraction for Print Signature

Feature extraction for *print signature* is performed during registration as well as authentication. It takes captured images of the *print signature* P as the input, and extracts the most descriptive features such as shapes, profiles, or spatial configuration of P as the output.

In our test setup, the "IntelPlay QX3" [4] computer microscope is used to capture the image of the *print signature* P. We have selected this cheap (costing 50 US dollars) microscope as a low–cost scanner. As a result, the quality of captured images is not always satisfactory. As shown in Figure 2(b-e), only the rough shape of the four dots is invariant under illumination and focus changes. Therefore, we binarize and segment the four dots before extracting the shape as the descriptor of the *print signature*.

- *Binarization:* Binarization involves the selection of an optimal threshold such that the major features are preserved during conversion from a grayscale to a binary image. A good binarization accuracy is very important for the registration process which requires a precise description of the *print signature*. Illumination conditions and focus have a major influence on this. In a bad–illuminated and out–of–focused image, the edges of the four dots will be blurred, thus destroying the accuracy of binarization and later shape retrieval. To overcome this in the registration procedure, we first capture a set of images under different focus and illumination conditions. Then the images are binarized using the optimal threshold value defined by Otsu's algorithm [18] which attempts to maximize the inter–class variance between the class of pixels above the threshold, and the class of pixels below. The average form of these images is used as binarization result. During authentication, we use Otsu's threshold directly.

- *Segmentation:* Since we have the a priori knowledge of how the *secure pattern* looks like, we can easily segment the four dots from the background. In case the four dots cannot be segmented successfully, this information is fed back to the binarizer, instructing it to adjust the threshold and redo the binarization. A set of segmented images for Figure 2(b) are shown in Figure 7.

[4]http://www.intel.com/support/intelplay/qx3/

- *Feature extraction:* The security of our method relies on the performance of feature extraction and matching for the *print signature*. In the test setup, shapes of the four dots have been identified as the main feature of *print signature*, so the problem is reduced to a shape matching problem. In the computer vision literature, shape matching methods have been studied for a long time, and a variety of solutions have been discovered. [24] provides a good review of these developments. However, our requirements are a bit different. Traditional shape matching algorithms are usually robust against affine transforms like translation, scaling or rotation. In our method, since landmarks are used to assist alignment, this robustness is not required. We can therefore use a simple radius profile as the descriptor for the shape of the dots.

  For each dot, the radius is calculated from its centroid to the perimeter. Typical perimeter length of a dot in our experiment is between 180 to 200 pixels. Using all of them as the shape descriptor is not acceptable because:

  1. The radius values are not independent of each other. Instead, the position of each pixel on the perimeter is determined by the pixels beside it. If the radius values of all pixels along the the perimeter are used, there will be a lot of redundancy in the data set.
  2. Finally the descriptor will be encrypted and encoded using symbolic encoding methods such as bar codes or OCR fonts. These methods have limited storage capacity.

  Therefore, we partition the radius profile proportionally into several segments using polar coordinates and then calculate the average radius value r' for each segment, as shown in Figure 8. Considering that the correlation between adjacent radius values of fan–shaped segments is low, we assume the obtained set of r' are independent variables. Thus, the profile can be repre-

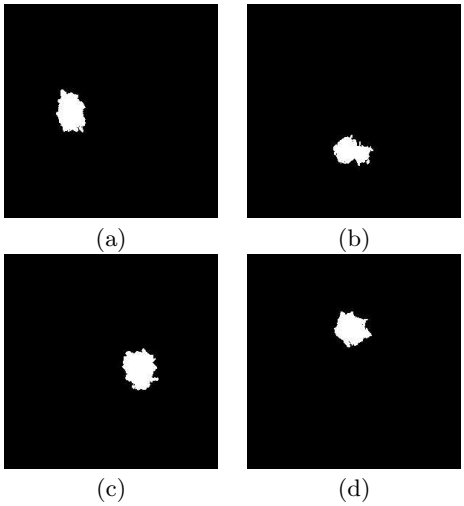


Figure 7: Segmented Secure Pattern

sented as:

$$\vec{R'} = (r'_1, r'_2, \ldots, r'_{N-1}, r'_N)^T$$

Since we are only interested in the shape of the dots but not the size, we normalize the profile using its mean value:

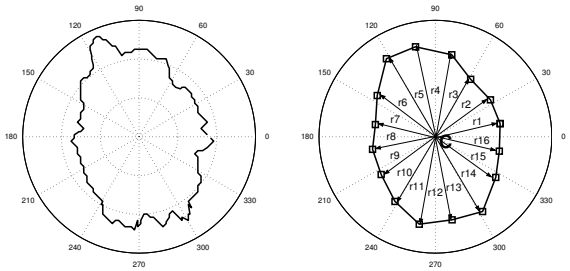
$$\bar{r} = \frac{1}{N}\sum_{i=1}^{N} r'_i$$

$$\vec{R} = (\frac{r'_1}{\bar{r}}, \ldots, \frac{r'_N}{\bar{r}})^T = (r_1, r_2, \ldots, r_{N-1}, r_N)^T \quad (3.1.1)$$

Here N is the number of segments. It is a critical parameter for the overall security. We leave the discussion about N for the next section. Thus, the feature description of our *print signature* is represented as:

$$F(P) = \{\vec{R}_1, \vec{R}_2, \ldots, \vec{R}_M\} \quad (3.1.2)$$

where M is the number of dots used in the secure pattern.



(a) Profile of Dot (b) Average profile

Figure 8: Profile of Printing Signature

3.2 Profile Matching

Referring to the matching function 2.3.2, DM is a discriminative decision function that measures the similarity between extracted profiles of the *print signature* and the registered profiles. It must be carefully selected so that no authentic document is rejected (false–alarm rate is low) and no forged document is accepted (false–acceptance rate is also low). Given a reference profile $\vec{R}_{ref}$ for one dot, we use the following Euclidean distance classifier to differentiate profiles:

Define Euclidean distance as:

$$\begin{aligned} D(R, R_{ref}) &= \|\vec{R} - \vec{R}_{ref}\|^2 \\ &= \sum_{i=1}^{N} (r_i - r_{ref,i})^2 \end{aligned} \quad (3.2.1)$$

For threshold T, we consider:

$$\begin{aligned} D(R, R_{ref}) < T &\rightarrow R \text{ and } R_{ref} \text{ are the same} \\ D(R, R_{ref}) \geq T &\rightarrow R \text{ and } R_{ref} \text{ are different} \end{aligned} \quad (3.2.2)$$

and DM:

ACCEPT, if:

R and R_{ref} are the same for all M dots.

REJECT, if:

R and R_{ref} are not the same for any of M dots. (3.2.3)

Here, T, N and M are to be determined together with their performance analysis.

Matching using profiles can also be used for other types of *print signatures* such as the one detected on characters as shown in Figure 3. The profile of an arbitrary shape is obtained by calculating the distance from its outermost perimeter to its centroid. But since there is no fixed location for the *print signature* to be detected, the location of a specific *print signature* must be encoded into the barcode, and a positioning mechanism is required to locate the *print signature* precisely.

3.3 Performance Analysis

In equation 3.2.3, T and N are two very important parameters which determine the performance of the classifier. When T increases, the classifier becomes more robust against noise, but the false–acceptance rate increases. Conversely, when N increases, D also increases, the classifier becomes more sensitive to variance, but the false–alarm rate increases as well. We regard the radius values of the dots as random variables and then use a statistical model to estimate the optimal values for T and N.

To simplify the analysis, let us assume $M = 1$, that is, only one dot is used in the secure pattern. As shown in Equation 3.1.1, for a single dot A, profile $\vec{R}$ can be considered as a joint distribution of independent random variables $r_1 \ldots r_N$. The randomness of these variables comes from the environmental conditions when capturing the image, and the threshold value used to binarize the image. To study the distribution of these variables $r_1 \ldots r_N$, we captured 100 images for the same dot using different illumination and focusing conditions. For each image, we use 5 distinct threshold values for binarization. So altogether we obtained 500 binarized images. Then for each different N from 8 to 64, we partition the profile into N segments and calculate the average profile $\vec{R}_{j,N}$, $j = (1 \ldots 500)$. The distribution of each radius value r_i in $\vec{R}_{j,N}$ can be determined by computing the histogram for $r_{j,N,i}$, $j = (1 \ldots 500)$, $i = (1 \ldots N)$ for each i. In our experiment, the shapes of the histograms had a Gaussian profile, so we assume r_i obeys a Gaussian distribution. By using the "Bera-Jarque Normality Test"[9], our hypothesis is verified with a P–value (significance level, the larger the better) of 0.4 (as oppose to confidence level of 95%).

Another useful result that we obtained from the experiments is that the standard deviation for each set of r_i is almost the same. This can be explained as follows: the randomness is homogeneous for all directions. We use symbol σ to denote the standard deviation hereafter.

Let

$$\bar{R} = (\bar{r}_1, \bar{r}_2, \ldots, \bar{r}_{N-1}, \bar{r}_N)^T$$

denote the mean profile of dot A from the above experiments, and

$$R = (r_1, r_2, \ldots, r_{N-1}, r_N)^T$$

denote the profile of dot A obtained from one test, we have

$$\frac{r_i - \bar{r_i}}{\sigma} \sim N(0, 1), \qquad i = (1 \ldots N)$$

which means that $\frac{r_i - \bar{r_i}}{\sigma}$ obeys the Gaussian normal distribution.

Now consider

$$D(R, R_{ref}) = \sum_{i=1}^{N} (r_i - r_{ref,i})^2$$

Note that the reference value R_{ref} is obtained from the registration process where the average value of multiple images is used. So $r_{ref,i}$ must be very close to $\bar{r_i}$. Thus,

$$D(R, R_{ref}) \doteq \sum_{i=1}^{N} (r_i - \bar{r_i})^2$$

It therefore follows that:

$$\frac{D(R, R_{ref})}{\sigma^2} = \sum_{i=1}^{N} (\frac{r_i - \bar{r_i}}{\sigma})^2$$

So

$$\frac{D(R, R_{ref})}{\sigma^2} \sim \chi^2(N)$$

is a chi-square cumulative distribution with N degrees of freedom.
We demand that the "false-alarm" rate be lower than 0.5%, or,

$$P(D(R, R_{ref}) > T) < 0.005$$

Table 3.3.1 shows the acceptable T and N values under this requirement.

N	σ^2	$\chi^2(N) = 0.995$	T
72	0.001589	106.7	0.16940
36	0.001241	61.58	0.07641
32	0.001246	56.33	0.07017
24	0.001063	45.56	0.04841
16	0.001015	32.27	0.03478
8	0.000821	21.96	0.01801

Table 3.3.1: Choice of N and T under false-alarm rate $< 0.5\%$

In Section 2.3, we raised the question that whether it is possible to find another profile P' such that $F'(P') = F(P)$. We rephrase the problem under the current context to ask the question: given a discriminative function D and parameters N, T, how large is the probability for two distinct profiles R and R', to have $D(R, R') < T$.

To answer this question, we must know the distribution of the radius r across different dots. Our experiment on 400 different dots shows that the distribution of r is Gaussian, with an average P-value of "Bera-Jarque Normality Test" of 0.2. The result is explained as: the average radius value of the dots is our designed dot's radius in the "secure pattern". By the central limit theorem, the distribution of radius value must conform to a Gaussian distribution under the sum of a large number of random influences.

Now consider two distinct profiles R and R', and

$$r_i, r'_i \sim N(\bar{r}, \sigma_r^2), \qquad i = (1 \ldots N)$$

where $\bar{r}$ is the average radius value and σ_r is the standard deviation of radius r, we have

$$(r_i - r'_i) \sim N(0, 2\sigma_r^2),$$
$$\frac{r_i - r'i}{\sqrt{2\sigma_r^2}} \sim N(0, 1), \qquad i = (1 \ldots N)$$

Obviously, for

$$D(R, R') = \sum_{i=1}^{N} (r_i - r'_i)^2$$
$$\frac{D(R, R')}{2\sigma_r^2} \sim \chi^2(N)$$
$$P(D(R, R') < T) = P(\frac{D(R, R')}{2\sigma_r^2} < \frac{T}{2\sigma_r^2})$$

Substituting T and N using the values shown in Table 3.3.1, we have the false–acceptance rate as shown in Table 3.3.2.

N	σ_r^2	$P(D(R, R') < T)$
72	0.02387	4.468E-34
36	0.02228	1.333E-18
32	0.01846	8.636E-15
24	0.02078	3.787E-13
16	0.01846	3.950E-08
8	0.01291	4.680E-04

Table 3.3.2: False–acceptance rate

We can find that the probability of successfully creating a *print signature* whose profile P' can pass our authentication process is very low even for only one dot. For a "secure pattern" with M dots, since we accept a document to be authentic only when the profiles of all M dots are matched, the false–alarm rate will be

$$1 - (1 - P_{f.alarm})^M,$$

and the possibility of false–acceptance wherein all M faked dots are matched will be

$$P_{f.accpt}^M$$

Use $P_{f.alarm} = 0.5\%$ and $N = 32$ as an example, the false–alarm rate becomes 1.98% and the false–acceptance rate is reduced to $5.562E - 57$. The false–alarm rate is slightly increased but the false-acceptance rate is greatly decreased. Of course, the use of more dots requires more computation. The choice of M should be balanced between security concerns and acceptable resource costs.

It must be pointed out that our reasoning is based on the assumption that the radius r is a continuous random variable. In practice, since the value must be quantized, the false–alarm and false–acceptance rate will be amplified. However, this problem can be mitigated by the use of high resolution image sensors.

4. EXPERIMENTAL RESULTS

In this section, we will present extensive experimental results to demonstrate the feasibility of our proposed method for document authentication. Our experiments are intended to test whether an authentic document can successfully pass the authentication process, and a forged document can be successfully rejected. We use $N = 32/T = 0.07017$ as shown in Table 3.3.1, because these values seem to be optimal in terms of good discriminative power as well as low storage requirement. The "secure patterns" used in the experiments are composed of 1–4 small dots respectively. The printers used are a HP LaserJet 8100 monochrome printer and a HP LaserJet 4600 color printer. For the color printer,

the "secure pattern" was printed using the black channel. The paper we used in the experiments include plain paper, color paper, translucent paper and card paper. These types of paper are widely used in all kinds of business and administrative documents such as certificates, bills of lading, invoices, licenses and checks. We printed authentic docu-

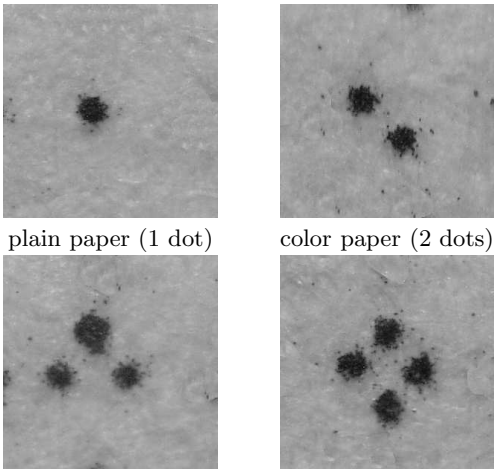


Figure 9: Experimental Print Signatures

ments with combinations of "secure patterns", printers and papers. Some captured pattern images are shown in Figure 9. The reference profile for a "secure pattern" was obtained by averaging the results from multiple tests as was discussed in Section 3.1. For each authentic document, we capture another 50 images by changing illumination, focusing, and by applying some small mis–alignments. The Euclidean distances between the profile of these images and the reference profile are marked using 'o' in Figure 10. We also created 50 forged copies for each authentic document by reprinting/scanning–reprinting the same document. We did not perform the photocopying test because our initial experiments had shown that the quality of photocopied documents is very bad. The *print signature* was destroyed to such an extent that we could not even segment the dots. The Euclidean distances between the profile of these forged images and the reference profile are marked using 'x' in Figure 10.

As we can see from the results, no forged document has been accepted. But as the number of dots increases, there have been a few occasions that authentic documents are rejected. This result conforms to our designed false–alarm rate of 0.5% for one dot and 1.98% for four dots. Rejecting almost all forged documents at the expense of erroneously rejecting a few authentic documents is acceptable since in most cases forged documents can cause a lot more damage. However, if we perform another round of validation when a document is rejected, the false–alarm rate can be greatly reduced.

5. APPLICATIONS OF PRINT SIGNATURE

Print signature is particularly suitable for applications that require documents to be protected against unauthorized duplications while allowing the verification of the document to be convenient as well. Such applications include *Bill of Lading*, *On–line Ticketing*, *Lottery Tickets*, *Voting*

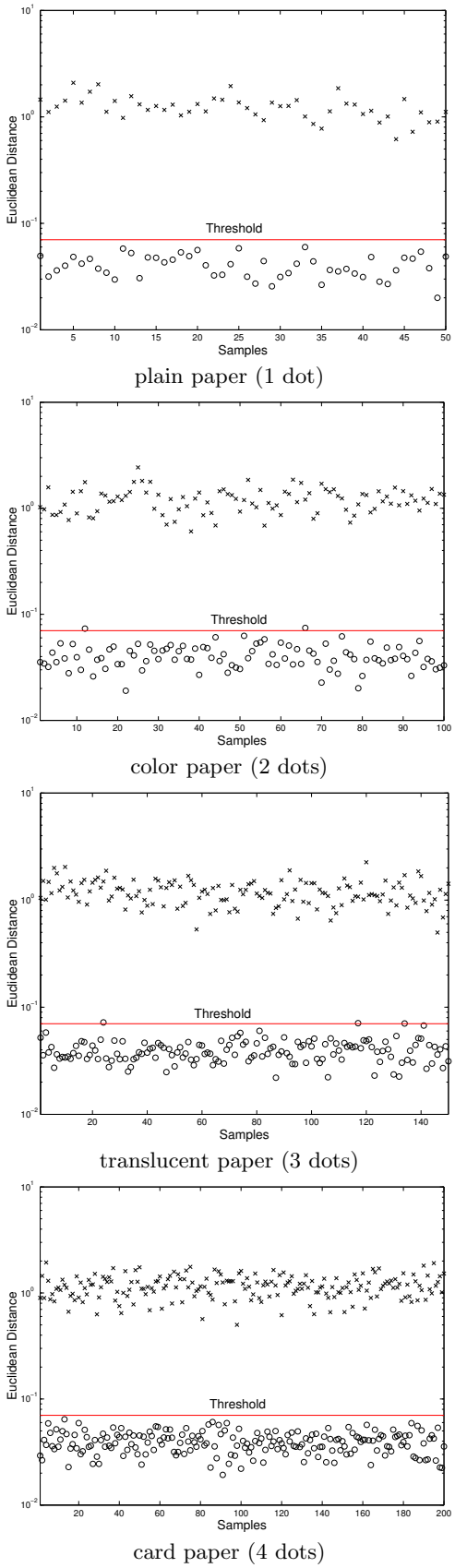


Figure 10: Experimental Results

Ballot Paper, etc. Let us take bill of lading (B/L) for shipping industry as an example to see how *print signature* can be embedded into a practical workflow.

The shipping industry relies heavily on the use of original paper document for (B/L). In a typical workflow, the original copies of (B/L) have to be printed by the shipping company, then couriered to the customer, which results in courier costs and additional delay in the entire transaction process. The current workaround is that some shipping companies give their valued customers pre–printed blank (B/L) so that their customers can print the details and sign on behalf of the shipping company. However the shipping companies are unable to control or track what is actually printed on these forms, and who has access to the forms. Alteration, forgeries and fraud using such documents are common. When there are discrepancies suspected in (B/L), document verification involves cross–checking of the details with various parties and the verification of the authenticity of the document. This is often a resource–consuming and costly process. Handling (B/L) electronically is a global trend, but its acceptance and implementation is still far away. (B/L), as an official contract between shipping company and its customer, is required to be presented in a paper form by the legal system in many countries.

By incorporating *print signature* into the workflow, a Internet based (B/L) delivery system can be built. Instead of giving their customer pre–printed blank (B/L), shipping companies give/loan their customers specialized printers and computer systems capable of printing (B/L) and making *print signatures*. The printing process in done interactively with shipping companies' central database server. *Print signatures* are generated for each printed page together with the page content (retrieved from database). Digital signatures are made by the server, so that the customers cannot perform any forgery attacks. Holding this (B/L) with valid *print signature*, customers are ready to claim goods shipped to them.

At the verification side, the warehouse keepers have been equipped with devices capable of authenticating (B/L) by validating the *print signature*. The verification process is done off–line with no need to access any database. The detailed procedure is the same as that was discussed in Section 2.2. It is possible that different shipping companies can share one single verification device, because digital signatures can sufficiently differentiate them apart.

As can be seen from the (B/L) workflow, *print signature* has the ability to successfully close the authenticity gap between the electronic world and the paper world. We may consider *print signature* as an authenticating technique in the paper–based world, but it does create a lot of new possibilities in the electronic world, which is beneficial for the global e–commerce drive.

6. CONCLUSIONS AND FUTURE WORK

In this paper, we have proposed a novel authentication technique for printed paper document. The *print signature* is based on the inherent randomness present in the physical printing process. The security of the method is guaranteed by both the digital signature and the *print signature*. The method has been demonstrated to be secure against forgery and duplication attacks.

As the laser printing technology improves, the printing resolution will become even higher. However, as long as the underlying mechanism is unchanged, we still expect to see the random phenomenon on each copy of printed paper. This will only entail the use of microscopes of even higher resolution.

This method can be readily extended to other document types such as offset–printed documents, ink–jet printed documents, or manually signed documents. It basically reduces to the task of finding unique random phenomenon in each copy of the document to be used as a signature. For example, the ink trail for each manually signed document is unique. As long as the uniqueness is found, a new document authentication method based on the same principle can be developed.

7. REFERENCES

[1] J. Abigail and R. H. R. Harper. *The Myth of the Paperless Office*. MIT Press, Cambridge, MA, 2001.

[2] S. J. Bigelow and E. Kuaimoku. *Easy Laser Printer Maintenance and Repair*. Windcrest/McGraw–Hill, Blue Ridge Summit, PA, 1994.

[3] Borowski Jr *et al.* Surface treated security paper and method and device for producing surface treated security paper. US Patent Number 5,193,854, 1993.

[4] J. Brassil. Tracing the source of a shredded document. In *5th International Workshop on Information Hiding*, Oct. 7–9, 2002.

[5] Constant and N. James. Holographic identification system using incoherent light. US Patent Number 4,820,006, 1989.

[6] I. Cox, J. Kilian, T. Leighton, and T. Shamoon. Secure spread spectrum watermarking for multimedia. *IEEE Transactions on Image Processing*, 6(12):1673–1687, 1997.

[7] N. Degara-Quintela and F. Pérez-González. Visible encryption: Using paper as a secure channel. In P. W. Wong and E. J. Delp, editors, *Security and Watermarking of Multimedia Contents V, Proc. of SPIE*, volume 5020, 2003.

[8] E. B. Greene *et al.* Coatings and ink designs for negotiable instruments. US Patent Number 6,155,604, 2000.

[9] G. G. Judge *et al. Introduction to the Theory and Practice of Econometrics, 2nd Edition*. John Wiley & Sons, New York, 1988.

[10] E. B. Greene. Negotiable instrument. US Patent Number 4,634,148, 1987.

[11] E. B. Greene and D. Jonathan. Security document. US Patent Number 6,089,610, 2000.

[12] S. Huang and J. K. Wu. Optical watermark. WIPO-PCT Patent Number WO 0,223,481, 2000.

[13] Kimura and Yoshihiro. Woven security label. US Patent Number 6,068,895, 2000.

[14] A. J. Menezes, P. C. van Oorschot, and S. A. Vanstone. *Handbook of Applied Cryptography*. CRC Press, 1997.

[15] E. Métois, P. Yarin, N. Salzman, and J. R. Smith. Fiberfingerprint identification. In *Third Workshop on Automatic Identification*, Mar.14–15, 2002.

[16] M. Naor and A. Shamir. Visual cryptography. In *Proc. EUROCRYPT 94*. Springer, 1994. Lecture Notes in Computer Science No. 950.

[17] A. D. Narasimhalu, W. Wang, and M. S. Kankanhalli. Method for utilizing medium nonuniformities to minimize unauthorized duplication of digital information. US Patent Number 5,412,718, 1993.

[18] N. Otsu. A threshold selection method from gray-level histogram. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-9(1):62–66, 1979.

[19] T. Pavlidis, J. Swartz, and Y. P. Wang. Fundamentals of bar code information theory. *Computer*, 23(4):74–85, Apr. 1990.

[20] T. Pavlidis, J. Swartz, and Y. P. Wang. Information encoding with two–dimensional bar codes. *Computer*, 24(6):18–28, June 1992.

[21] R. L. van Renesse, Ed. *Optical Document Security, Second Edition*. Artech House, Inc, Norwood, MA, 1998.

[22] Rhoads and B. Geoffrey. Identification/authentication system using robust, distributed coding. US Patent Number 5,745,604, 1998.

[23] A. Shamir. Method and apparatus for protecting visual information with printed cryptographic watermarks. US Patent number 5,488,664, 1996.

[24] R. C. Veltkamp and M. Hagedoorn. State-of-the-art in shape matching. Technical Report UU-CS-1999-27, Utrecht University, the Netherlands, 1999.

[25] N. R. Wagner. Fingerprinting. In *Proceedings of the 1983 IEEE Symposium on Security and Privacy*, pages 18–22, 1983.

[26] Zeira *et al.* Verification methods employing thermally-imageable substrates. US Patent Number 6,107,244, 2000.