Regular Paper

An Evaluation Method for a Magnetic Artifact-metric System

Hiroyuki Matsumoto[†,††] and Tsutomu Matsumoto[†,†††]

We have studied an individual authentication systems that authenticate artifacts by verifying the inherent patterns randomly created on them, and to which we refer as "*artifact-metric* systems." For example, magnetic texture can be created on documents by dispersing magnetic material randomly throughout their substrate. In this paper we describe performance of such a magnetic *artifact-metric* system that verifies the magnetic texture on the document. We illustrate how to efficiently evaluate the accuracy of authentication for the system, and then discuss changes in accuracy with variations in size of intrinsic patterns. Enhancement of the performance is also examined.

1. Introduction

Recent high-tech counterfeiting with desktop publishing techniques has stimulated research and/or extension efforts on document security. In order to achieve a secure anti-counterfeiting system, we have focused on such individual authentication systems that authenticate intrinsic patterns from inherent texture randomly created on artifacts, and have classified them as *artifact-metric* systems [4),5)]. Security of the *artifact-metric* systems is based on difficulty in reproducing the intrinsic patterns which provide evidence of genuineness. We have utilized intrinsic patterns from magnetic texture which can be inherently created by scattering magnetic micro-fibers randomly throughout an artifact, and developed a magnetic *artifact-metric* system which we call "FibeCrypt" [2)~5)].

In this paper, we evaluate the accuracy of authentication of a magnetic *artifact-metric* system with such indicators as the false non-match rate (FNMR) and false match rate (FMR), or the receiver operating characteristic (ROC) curve [1),6)]. This paper discusses the accuracy of authentication without using *clones*, which mean the things produced by dishonest ways such as counterfeiting, alteration, duplication, simulation or substitution.

The performance evaluation usually require many artifact samples and troublesome, repetitive operations. However, in the development stage of an *artifact-metric* system, it is often difficult for us to evaluate its performance precisely due to either a shortage of samples or their durability. In this paper, we describe how to evaluate accuracy of authentication for a magnetic *artifact-metric* system, using a small number of artifact samples and a magnetic *artifact-metric* system to coordinate its algorithm.

We measure the actual paper documents to acquire intrinsic patterns, and then analyze their errors distribution. We extend the results by adding calculated errors to the template. The FNMR and FMR curves, and the ROC curves are presented to examine the relationship between the accuracy of authentication and the number of elements used for verifying the intrinsic patterns. Enhancement of the performance is also examined.

2. The Magnetic *Artifact-metric* System

The magnetic *artifact-metric* system which we discuss in this paper is shown in **Fig. 1**. We use paper documents throughout which magnetic micro-fibers, containing iron oxide particles at the rate of 70 wt.%, are randomly dispersed. The diameter and length of fiber are respectively around 0.03 mm and 5 mm. The average density of fibers in a square meter is one gram, and the size is 210 × 75 mm. We call the paper documents "*F-papers*." Intrinsic patterns of an *F-paper* are captured by a magneto-resistive sensor in the magnetic reader as a magnetic signal, and can be quantized into 256 numbers by an analog to digital converter, and then transferred to the personal computer (PC) via the RS232-type serial interface. Thus, the magnetic reader outputs an intrinsic pattern according to magnetic texture while scan-

† Graduate School of Engineering, Yokohama National University
†† Information & Security Systems Division, NHK Spring Co., Ltd.
††† Graduate School of Environment and Information Sciences, Yokohama National University

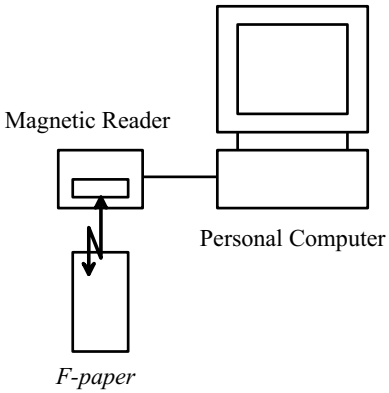

Fig. 1 The magnetic *artifact-metric* system consists of a magnetic reader and a personal computer.

ning the *F-paper*. Finally, the PC authenticates the *F-paper* by verifying the intrinsic pattern.

3. The Accuracy of Authentication

3.1 Performance Evaluation

In repetitive authentication of an artifact, intrinsic patterns from the artifact vary due to errors in capturing and preprocessing. These include errors in captured texture, motion of the artifact, detection with sensors, signal conversion and data compression. To achieve repeatable measures in the authentication, *artifact-metric* systems commonly compensate for the errors mechanically, electrically and logically in their processes. Although the compensation provides stable authentication, it brings about the fact that not all intrinsic patterns are distinctive.

Artifact-metric systems are similar to the biometric systems in which the inevitable errors occur in authentication. Therefore, the accuracy of *artifact-metric* systems can be evaluated by the same method as that for the biometric systems, and accordingly indicated by the false non-match rate (FNMR) and false match rate (FMR) [1]. The FNMR and FMR are functions of the decision threshold which the system applies to its pattern matching algorithm. The FNMR is the probability that an *artifact-metric* system will fail to verify the identity of a legitimate artifact, and the FMR is the probability that the *artifact-metric* system will incorrectly identify an artifact. We are using the FNMR and FMR, in this paper, to refer to the primary accuracy without any enhancement of performance.

Other parameters, i.e., the false rejection rate (FRR) and the false acceptance rate (FAR) have become well-known to public, and are often used as the indicators of the accuracy of authentication for an individual authentication system. We may be able to improve the accuracy by applying some protocols, e.g., a retrying protocol, to the system. Therefore, we are using the FRR and FAR to refer to the ultimate accuracy regardless of performance enhancement.

Furthermore, the equal error rate (EER) is defined as the probability of errors when the decision threshold is set such that the FNMR equals to the FMR (or the FRR equals to the FAR), and commonly used as a representative indicator of the accuracy. Also, the receiver operating characteristic (ROC) curves, which indicate the relation between the FNMR and FMR (or, the FRR and FAR), are used for evaluation of the accuracy as well as for comparison of several evaluation tests.

3.2 Samples for Performance Evaluation

Performance of *artifact-metric* systems should be evaluated both for usability and for security. If we examine the accuracy of authentication for the system without using *clones*, we can evaluate the primary performance of the system mainly from the point of view of usability. This examination also enables us to evaluate security of the system against unregistered samples, namely, non-effort forgery. If we examine the accuracy of authentication of the system using *clones*, we can evaluate *clone* resistance of the system mainly from the point of view of security. This paper concerns the primary performance evaluation of the magnetic *artifact-metric* system without using *clones*.

3.3 Problems in Performance Evaluation

The FNMR curve can be obtained by operating repetitive authentication using the artifact samples and their templates which are previously registered with the system. The FMR curve can be done similarly, but using unregistered samples and templates that are enrolled by other samples. While these curves are suitable for performance evaluation, it will require a great deal of labor to improve their precision. For example, we designed a stored-value card system, and have reported that the EER of the system is 1.5×10^{-4} by operating a total of 360,000 times over 200 cards and 3 card terminals, and 600,000 times over 50 templates and 3 card terminals, for the FNMR and FMR respectively [5].

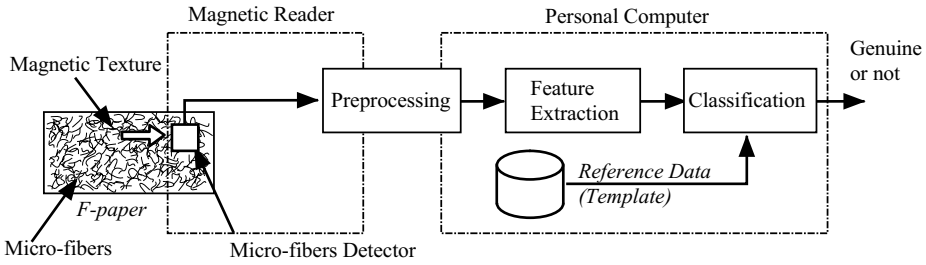

Fig. 2 The procedures in the magnetic *artifact-metric* system.

Even if we can accomplish such a troublesome work, there is yet another problem in the development stage of an *artifact-metric* system. It is often difficult for us to evaluate its performance precisely because we usually have a small number of artifact samples. Furthermore, if we acquire enough samples, it may still be difficult because of their lack of durability against many thousand times of operations. Hence, the following sections de-scribe how to evaluate the accuracy of authentication for the magnetic *artifact-metric* system using a small number of artifact samples to coordinate its algorithm.

4. Authentication

4.1 Authentication Procedures

The magnetic *artifact-metric* system involves authentication procedures shown in **Fig. 2**. The magnetic texture is captured, and then preprocessed by the system. The authentication procedures includes both a feature extraction procedure and a classification procedure.

4.1.1 Feature Extraction

From the magnetic reader, the PC receives a raw data,

$$\boldsymbol{r} = (r_1, r_2, \ldots, r_n)^t, \tag{1}$$

where r_i $(i = 1, 2, \ldots, n)$ represents i-th raw data.

The raw data will be averaged and compressed in order to remove glitches or rapid noises. By sequentially averaging every $a_0 \geq 1$ elements of the raw data $\boldsymbol{r}$, the PC compresses $\boldsymbol{r}$ into the compressed pattern,

$$\boldsymbol{c} = (c_1, c_2, \ldots, c_m)^t, \tag{2}$$

where c_j $(j = 1, 2, \ldots, m)$ represents the mean value of the block j, and is given by

$$c_j = \frac{1}{a_0} \sum_{i=(j-1)a_0+1}^{j \cdot a_0} \tag{3}$$

Finally, by extracting d $(1 \leq d \leq m)$ sequential elements of $\boldsymbol{c}$, we can obtain an intrinsic pattern,

$$\boldsymbol{P}_{d,k} = (c_k, c_{k+1}, \ldots, c_{k+d-1})^t, \tag{4}$$

where $1 \leq k \leq m$ and $k + d - 1 \leq m$.

4.1.2 Registration

A template $\hat{\boldsymbol{P}}_{d,r}$, where the subscript r indicates a reference point, i.e. $k = r$, can be created by capturing the intrinsic patterns from the same *F-paper* $M \geq 1$ times. In the magnetic *artifact-metric* system, the PC calculates the template $\hat{\boldsymbol{P}}_{d,r}$ as the average of the intrinsic patterns $\boldsymbol{P}^i_{d,r}$, where the subscript $i = 1, 2, \ldots, M$ indicates the multiple samples from the same *F-paper*. We define a mean value of the k-th elements of $\boldsymbol{P}^i_{d,r}$ as

$$p_k = \frac{1}{M} \sum_{i=1}^{M} c^i_k, \tag{5}$$

where $k = r, r + 1, \ldots, r + d - 1$.

Finally, we can write the template as

$$\hat{\boldsymbol{P}}_{d,r} = (p_r, p_{r+1}, \ldots, p_{r+d-1})^t. \tag{6}$$

4.1.3 Classification

The PC classifies an *F-paper* whether genuine or not by checking its intrinsic pattern in the subsequent authentication procedure to which we apply a pattern-matching scheme based on the correlation. Every time the PC examines an *F-paper*, a compressed pattern $\boldsymbol{c} = (c_1, c_2, \ldots, c_m)^t$ is captured, and then an intrinsic pattern $\boldsymbol{P}_{d,r} = (c_r, c_{r+1}, \ldots, c_{r+d-1})^t$, will be extracted from $\boldsymbol{c}$. Simultaneously, a template, $\hat{\boldsymbol{P}}_{d,r} = (p_r, p_{r+1}, \ldots, p_{r+d-1})^t$ at the corresponding reference point can be obtained from the templates which are previously recorded. If we define the degree of similarity between $\boldsymbol{P}_{d,r}$ and $\hat{\boldsymbol{P}}_{d,r}$ as $S(\boldsymbol{P}_{d,r}, \hat{\boldsymbol{P}}_{d,r})$, which can be calculated as follows:

$$S\left(\boldsymbol{P}_{d,r}, \hat{\boldsymbol{P}}_{d,r}\right) = \frac{\sum_{i=r}^{r+d-1} (c_i - c_r) \cdot (p_i - \bar{p})}{\sqrt{\sum_{i=r}^{r+d-1} (c_i - \bar{c}_r)^2 \sum_{i=r}^{r+d-1} (p_i - \bar{p})^2}}, \quad (7)$$

where $\bar{c}_r$ and $\bar{p}$ are mean values of all the elements of the patterns $\boldsymbol{P}_{d,r}$ and $\hat{\boldsymbol{P}}_{d,r}$, respectively. Actually, in the classification process, the intrinsic pattern is captured redundantly to compensate for position errors of the reference point. Every time the PC examines an *F-paper*, $(2s+1)$ redundant patterns $\boldsymbol{P}_{d,(r-s)}$, $\boldsymbol{P}_{d,(r-s+1)}, \ldots, \boldsymbol{P}_{d,r}, \ldots, \boldsymbol{P}_{d,(r+s-1)}, \boldsymbol{P}_{d,(r+s)}$, where $s \geq 0$ is the number of shifts, will be extracted from $\boldsymbol{c}$. The PC calculates the minimum value of $S(\boldsymbol{P}_{d,r}, \hat{\boldsymbol{P}}_{d,r})$ by

$$S_{\min}(\boldsymbol{P}_{d,r}, \hat{\boldsymbol{P}}_{d,r}) \stackrel{def}{=} \min_{-s \leq k \leq s} S(\boldsymbol{P}_{d,(r+k)}, \hat{\boldsymbol{P}}_{d,r}), \quad (8)$$

where the value of the reference point r is limited as $s+1 \leq r \leq m-d-s+1$.

Finally, the PC classifies the *F-paper* as acceptable, i.e., genuine, if $S_{\min}(\boldsymbol{P}_{d,r}, \hat{\boldsymbol{P}}_{d,r}) > \alpha$, otherwise not, according to a fixed threshold value, α.

5. Experimental Results

5.1 Experimental Conditions

Our experiments are conducted on the following conditions;

(1) The number of raw data per sampling is 11,400; $\boldsymbol{r} = (r_1, r_2, \ldots, r_{11,400})$.

(2) The number of raw data per averaging is 10; $a_0 = 10$.

(3) The number of compressed patterns per sampling is 1,140; $\boldsymbol{c} = (c_1, c_2, \ldots, c_{1140})$.

(4) The number of samples for creating a template is 3; $M = 3$.

(5) We used the database to query a decided template $\hat{\boldsymbol{P}}_{d,r}$, in each experiment. The number of patterns in the template database is 1,140; the database Ψ defined as

$$\Psi \stackrel{def}{=} \hat{\boldsymbol{P}}_{1140,1} = (p_1, p_2, \ldots, p_{1140})^t. \quad (9)$$

While 1,140 compressed patterns are extracted, raw data from the rear position of the *F-papers* are unstable in the repetitive verification because of positioning errors. Accordingly, the stable results, $\{\boldsymbol{P}_{d,r} | r \leq 400\}$, from the lead position of the *F-paper* are fixed for the measurements of the FNMRs.

(6) The number of shifts is 7; $s = 7$. We examined 100 sheets of *F-papers* to fix the number of shifts s for $d = 20$, 30 and 40. **Figure 3** shows the FNMR curves when we set the number of elements $d = 40$ as an example. In the graph, we show curves when we set the number of shifts s from 0 to 8. We found that error rates generally decrease with an increase in s, and, however, the curves are perfectly overlapped and undistinguishable among the cases of $s = 6$, 7 and 8 for every d. Accordingly, we set the number of shifts s as $s = 7$ in our examination.

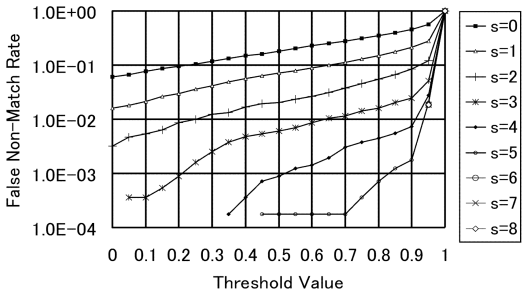

Fig. 3 The FNMR curves in the cases of $d = 40$, when changing the number of shifts from 0 to 8.

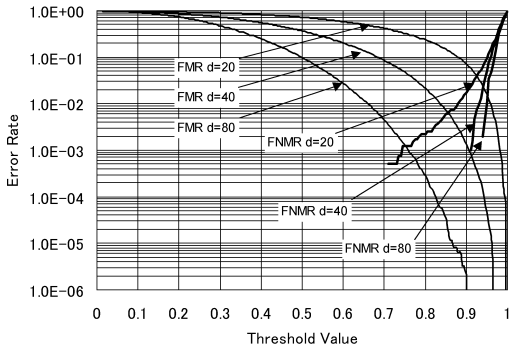

Fig. 4 The FNMR and FMR curves are measured for magnetic *artifact-metric* system when we set the number of elements as $d = 20$, 40 and 80.

5.2 The Experimental Results

We examined the accuracy of authentication for the magnetic *artifact-metric* system when changing the number of elements d, and the reference point r. The FNMR and FMR curves are shown in **Fig. 4** when we set the number of elements as $d = 20$, 40, and 80, which are correspond to around 3.5 mm, 7.0 mm, and 14.0 mm respectively. The FNMR curves are plotted the

Table 1 The number of samples for each curve.

number of elements	FNMR	FMR	Simulated FNMR
$d = 20$	4.0×10^3	2.5×10^6	2.9×10^5
$d = 40$	2.0×10^3	1.2×10^6	8.4×10^4
$d = 80$	1.0×10^3	6.1×10^5	7.0×10^4

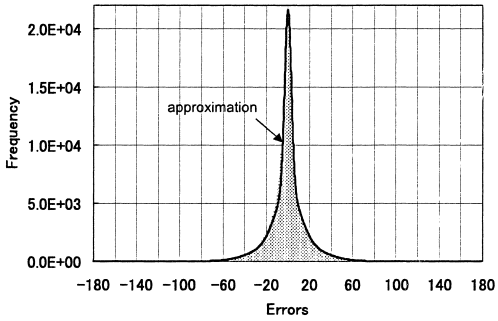

Fig. 5 The histogram of the errors in 114,000 raw data is presented, and can be suitably approximated by adding more than three normal distributions.

results of 200 times operation of the *F-paper*, which we used for extracting templates. Meanwhile, the FMR curves are plotted by mutually verifying the intrinsic patterns from 210 sheets of *F-papers*. The number of samples for each curve is shown in **Table 1**. It is quite obvious that the FMR curves are precise enough for the performance evaluation, but the FNMR curves are not.

5.3 Simulation Results

In order to make up the shortage of the data for the FNMR evaluation, we established a technique to extend the experimental results. We calculated errors of 114,000 raw data in a total of 100 repetitive verifications for an *F-paper*. In this calculation, we use not $\{\boldsymbol{P}_{d,r} | r \leq 400\}$ but $\{\boldsymbol{P}_{d,r} | r \leq 1140\}$ to estimate the least upper bound of the FNMR with the maximum positioning error. **Figure 5** shows the histogram of the errors in the raw data. We simulated a repetitive verification with the patterns which were produced by adding calculated noises according to this error distribution, and extended the FNMR curves as shown in **Fig. 6**. The number of samples for each simulated curve is also shown in Table 1. We can see from this graph that the extended FNMR curves are well simulated the least upper bounds of the FNMRs for the actual curves. The extended results are so precise that we can also plot them as the ROC curves, and shown in **Fig. 7**.

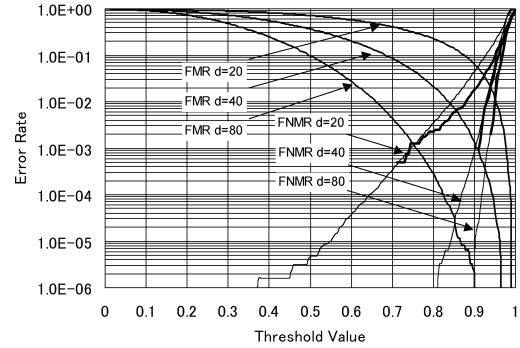

Fig. 6 The FMR curves are enhanced by using simulated data when we set the number of elements as $d = 20$, 40 and 80.

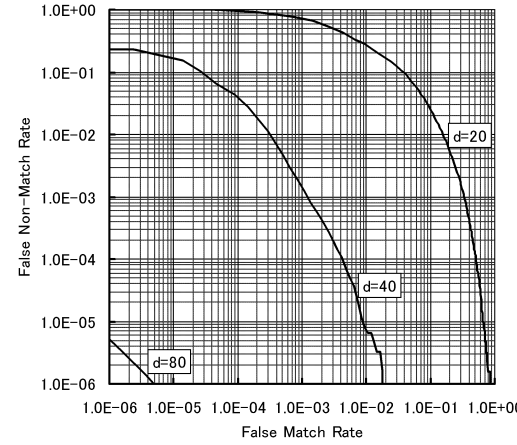

Fig. 7 The ROC curves are plotted for the magnetic *artifact-metric* system when we set the number of elements as $d = 20$, 40 and 80.

6. Examination

6.1 The Errors in Raw Data

We have found that the distribution of errors in raw data can be suitably approximated by adding more than three normal distributions. In Fig. 5, the thick solid line is an example of approximation by adding three normal distributions. This fact may indicate that the errors in raw data come from a distribution, of which probability density function consists of those with some normal distributions. It seems reasonable to suppose that the distribution of errors in raw data is caused by normal distributed errors such as errors in the texture itself, motion of the artifact, detection with sensors, and signal conversion. Additionally, the errors in motion of the artifact should include those in

its vertical/horizontal and rotational position.

6.2 The Accuracy of Authentication

It is clear in Fig. 6 and Fig. 7 that both the FNMRs and FMRs, are decreased, i.e. the accuracy of authentication increases, with an increase in the number of elements d which are used for verifying in the pattern matching process. We can see form Fig. 6 or Fig. 7 that the EERs of the magnetic *artifact-metric* system are 1.5×10^{-2}, 1.0×10^{-3} and 1.1×10^{-6}, when $d = 20$, 40 and 80, respectively. Although we can see from Fig. 4 that the EER is around 1.3×10^{-2} when $d = 20$, it is appropriate to estimate the ERR to be the former, i.e., 1.5×10^{-2}, as the least upper bound of the EER.

6.3 Protocols

6.3.1 A Retrying Protocol

To meet the requirements of a higher level of security, there are some protocols to enhance performance of the systems. When we apply a retrying or alternative-checking protocol, the PC retries once again only if the *F-paper* was not accepted at the first attempt. The EER is expected to be decreased with the retrying protocol [5]. If we apply this protocol, the FRR, FR_r, and the FAR, FA_r, will be respectively estimated as

$$FR_r = FR^2, \tag{10}$$

and

$$FA_r = 1 - (1 - FA)^2. \tag{11}$$

In Eqs. (10) and (11), FR and FA are the FNMR and FMR respectively of the system without a retrying protocol.

The experimental result, where we apply the retrying protocol to the classification setting the number of elements as $d = 20$, is shown as the thick solid lines in **Fig. 8**. The FRR and FAR of the system would be respectively estimated by Eqs. (10) and (11), and are indicated by the thin solid lines in Fig. 8. We see from the graph that there are differences between the experimental results and the estimated curves for the FRR and FAR. The following gives the reasons for being different, and a detailed account for the estimate of the FRR and FAR when applying the retrying protocol.

False Rejection Rate

We obtain a degree of similarity, $S_{\min}(\boldsymbol{P}_{d,r}, \hat{\boldsymbol{P}}_{d,r})$ whenever the PC verifies a pattern $\boldsymbol{P}_{d,r}$ with a template, $\hat{\boldsymbol{P}}_{d,r}$. Assuming that the random variable $\tau = S_{\min}(\boldsymbol{P}_{d,r}, \hat{\boldsymbol{P}}_{d,r})$ is obtained with a probability density function $f_s(\tau)$, the FNMR can be given by

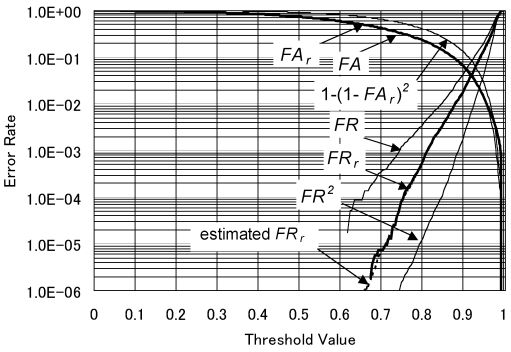

Fig. 8 The experimental results and the estimates, which we apply the retrying protocol to the classification setting the number of elements as $d = 20$ are shown.

$$FR = \int_{-1}^{\alpha} f_s(\tau) d\tau = F_s(\tau)|_{-1}^{\alpha}, \tag{12}$$

where α is a fixed threshold value, and

$$\int_{-1}^{1} f_s(\tau) d\tau = F_s(\tau)|_{-1}^{1} = 1. \tag{13}$$

In Eq. (13), the subscript s stands for "*self*." We define the FNMR of the system as the mean value of the FNMRs for K patterns $\boldsymbol{P}_{d,r}^k$ $(k = 1, 2, \ldots, K)$. If we know that the random variable τ comes from the pattern $\boldsymbol{P}_{d,r}^k$, we can write:

$$\begin{aligned} FR &= \frac{1}{K} \sum_{k=1}^{K} \int_{-1}^{\alpha} f_s(\tau | \boldsymbol{P}_{d,r}^k) d\tau \\ &= \frac{1}{K} \sum_{k=1}^{K} F_s(\tau | \boldsymbol{P}_{d,r}^k)|_{-1}^{\alpha}, \end{aligned} \tag{14}$$

where $f_s(\tau | \boldsymbol{P}_{d,r}^k)$ is the probability density function of τ for the given pattern $\boldsymbol{P}_{d,r}^k$.

Using Bayes theorem, the probability of $\boldsymbol{P}_{d,r}^k$ given τ, which is called the posterior probability of τ, can be given by

$$\Pr\left(\boldsymbol{P}_{d,r}^k | \tau\right) = \frac{\Pr(\boldsymbol{P}_{d,r}^k) f_s(\tau | \boldsymbol{P}_{d,r}^k)}{f_s(\tau)}, \tag{15}$$

where

$$\sum_{k=1}^{K} \Pr\left(\boldsymbol{P}_{d,r}^k | \tau\right) = 1. \tag{16}$$

In Eq. (15), $\Pr(\boldsymbol{P}_{d,r}^k)$ is the unconditional probability of $\boldsymbol{P}_{d,r}^k$, and

$$f_s(\tau) = \sum_{k=1}^{K} \Pr\left(\boldsymbol{P}_{d,r}^k\right) f_s\left(\tau | \boldsymbol{P}_{d,r}^k\right), \tag{17}$$

is the probability density function of τ. By Eqs. (12) and (15), the FNMR at the second attempt, which is posterior to τ, can be given by

$$FR_2 = \sum_{k=1}^{K} \Pr\left(\boldsymbol{P}_{d,r}^k|\tau\right)\left\{F_s(\tau|\boldsymbol{P}_{d,r}^k)|_{-1}^{\alpha}\right\}. \quad (18)$$

Finally, the FRR when applying the retrying protocol can be estimated from Eqs. (14) and (18), as

$$\begin{aligned} FR_r &= FR \cdot FR_2 \\ &= \frac{1}{K}\sum_{k=1}^{K}\left\{F_s(\tau|\boldsymbol{P}_{d,r}^k)|_{-1}^{\alpha}\right\} \\ &\quad \times \sum_{k=1}^{K}\Bigg[\Pr(\boldsymbol{P}_{d,r}^k|\tau) \\ &\qquad \times \left\{F_s(\tau|\boldsymbol{P}_{d,r}^k)|_{-1}^{\alpha}\right\}\Bigg] \end{aligned} \quad (19)$$

This gives us a reasonable estimate because the probability density functions, i.e. $f_s(\tau|\boldsymbol{P}_{d,r}^k)$, are highly diversified as shown in **Fig. 9**. By Eq. (19), we can reasonably calculate the FRR as the dotted line, which is indicated by "estimated FR_r" in Fig. 8.

False Acceptance Rate

Assuming that the random variable $\tau = S(\boldsymbol{P}_{d,r}, \hat{\boldsymbol{P}}_{d,r})$ is obtained with a probability density function $f_n(\tau)$, the FMR can be given by

$$\begin{aligned} FA &= \int_{\alpha}^{1} f_n(\tau)d\tau \\ &= 1 - \int_{-1}^{\alpha} f_n(\tau)d\tau \\ &= 1 - F_n(\tau)|_{-1}^{\alpha}, \end{aligned} \quad (20)$$

where α is a fixed threshold value, and

$$\int_{-1}^{1} f_n(\tau)d\tau = F_n(\tau)|_{-1}^{1} = 1. \quad (21)$$

In Eqs. (20) and (21), the subscript n stands for "*nonself*" to distinguish it from the probability density function $f_s(\tau)$ of which subscript stands for "*self*."

We generally define the FMR of the system as the mean value of the FMRs which are obtained from cross-verification of K patterns. In the same way as the calculation of the FRR, we reach

$$\begin{aligned} FA_r &= 1 - (1 - FA) \cdot (1 - FA_2) \\ &= 1 - \left[1 - \frac{1}{K}\sum_{k=1}^{K}\left\{F_n(\tau|\boldsymbol{P}_{d,r}^k)|_{-1}^{\alpha}\right\}\right] \\ &\quad \times \left[1 - \sum_{k=1}^{K}\Pr(\boldsymbol{P}_{d,r}^k|\tau) \right. \\ &\qquad \left. \times \left\{F_n(\tau|\boldsymbol{P}_{d,r}^k)|_{-1}^{\alpha}\right\}\right], \end{aligned} \quad (22)$$

where FA_2 is the FMR at the second attempt. Thus, by Eq. (22), we may theoretically estimate the FAR when we apply the retrying protocol, assuming that the same pattern was presented in the two attempts.

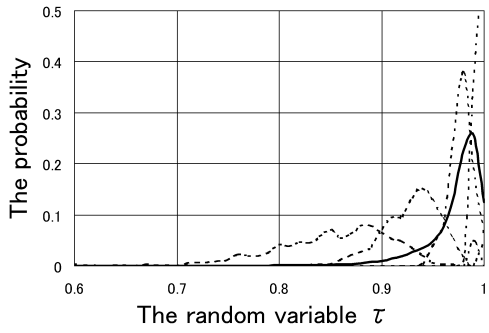

Fig. 9 Examples of the probability density function are highly diversified. The probability density function $f_n(\tau)$ is indicated by the solid line in this graph.

In spite of that, we encounter difficulties to apply this estimation method to practical evaluation because it requires hard work to find the probability density function $f_n(\tau|\boldsymbol{P}_{d,r}^k)$, and the probability $\Pr(\boldsymbol{P}_{d,r}^k|\tau)$. In **Fig. 10**, the solid line indicates the probability density function, $f_n(\tau)$. We give an example of the probability density function, $f_n(\tau|\boldsymbol{P}_{d,r}^k)$ as a dotted line in this figure. Based on our analysis, $f_n(\tau|\boldsymbol{P}_{d,r}^k)$ is almost the same as $f_n(\tau)$. Therefore, there is no distinguishable change in the probability. Accordingly, we presume the FAR to be no different as from the first attempt. Put another way, the random variable τ, which will occur at the second attempt, will occur with the same probability as the first attempt, i.e. $FA_2 = FA$, assuming that the patterns presented in the two attempts are different from each other. On the other hand, the experiment results were obtained when we assumed that the same pattern was presented in the two attempts. Therefore the random variable τ, which will occur at the second attempt, will not

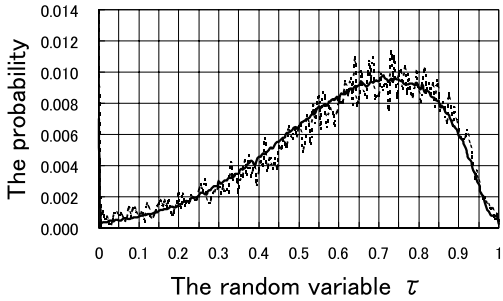

Fig. 10 An example of the probability density function $f_n(\tau|\boldsymbol{P}^k_{d,r})$ is nearly the same as the probability density function $f_n(\tau)$ indicated by the solid line.

occur with the same probability as the first attempt, but with the probability corresponding to the histogram in Fig. 5. Finally, we can estimate the FAR of the system almost as it is, provided that the same pattern was presented in the two attempts. Consequently, the FAR when applying the retrying protocol, is estimated and indicated by FA_r in Fig. 8.

6.3.2 A Double-check Protocol

The double-check protocol, where the PC checks an *F-paper* by verifying the two intrinsic patterns $\boldsymbol{P}_{n,i}$ and $\boldsymbol{P}_{n,j}$ $(i \neq j)$, and then judges the *F-paper* to be acceptable if both patterns are acceptable, i.e., $S_{\min}(\boldsymbol{P}_{n,i}, \hat{\boldsymbol{P}}_{n,i}) > \alpha$ and $S_{\min}(\boldsymbol{P}_{n,j}, \hat{\boldsymbol{P}}_{n,j}) > \alpha$, will also decrease the EER. We can estimate the performance in the similar way as the retrying protocol. The FRR and the FAR of this double-check protocol, will be respectively estimated as

$$FR_d = 1 - (1 - FR)^2, \tag{23}$$

and

$$FA_d = FA^2. \tag{24}$$

Figure 11 shows the results when we applied the double-check protocol to the authentication procedure. In the graph, the curved lines indicated by $d = n \times 2$ $(n = 20,\ 40)$ show the experimental results. Similarly, the curved line indicated by $d = 20 \times 4$, shows the experimental results, where the PC checks an *F-paper* by verifying the four intrinsic patterns, and then judges the *F-paper* to be acceptable if every pattern is acceptable. We can estimate the performance in a similar way as Eqs. (23) and (24). The FRR and FAR of this system, when we applied this fourfold-check protocol to the authentication procedure, can be respectively estimated as

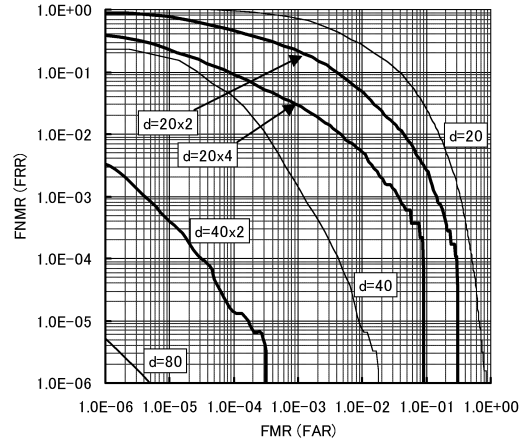

Fig. 11 The ROC curves are plotted for the magnetic *artifact-metric* system when we set the number of elements as $d = 20$, 40 and 80. The results to which we apply the multi-check protocols are also plotted.

$$FR_f = 1 - (1 - FR)^4, \tag{25}$$

and

$$FA_f = FA^4 \tag{26}$$

In Fig. 11, we did not plot the ROC curves, which can be theoretically calculated by Eqs. (23) and (24), or by Eqs. (25) and (26), because the result of the experiment gave good agreement with the value that had been obtained by theoretical calculation, and thus they overlapped with the experimental curves. This fact indicates that we would be able to expansively estimate the performance the FRR and FAR of a multi-check system, respectively, as

$$FR_{\text{multi}} = 1 - (1 - FR)^m, \tag{27}$$

and

$$FA_{\text{multi}} = FA^m, \tag{28}$$

where m is the number of multiple checks. We can see from Fig. 11, the FARs of continuous patterns, i.e. $d = 40$ and $d = 80$, are lower than those of divided patterns, $d = 20 \times 2$, and $d = 20 \times 4$ or $d = 20 \times 2$, respectively, even if the total number of elements is the same.

7. Conclusions

In this paper, we demonstrated a practical measure for performance evaluation of *artifact-metric* system. A case study clarified the accuracy of authentication of the magnetic *artifact-metric* system with such indicators as the FNMR and FMR, and the ROC curves. Based on measurements for a small number of paper documents, we calculated extend the FNMR curves. We found that the extended

FNMR curves are well simulated the least upper bounds of the FNMRs. We also demonstrated that an increase in the number of elements decreases the EER. In addition, we demonstrated and analyzed enhancement by a retrying protocol or a multi-check protocol. It was found through the examination that the enhanced curves are reasonably calculated by taking account of probability density for the degree of similarity arising from each pattern.

While we detailed how to evaluate the accuracy of authentication of *artifact-metric* systems assuming that no attacker or *clone* exists, the FMR where *clones* do exist will be higher than the original one. From this point we might go on to a security examination against cloning, since the primary consideration in evaluation of authentication system should be given into its security.

Acknowledgments This research was partially supported by MEXT Grant-in-Aid for Scientific Research on Priority Areas 13224040 (Tsutomu Matsumoto).

References

1) ANSI A9.84-2001, Biometrics Information Management and Security (2001).
2) Matsumoto, H., Suzuki, K. and Matsumoto, T.: A clone preventive authentication technique which features magnetic micro-fibers and cryptography, *Proc. SPIE*, Vol.3314, pp.275–286 (1998).
3) Matsumoto, H., Yamamotoya, K. and Matsumoto, T.: Document Protection by Micro-Fibers and Cryptography, *Proc. PISEC '99*, Barcelona, Spain (1999).
4) Matsumoto, H. and Matsumoto, T.: Artifact-metric systems, Technical Report of IEICE, ISEC2000-59, October 2000, pp.7–14 (2000).
5) Matsumoto, H., Takeuchi, I., Hoshino, H., Sugahara, T. and Matsumoto, T.: An Artifact-metric System Which Utilizes Inherent Texture, *IPSJ Journal*, Vol.42, No.8, pp.139–152 (2001).
6) Wayman, J.L.: Technical Testing and Evaluation of Biometric Identification Device, *Biometrics: Personal Identification in Networked Society*, The Kluwer Academic, International Series in Engineering and Computer Science, Jain, A.K., Bolle, R. and Pankanti, S. (Eds.), Vol.479, Chapter 17, pp.345–368 (1999).

(Received November 30, 2001)
(Accepted June 4, 2002)

Hiroyuki Matsumoto received the B.E. degree in mechanical engineering from the University of Electro-Communications, Tokyo, Japan, and joined NHK Spring Co., Ltd., Yokohama, Japan, in 1982. He had been engaged in pattern recognition systems, and received the M.E. degree in electronic information engineering from Toyota Technological Institute, Nagoya, Japan, in 1987. From 1999 to 2002, he was a Ph.D. student of Yokohama National University, Yokohama, Japan, and received the Ph.D. degree in artificial environment and systems from Yokohama National University in 2002. He is a manager of the Information & Security Systems Division of NHK Spring Co., Ltd. His research interests include document security and biometrics.

Tsutomu Matsumoto was born in Maebashi, Japan, on October 20, 1958. He received the Dr.Eng. Degree from the University of Tokyo in 1986 and since then his base has been in Yokohama National University where he is enjoying research and teaching in the field of cryptography and information security as a Professor in Graduate School of Environment and Information Sciences. He is a member of Cryptography Research and Evaluation Committee of Japan. He served as the general chair of ASIACRYPT 2000. He is an associate editor of Journal of Computer Security and is on the board of International Association for Cryptologic Research. He is a member of IEICE Technical Group on Information Security and of IPSJ Special Interest Group on Computer Security. He received Achievement Award from the IEICE in 1996.