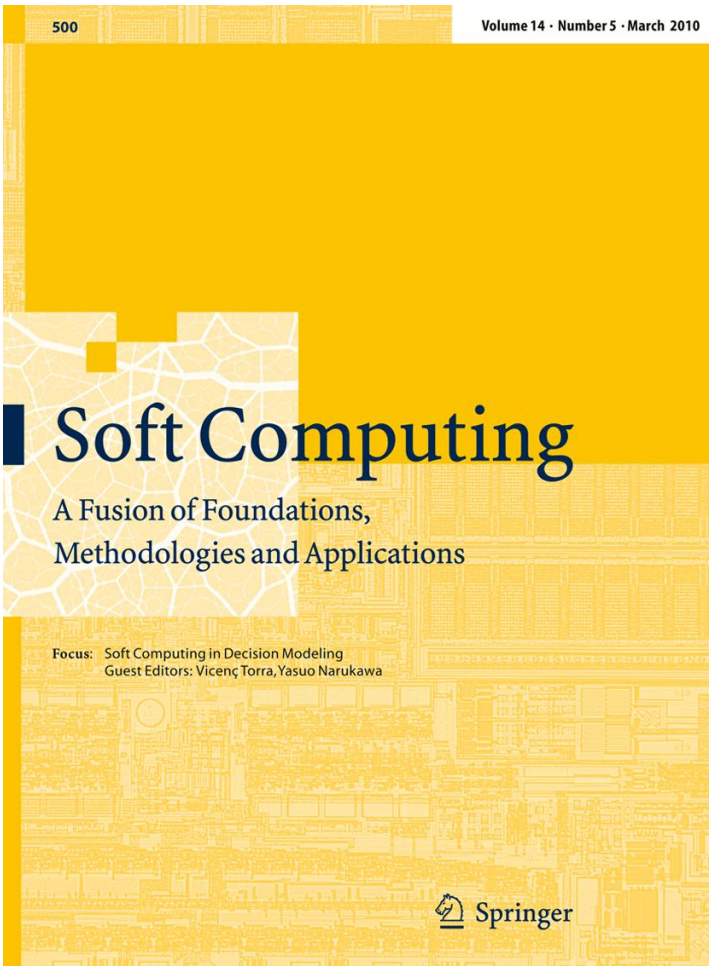
ISSN 1432-7643, Volume 14, Number 5



Soft Comput (2010) 14:529–536
DOI 10.1007/s00500-009-0449-6

FOCUS

Privacy-preserving similarity evaluation and application to remote biometrics authentication

Hiroaki Kikuchi · Kei Nagai · Wakaha Ogata · Masakatsu Nishigaki

Published online: 4 June 2009

Abstract In this paper, a new method for secure remote biometric authentication preventing the vulnerability of compromised biometrics is presented. The idea is based on a public-key cryptographical protocol, referred as *zero-knowledge proof*, which allows a user to prove that she has surely a valid biometric data without revealing the data. Hence, the scheme is free from the risk of disclosure of biometric data. Even if a malicious administrator has a privilege access to the private database, it is infeasible for him to learn the private template. This paper studies two well-known definitions, the *cosine correlation* and the *Euclidean distance* as similarities of given two feature vectors. Both similarities are defined with some multiplications and additions, which can be performed in privacy-preserving way because of the useful property of public-key commitment scheme, *additive homomorphism*. The estimation based on the experimental implementation shows that the private Euclidean distance scheme archives better accuracy in terms of false acceptance and rejection than the private cosine coloration scheme, but it requires about $5/2n\ell$ overhead to evaluate n-dimension feature vectors consisting of ℓ-bit integers.

H. Kikuchi (✉) · K. Nagai
Department of Communication and Network Engineering,
School of Information and Telecommunication Engineering,
Tokai University, 1117 Kitakaname, Hiratsuka,
Kangawa 259-1292, Japan
e-mail: kikn@tokai.ac.jp

W. Ogata
Graduate School of Innovation Management,
Tokyo Institute of Technology, Tokyo, Japan
e-mail: wakaha@mot.titech.ac.jp

M. Nishigaki
Graduate School of Science and Technology,
Shizuoka University, Shizuoka, Japan
e-mail: nishigaki@inf.shizuoka.ac.jp

1 Introduction

Biometrics identifiers are now commonly used to identify individuals in more secure and more efficient ways than the conventional password-based methods. Typically, the biometric identifiers including fingerprint, vein, iris, facial images are scanned and processed in appropriate algorithm to extract a *feature vector*, which is stored as a *template* in registration (Maltoni et al. 2003). In authentication, a feature vector extracted from a newly scanned image will be compared to the template to verify that the owner of the biometric data is legitimate or not.

The biometric recognition, however, is mostly made in *local* environment, e.g., a matching with the template data stored in secure smartcard (in ATM cards), or a user authentication at personal laptop PCs. The reason of limitation in local is the known vulnerabilities of *remote biometric authentication* that once a biometric template is stolen, it is stolen forever and can not be recovered. If we store our biometric data to some service provider, we immediately face risks that the server may be compromised, or a malicious administrator of the server may learn our highly private data and can disclose it.

Many researchers pointed out the issue in remote biometrics authentication and several attempts addressing it have been made. Ratha et al. (2001) proposed a "cancelable biometrics", using a morphing technique to transform biometric data into a randomized form, which depends on given morphing function. Jeong et al. (2006) proposed a changeable biometrics for face recognition using the

principal component analysis (PCA) and the independent component analysis (ICA). Given two vectors chosen from PCA and ICA coefficients, they extract from an input face image the transformed vector according to a scrambling rule. When the transformed template is compromised, the scrambling rule is replaced by a new one. Juels and Sudan's "fuzzy vault scheme" (Juels and Sudan 2002) is an improvement upon the previous work by Juels and Wattenberg (1999). In Juels and Sudan (2002), they use the polynomial reconstruction problem based on an error-collection code such as the Reed-Solomon. Clancy et al. (Clancy and Kiyavash 2003) proposed a "fingerprint vault system" based on the fuzzy vault. Using multiple minutiae location sets, they use canonical positions of minutiae, as the elements of a set. Uludag and Jain (2004) proposed a fuzzy vault system for fingerprint using the Lagrange interpolation and the cyclic redundancy check (CRC) for testing polynomial reconstruction instead of the error-collection step.

Many cryptographical primitives are introduced to construct secure protocols. Bringer et al. proposed a scheme using a group signature in Julien et al. (2008). In Julien et al. (2007), an extended private information retrieval is used and then improved in Julien and Hervé (2008), too. Socek et al. used a set intersection as the degree of similarity in Daniel and Vladimir (2008). Barbosa et al. proposed a hybrid approach based on the support vector machine classifier and the Paillier public key encryption in Manuel et al. (2008).

Studies on security model of remote biometric authentications are made in Qiang et al. (2008). In Qiang et al. (2008), Tang et al. proposed a new formal security model for biometric-based remote authentication schemes so that several privacy concerns can be covered in it. Une et al. proposes a measure for evaluating schemes, called "wolf attack probability" in Masashi et al. (2007).

In this paper, we present a new method for secure remote biometric authentication preventing the vulnerability of compromised biometrics. Our idea is based on a public-key cryptographical protocol, referred as *zero-knowledge proof*, which allows a user to prove that she has surely a valid biometric data without revealing the data. Hence, the scheme is free from the risk of disclosure of biometric data. Even if the administrator with privilege access to the private database is malicious, it is infeasible to learn the private template. Without learning the template stored at the server, he performs an evaluation of similarities between the template and the new input in privacy-preserving way.

The zero-knowledge proof is generally "expensive" in terms of communication and computation costs. The performance of schemes depends on what similarity measure is used for the secret evaluation. In this paper, we study two well-known definitions, the *cosine correlation* and the *Euclidean distance* as similarities of given two feature vectors. Both similarities are defined with some multiplications and some additions, which can be performed in privacy-preserving way because of the useful property of public-key commitment scheme, *additive homomorphic*. The estimation based on the experimental implementation shows that the private Euclidean distance scheme achieves better accuracy in terms of false acceptance and rejection than the private cosine correlation scheme, but it requires about $5/2n\ell$ overhead to evaluate n-dimension feature vectors consisting of ℓ-bit integers.

The remainder of this paper is organized as follows. After giving the definitions for some fundamental building blocks, e.g., similarities and commitment functions, in Sect. 2, we construct two protocols for secure private similarity evaluation in Sect. 3. In Sect. 4, we evaluate our two proposed protocols from several viewpoints, including accuracy, performance and security. Section 5 concludes our study on secure remote biometric authentication.

2 Preliminaries

2.1 Similarities

Let $\boldsymbol{a} = (a_1, \ldots, a_n)$ and $\boldsymbol{b} = (b_1, \ldots, b_n)$ be n-dimensional vectors of R^n. We consider the following two well-known similarities between $\boldsymbol{a}$ and $\boldsymbol{b}$, which will be evaluated in privacy-preserving way in a later section.

Definition 1 A cosine correlation is a similarity between $\boldsymbol{a}$ and $\boldsymbol{b}$ defined as

$$\cos(\boldsymbol{a}, \boldsymbol{b}) = \frac{\boldsymbol{a} \cdot \boldsymbol{b}}{||\boldsymbol{a}|| \cdot ||\boldsymbol{b}||} = \frac{a_1 b_1 + \cdots + a_n b_n}{\sqrt{a_1^2 + \cdots + a_n^2}\sqrt{b_1^2 + \cdots + b_n^2}}$$

where $||\boldsymbol{a}||$ is a norm of $\boldsymbol{a}$.

Definition 2 An Euclidean distance, $d(\boldsymbol{a}, \boldsymbol{b})$, is defined as

$$d(\boldsymbol{a}, \boldsymbol{b}) = ||\boldsymbol{a} - \boldsymbol{b}|| = \sqrt{\sum_{i}^{n} (a_i - b_i)^2}.$$

For simplification, taking the normalization of $\boldsymbol{a}$ and $\boldsymbol{b}$, we can reduce the computational cost of cosine correlation as $cos(\boldsymbol{a}/||\boldsymbol{a}||, \boldsymbol{b}/||\boldsymbol{b}||) = \boldsymbol{a} \cdot \boldsymbol{b}$. Taking squared as $d(\boldsymbol{a}, \boldsymbol{b})^2$, we can omit the computation of square root for Euclidean similarity.

2.2 Secure commitment

A *commitment* is a cryptographical primitive to commit to a value while keeping it hidden and then reveal the committed value later.

A function $E(m, r)$ is considered as secure commitment to message m, where r is a random number, if it satisfies

1. No information reveals from $E(m, r)$, and
2. No one finds $m' \neq m, r$ and r' such that $E(m, r) = E(m', r')$.

Fujisaki and Okamoto proposed (Fujisaki and Okamoto 1997) a probabilistic commitment scheme based on the integer factorization problem as follows.

Definition 3 Let n be a composite number that no one knows the factors, and g and h be elements of Z_N such that $\log_g h$ is not known by anybody. A commitment to m is

$$E(m, r) = g^m h^r \bmod N,$$

where r is a random number.

The Fujisaki–Okamoto commitment has an *additive homomorphism*, a useful property for privacy-preserving computation, satisfying

$$E(m, r) \times E(m', r') = E(m + m', r + r') \text{ and } E(m, r)^x = E(mx, rx),$$

where the addition $m + m'$ is an ordinary arithmetic (not modular arithmetic) since we do not know the order of g and h.

We often write $E(m)$ to mean $E(m, r)$ when we do not necessary specify the random value.

2.3 Zero-knowledge proof of commitment

We introduce a cryptographical protocol for proving that a committed value m lies in a specific interval $[a, b]$ without revealing m, often known as Boudot's Range Proof (Boudot 2000).

Definition 4 Let F be a commitment $E(m, r)$ to message m. A range proof of knowledge of commitment is a cryptographical protocol allowing a prover to show that committed m is in $[a, b]$ without revealing m to a verifier, denoted by

$$PK\{m, r \mid F = E(m, r) \wedge m \in [a, b]\}$$

where r is uniformly chosen over $[-2^s N + 1, 2^s N - 1]$ and s is a security parameter [e.g., $s = 160$ (bit)].

The range proof takes about five times of overhead of a standard zero-knowledge proof of the committed value $PK\{m \mid F = E(m, r)\}$. Namely, it is expensive in terms of both computation and communication.

3 Private similarity evaluations

3.1 Overview and assumption

In our model, Alice is a user who tries to prove her identity to a server. Bob is the server who authenticates Alice based on the data that Alice has already registered. Assume that Alice does not fully trust Bob, that is, in so-called "*honest-but-curious*" model, where all players follow the protocol *honestly*, but are *curious* in that they try to find out as much as possible about the other input. In other words, no party can intentionally forget knowledge that it learns during the protocol. Therefore, instead of her private biometric data $\boldsymbol{x} = (x_1, \ldots, x_n)$, Alice registers the commitment to $\boldsymbol{x}$, $E(\boldsymbol{x})$, from which Bob cannot learn $\boldsymbol{x}$. To authenticate her to Bob, Alice scans her fresh biometric data $\boldsymbol{y} = (y_1, \ldots, y_n)$ and proves $\boldsymbol{x} \approx \boldsymbol{y}$ to Bob without revealing $\boldsymbol{y}$ (nor $\boldsymbol{x}$) in the zero-knowledge proof of similarities between $\boldsymbol{x}$ and $\boldsymbol{y}$.

There are many efficient protocols for proving several kinds of equalities in zero-knowledge way, and we need to prove privately that $\boldsymbol{y}$ is "close" to $\boldsymbol{x}$. It is not so hard to implement the fuzzy matching if Alice is allowed to access her tamper-proof device to recover $\boldsymbol{x}$ to be compared with new one $\boldsymbol{y}$. In the next section, we will show that the state-of-the-art cryptographical protocols allow us to evaluate similarities between any given committed vectors and to show the difference is within a range, without disclosing private biometric data to anyone. Hence, the protocol is free from the risk of private information disclosure.

3.2 Private cosine correlation evaluation

We show a protocol for secure evaluation of a cosine correlation given $\boldsymbol{x}$ and $\boldsymbol{y}$ in Fig. 1.

First of all, Alice needs to compute the commitment to her true private input $\boldsymbol{x}$ using random values $r_1, \ldots, r_n$ chosen uniformly over Z_N, as $E_i = E(x_i/c, r_i)$ for $i = 1, \ldots, n$. For reducing computational cost, we use the norm $c = ||\boldsymbol{x}||$ to normalize the committed input x_i. The random values are used for making the commitment indistinguishable against Bob in a sense that he can not distinguish two messages with non-negligible probability.

The key idea of the protocol is to evaluate the cosine correlation between template data $\boldsymbol{x}$ and an input data $\boldsymbol{y}$ without revealing private $\boldsymbol{x}$ and $\boldsymbol{y}$. The additive homomorphic property of the commitment scheme allows Bob to compute the commitment of the cosine correlation between hidden $\boldsymbol{x}$ and $\boldsymbol{y}$ at the third step as follows:

$$D_C = \prod_{i=1}^{n} G_{i=1}^{n} = \prod_{i=1}^{n} E(x_i/c, r_i)^{y_i/c'} = \prod_{i=1}^{n} E(x_i y_i/cc', r_i y_i/c')$$
$$= E\left(\frac{1}{||\boldsymbol{x}||||\boldsymbol{y}||}\sum_{i=1}^{n} x_i y_i, \sum_{i=1}^{n} r_i y_i/c'\right) = E(\cos(\boldsymbol{x}, \boldsymbol{y}), R_C)$$

where R_C is a random element computed as $\sum_{i=1}^{n} r_i y_i/c'$. Since Alice is allowed to access the tamper-proof device to obtain random values used to commit $\boldsymbol{x}$, she is able to learn R_C, and thereby get d_C. She also needs to prove to Bob that

Fig. 1 Protocol for cosine correlation evaluation

Protocol Private-Cosine
Input: $\boldsymbol{x} = (x_1, \ldots, x_n)$ and $\boldsymbol{y} = (y_1, \ldots, y_n) \in Z_N^n$.

1. (Registration) Alice sends to Bob the commitment to her private input $\boldsymbol{x}$, $E_1, \ldots, E_n$, where E_i is defined $E(x_i/c, r_i) = g^{x_i/c_x} h^{r_i}$ with random $r_i \in Z_N$ for $i = 1, \ldots, n$, and c is the norm of $\boldsymbol{x}$, i.e., $c = ||\boldsymbol{x}||$.
2. (Authentication) Alice computes a commitment to her scanned input $\boldsymbol{y} = (y_1, \ldots, y_n)$, $G_1, \ldots, G_n$ such that $G_i = E_i^{y_i/c'}$ for $i = 1, \ldots, n$, where $c' = ||\boldsymbol{y}||$. She proves to Bob that she knows the committed input $\boldsymbol{y}$ in the zero-knowledge proof
$$PK_1 = PK\left\{ y_i/c' \mid G_i = E_i^{y_i/c'} \right\}$$
for $i = 1, \ldots, n$.
3. Bob verifies PK_1 for all i and then computes $D_C = \prod_i^n G_i$ if PK_1 is valid.
4. Alice computes a similarity $d_C = cos(\boldsymbol{x}, \boldsymbol{y})$ and proves to Bob that she knows $\boldsymbol{y}$ that nearly equals to $\boldsymbol{x}$ in the zero-knowledge proof
$$PK_2 = PK\left\{ d_C, R_C \mid D_C = E(d_C, R_C) \wedge d_C \in [\tau_1, 1] \right\},$$
where $R_C = \sum_{i=1}^n r_i y_i / c'$.
5. Bob authenticates Alice if he verifies PK_2.

the commitment G_i has been correctly computed as defined formula without revealing y_i in PK_1.

At the end of the protocol, using the Boudot's range proof (Boudot 2000) and the conjunctive proof of knowledge (Cramer et al. 1994) (PK_2), she can finally convince Bob that she has valid input $\boldsymbol{y}$ such that the similarities $d_C = cos(\boldsymbol{x}, \boldsymbol{y})$ is greater than pre-determined threshold τ_1, which means that Alice is surely a legitimate user.

3.3 Private Euclidean distance evaluation

Figure 2 shows the protocol private-Euclid for proving Alice's private identity $\boldsymbol{y}$ is within the distance τ_2 from registered $\boldsymbol{x}$. In addition to the protocol private-cosine, it requires Alice to commit to not only $\boldsymbol{x}$ but also to squared $\boldsymbol{x}$ as $\tilde{E}_i$ at the registration step. Implicitly, we use notation X for the commitment to x, and $\tilde{X}$ for the commitment to x^2 in the figure.

The additive homomorphic property allows us to privately evaluate the Euclidean distance between $\boldsymbol{x}$ and $\boldsymbol{y}$ at side of Bob, at Step 3, as follows:

$$\begin{aligned} D_E &= \prod_{i=1}^n \tilde{E}_i \tilde{F}_i / G_i^2 = \prod_{i=1}^n E(x_i^2, r_i x_i) E(y_i^2, \tilde{r}_i') / E(2x_i y_i, 2r_i y_i) \\ &= \prod_{i=1}^n E(x_i^2 + y_i^2 - 2x_i y_i, r_i x_i + \tilde{r}_i' - 2r_i y_i) \\ &= E\left(\sum_{i=1}^n x_i^2 - 2x_i y_i + y_i^2, R_E \right) = E(||\boldsymbol{x} - \boldsymbol{y}||^2, R_E), \end{aligned}$$

letting R_E be a constant defined as $\sum_{i=1}^n r_i x_i - 2r_i y_i + \tilde{r}_i'$. For constructing zero-knowledge protocols PK_3, PK_4 and PK_5, we add a protocol proving that a committed number is a squared number, presented in Boudot (2000). If all proofs are valid, Bob is convinced that Alice is a legitimate user who has registered $\boldsymbol{x}$ and hence is able to show the correctly computed commitment of $||\boldsymbol{x} - \boldsymbol{y}||^2$ less than threshold τ_2.

4 Evaluation

Most zero-knowledge protocols are designed to be secure in the cost of communicational and computational overhead, which are not often considered as significant. There is a trade-off between performance and security, e.g., reducing a probability being impersonated by half requires double amount of bits to be computed. In addition, we claim that there is one more trade-off between accuracy and performance in secure biometric authentication. The accuracy (and the performance) depends on a function for similarity to be evaluated in zero-knowledge protocol. Hence, it is not trivial to identify the optimal function of similarity for the multiple objective requirements involved each other.

4.1 Feature vector

To compare two similarities, we performed some experiments using actual fingerprint images under the environment listed in Table 1. More than 500 live fingerprints are scanned and performed some sorts of image processing and extraction algorithms, which yield the feature vectors, called *ridge-valley orientation*.

The feature vector consists of a 18×18 matrix of orientations of ridges and valleys of the surface of finger,

Fig. 2 Protocol for Euclidean distance evaluation

Protocol Private-Euclid
Input: $\boldsymbol{x} = (x_1, \ldots, x_n)$ and $\boldsymbol{y} = (y_1, \ldots, y_n) \in Z_N^n$.

1. (Registration) Alice sends to Bob two sequences of commitments, $E_1, \ldots, E_n$, and $\tilde{E}_1, \ldots, \tilde{E}_n$, where
$$E_i = E(x_i, r_i), \qquad \tilde{E}_i = E(x_i^2, \tilde{r}_i),$$
r_i is a random value chosen from Z_N, and $\tilde{r}_i = r_i x_i$, for $i = 1, \ldots, n$. Alice proves to Bob that she knows the corresponding input $\boldsymbol{x}$, which is committed by E_i, and $\tilde{E}_i$, which is the commitment to x_i squared, in zero-knowledge proof
$$PK_3 = PK\left\{x_i, r_i \mid E_i = E(x_i, r_i) \wedge \tilde{E}_i = E_i^{x_i}\right\}.$$
Bob adds the commitments to template database if PK_3 is valid.
2. (Authentication) Alice computes three sequences of commitments with respect to her scanned input $\boldsymbol{y} = (y_1, \ldots, y_n)$,
$$F_i = E(y_i, r_i'), \quad \tilde{F}_i = E(y_i^2, \tilde{r}_i'), \quad \text{and } G_i = E_i^{y_i},$$
for $i = 1, \ldots, n$, where, r_i' and $\tilde{r}_i'$ are random values. She proves to Bob that the commitments have been properly computed in the zero-knowledge proof
$$PK_4 = PK\left\{y_i, r_i' \mid F_i = E(y_i, r_i'), \wedge \tilde{F}_i = F_i^{y_i} \wedge G_i = E_i^{y_i}\right\}$$
for $i = 1, \ldots, n$.
3. Bob computes $D_E = \prod_{i=1}^n \tilde{E}_i \tilde{F}_i / G_i^2$ if PK_4 is valid for all i.
4. Alice computes the Euclidean distance $d_E = d(\boldsymbol{x}, \boldsymbol{y})$ and proves to Bob that she knows $\boldsymbol{y}$ that nearly equals to $\boldsymbol{x}$ in a sense of Euclidean distance, using the zero-knowledge proof
$$PK_5 = PK\left\{d_E, R_E \mid D_E = E(d_E, R_E) \wedge d_E \in [0, \tau_2]\right\},$$
where $R_E = \sum_{i=1}^n r_i x_i - 2 r_i y_i + \tilde{r}_i'$.
5. Bob authenticates Alice if he verifies PK_5.

Table 1 Experiment environment

Item	Values
Fingerprint scanner	Digital Persona U. are .U4000
	Digital Persona Gold SDK 2.5.0
Fingerprint images	50 genuine and 450 imposter images
Resolution	300 × 300 (pixel)
Image processing	NIST NFIS2 (NIST FINGERPRINT IMAGE SOFTWARE 2 (NFIS2). http://fingerprint.nist.gov/NFIS/)
Software	Proprietary application with Java version 1.5.0_06,
Platform	Windows XP, 1.00 GHz, 512 MB

taking average for each local 16×16-pixel image. The ridge-valley orientation is quite stable against a transformation of images, thus good for the evaluation of similarities of high-dimension vectors. While, it needs to deal with empty portions of image caused by miss-scanning. To avoid some elements of feature from being zero, we take $n = L^2$ elements from the core of the 18×18 matrix. The accuracy of authentication depends on dimension n of the feature, and hence the optimal dimension is a significant issue. Figure 3 shows the variance of similarities (Euclidean distance) of two fingerprint images with respect to dimensions $n = 2 \times 2, 4 \times 4, \ldots, 18 \times 18$. From the observation of the result, we see that $n > 10 \times 10$ provides a good enough similarities to distinguish two images.

Figure 4 shows two distributions of cosine correlations; one between genuine and imposter images (labeled as "Imposter"), and the other one between two distinct images chosen from genuine images (as "Genuine"). The dimension of feature vector is $n = 8^2$. The genuine images are distributed within a narrow area of range, while the distribution of imposter images is broad. These distributions look almost disjoint, that is, the classification hardly ever fails.

The Euclidean distances of two feature vectors are distributed as well, shown in Fig. 5. In comparison of two similarities, the distribution of genuine images is quite separate from that of imposter images in the Euclidean distance, while these are distributed closely in cosine correlations. Therefore, the accuracy of Euclidean distance is likely to be better than that of the cosine correlation.

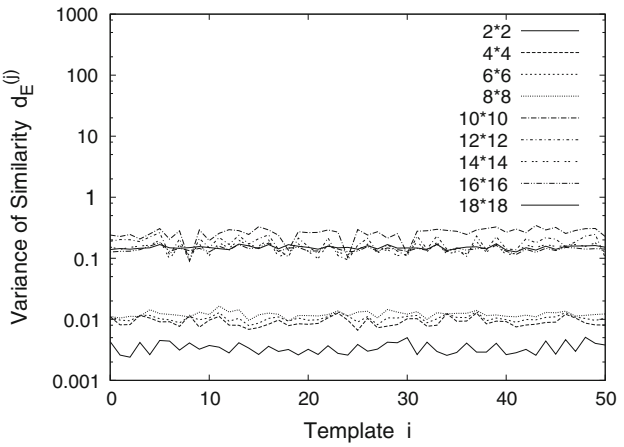


Fig. 3 Variance of similarity d_E

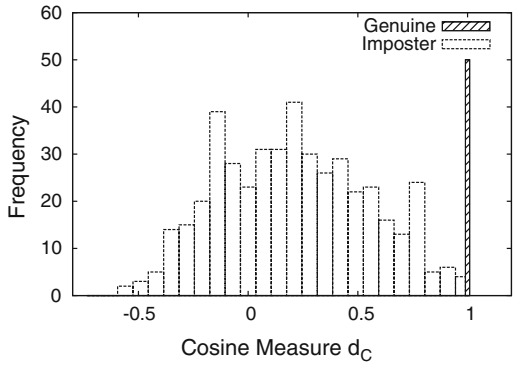


Fig. 4 Histogram of Cosine correlations $\cos(\boldsymbol{a}, \boldsymbol{b})$

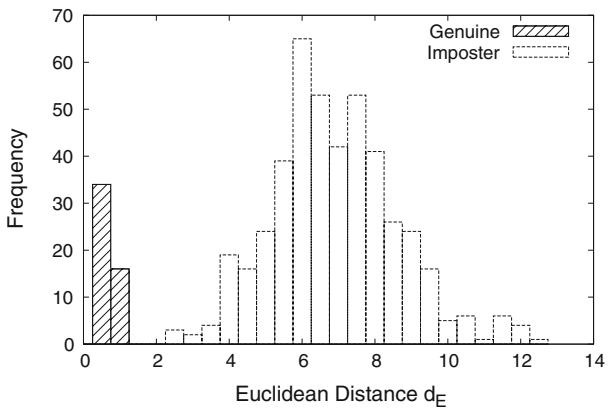


Fig. 5 Histogram of Euclidean distances $d(\boldsymbol{a}, \boldsymbol{b})$

4.2 Accuracy

We show the accuracy of authentication schemes based on the similarities in Fig. 6, where overall accuracy is given as equal error rate (ERR) of thresholds τ_1 and τ_2 with respects to n, the dimensions of feature vectors. An ERR is the rate at which both accept and reject errors are equal. Obviously,

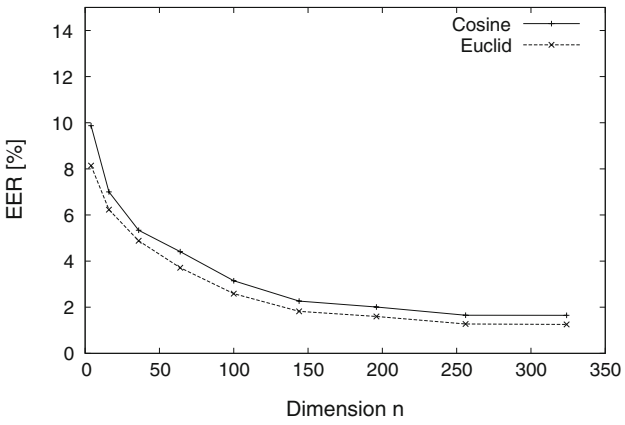


Fig. 6 Equal error rates (ERRs) with respect to dimension n

the experiment means that the Euclidean distance is superior in accuracy to the cosine correlation for all dimensions n. The result is compatible with the analysis of distributions studied in the above section.

Figure 7 shows the relative operating characteristic plot (ROC) for particular dimension $n = 18^2$, illustrating the change of false rejection rate (FRR) with respects to False Acceptance Rate (FAR). We observe that the tradeoff between these rates by varying thresholds, and the cosine correlation has higher error rate than the Euclidean distance. After all, the Euclidean distance is better similarity measure than the cosine correlation in terms of accuracy.

4.3 Performance

There are two factors for performance of protocols; the computational cost and the communication cost. The former is estimated as a number of modular exponentiations, which is the dominant factor of processing time, for each step in zero-knowledge proof. The latter is the function taking dimension n and size of modulus $\ell = |N|$, typically

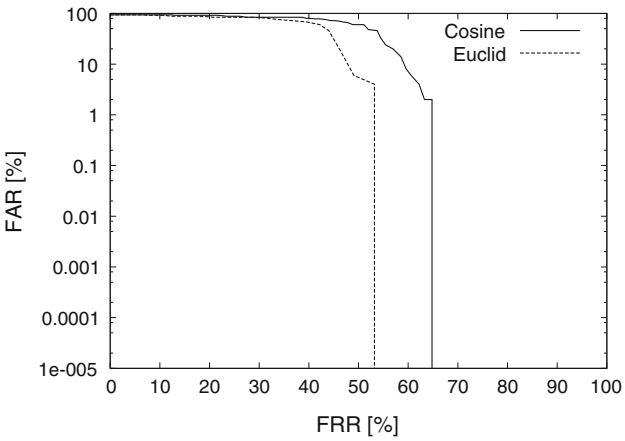


Fig. 7 Relative operating characteristic plot (ROC) for $n = 18^2$-dimension future vector

Table 2 Estimation of costs for each step in two protocols

	Private-Cosine		Private-Euclid	
Computation	1. PK_1	$3n$	1. PK_4	$11n$
	3. PK_2	19×2	3. PK_5	19×2
	Total	$3n + 38$		$11n + 38$
Communication	1. g	$n\ell$	1. $G, \tilde{F}$	$2n\ell$
	$PK1$	$n\ell$	PK_4	$3n\ell$
	3. PK_2	$5\ell \times 2$	3. PK_5	$5\ell \times 2$
	Total	$\ell(2n + 10)$		$\ell(5n + 10)$

$\ell = 1024$ bit. We summarize the estimation of both costs in Table 2. The estimation shows that the Euclidean distance requires about double, $5/2n\ell$ overhead of the cosine correlation to evaluate n -dimension vectors consisting of ℓ -bit integers, in theory.

In addition to the estimation from equations, we measure the processing time based on sample implementation of the protocols. Figure 8 shows the experimental results, where the size of modulus is $\ell = |N| = 1024$, security parameter in zero-knowledge protocol is $t = 160$ bit, and dimension of feature vector ranges from 2×2 to 18×18. We confirm that the estimation is compatible with the experimental result. Note that there is a constant amount of time at $n = 0$, which means the overhead caused at PK_2 and PK_5. In typical setting, say $n = 8^2$, protocol private-cosine and private-Euclid take $3,218$ and $8,078$ (ms), respectively.

4.4 Security

The security of the proposed protocols are based on the security of the strong RSA assumption, the difficulty of the decision Diffie–Hellman problem in the random oracle model. The probability to forge the commitments in PKs can be negligible as the security parameter increases. On the other hand, the common biometric features have less entropy than the commitment scheme. The probability of malicious party to impersonate someone without his biometric feature is fixed at a level determined by the entropy of the feature. Hence, the zero-knowledge protocol is secure enough to apply the biometric authentication.

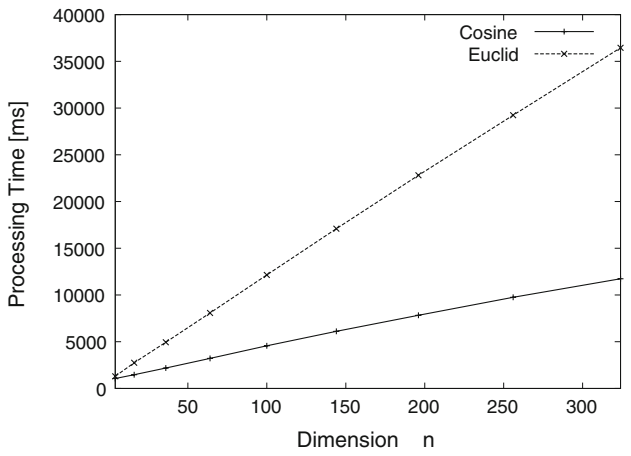


Fig. 8 Processing times for evaluating protocols with respect to dimension n

Definition 5 Let n be a secure RSA modulus. We say that RSA is a (t, ϵ)-secure one-way function if for any adversary A with running time less than t, we have $\Pr[A(N, e, x^e) = x] < \epsilon$, where $e \in Z^*_{\varphi(N)}, x \in Z^*_n$.

Theorem 1 (Security) *Assume that RSA is a* (t, ϵ_{RSA})*-secure one-way function. The Cosine correlation evaluation protocol is* (t', q, ϵ')*-secure against existential forgeries making at most q chosen-message queries and running in time at most t in the random oracle model.*

The security of the Euclidean distance evaluation protocol can be shown in the same way. The confidentiality of the committed value has been proved in Fujisaki and Okamoto (1997).

Theorem 2 (Fujisaki and Okamoto 1997) *If the probability that any polynomial-time adversary can factor n is negligible, there exists no probabilistic polynomial-time algorithm which given* $E(x/c, r)$ *can output* $(x'/c', r')$ *such that* $(x/c, r) \neq (x'/c', r')$ *and* $E(x/c, r) = E(x'/c', r')$.

Our model makes an assumption of tamper-freeness of secure device that stores the template feature vector with the random values used for commitment. We consider the assumption is reasonable in practical perspective since many secure devices are widely used in our daily life, e.g., the RFID and the smart cards. The requirement of secure device, however, is not useful from the usability point of view.

5 Conclusions

We have studied the protocols for secure similarity evaluation of vectors, private-cosine and private-Euclid, based on the zero-knowledge proof of range. The Private-Cosine allows a user to convince a server that the user has a secret similar to the data stored at server in a sense of the cosine correlation, while protocol Private-Euclid uses the Euclidean distance to evaluate similarity. The latter archives better accuracy in terms of false acceptance and rejection than the former in the cost of computational overhead. Our schemes are designed for secure remote biometric authentication that no malicious party including even server administrator can reveal private biometric data.

Acknowledgments We thank Dr. Yoichi Shibata, Mr. Taiki Sakashita and Mr. Takumi Yamamoto of Shizuoka University for their discussion on the topic with us. We thank Mr. Junichi Oshima of Tokyo Institute of Technology and Mr. Kenji Takahashi at Hitachi Limited.

References

Barbosa M, Brouard T, Cauchie S, Melo de Sousa S (2008) Secure biometric authentication with improved accuracy. In: Proceedings of ACISP 2008, LNCS 5107, pp 21–36

Boudot F (2000) Efficient proofs that a committed number lies in an interval. In: Proceedings EUROCRYPT 2000, LNCS 1807, pp 431–444, Springer, Heidelberg

Bringer J, Chabanne H (2008) An authentication protocol with encrypted biometric data. In: Proceedings of AFRICACRYPT 2008, LNCS 5023, pp 109–124

Bringer J, Chabanne H, Pointcheval D, Tang Q (2007) Extended private information retrieval and its application in biometrics authentications. In: Proceedings of CANS 2007, LNCS 4856, pp 175–193

Bringer J, Chabanne H, Pointcheval D, Zimmer S (2008) An application of the Boneh and Shacham Group signature scheme to biometric authentication. In: Proceedings of IWSEC 2008, LNCS 5312, pp 219–230

Clancy TC, Kiyavash N (2003) Secure smartcard-based fingerprint authentication. In: Proceedings ACM SIGMM 2003 Multim., Biom. Met. App., pp 45–52

Cramer R, Damgård I, Schoenmakers B (1994) Proofs of partial knowledge and simplified design of witness hiding protocols. In: Proceedings of CRYPTO '94, pp 174–187

Fujisaki E, Okamoto T (1997) Statistical zero knowledge protocols to prove modular polynomial relations. In: Proceedings of the CRYPTO '97, LNCS 1294, pp 16–30, Springer, Heidelberg

Jeong MY (2006) Changeable biometrics for appearance based face recognition. In: Proceedings of the biometric symposium. Biometric consortium conference, Baltimore, September

Juels A, Sudan M (2002) A fuzzy Vault scheme. In: Lapidoth A, Teletar E (eds) Proceedings of the IEEE international symposium information theory, p 408

Juels A, Wattenberg M (1999) A fuzzy commitment scheme. In: Tsudik G (ed) Sixth ACM conference computer and comm. Security, pp 28–36

Maltoni D, Maio D, Jain AK, Prabhakar S (2003) Handbook of fingerprint recognition. Springer Science+Business Media, Heidelberg

Ratha NK, Connell JH, Bolle RM (2001) Enhancing security and privacy in biometrics-based authentication systems. IBM Systems J 40(3)

Socek D, Božović V, Ćulibrk D (2008) Proceedings of ICETE 2007, CCIS 23, pp 139–151

Tang Q, Bringer J, Chabanne H, Pointcheval D (2008) A formal study of the privacy concerns in biometric-based remote authentication schemes. In: Proceedings of ISPEC 2008, LNCS 4991, pp 56–70

Uludag U, Jain AK (2004) Fuzzy fingerprint Vault. In: Proceedings workshop: biometrics: challenges arising from theory to practice, pp 13–16

Une M, Otsuka A, Imai H (2007) Wolf attack probability: a new security measure in biometric authentication systems. In: Proceedings of ICB 2007, LNCS 4642, pp 396–406