Proposal on Personal Authentication System in which Biological Information is embedded in Cryptosystem Key

Yukio Itakura [†,††] and Shigeo Tsujii [††]

Biometric personal authentication systems have a common problem— the biological information can easily be stolen by other individuals. In line with the process of the activities for the international standardization of the biometric system, this paper proposes a typical way to embed biological information, whatever its kind, into cryptographic keys as a measure for privacy protection and against unauthorized use. We believe that our proposal presents the following advantages: the improvement of protecting the privacy of biological information, economical effectiveness resulting from the practical use of the infrastructure of Public Key Infrastructure (PKI) as a biological information database, and humanity given to a man-machine interface by embedding an individual's biological information into a public key, an important element of the system. This paper also proposes how to build up a practical personal authentication system through the method proposed.

1. Introduction

Biometric personal authentication systems utilizing fingerprints or iris patterns have been studied and put to practical use. Biometric technology has become a familiar element of the security system.[1]

At the present time, the personal authentication system is mainly used at close range, such as for in-house room entry control, access to safes, and systems operation. In the future, it will be widely applied and diversified, particularly for a variety of approvals and settlements over networks, information access control, e-commerce via Internet, and remote personal identification. As a result, efforts for standardization are being made from a global point of view. In Europe and the U.S., for example, Biometrics Working Groups (BWG) are now studying biological authentication technologies to standardize the application program interface (API), common biometric exchange file format (CBEFF), matching accuracy evaluation, and IC card conjunction (with ISO/IEC SC17 WG4). In Japan, standardizing activities are also carried out in conjunction with them.[2]

In the future, these activities will require the higher-level privacy protection of biological information and the establishment of a precise and secure authentication method.

A problem common to biometric systems is that unauthorized use of biological information is very easy. For example, a fingerprint can be acquired from objects touched by the person, iris data can be obtained from the person's image captured by a camcorder, and DNA can be read out from a hair with a root.

Originally, biological information-based personal identification, that is, biometric personal authentication, was put to practical use on the precondition of a close range or face-to-face interface. Therefore, protecting the privacy of biological information and measures against unauthorized use have not been given sufficient consideration. We think that this problem will have to be more solved in the near future.

Since 1999, we have invented a personal authentication process using personal differences in DNA (deoxyribonucleic acid) to propose a public-key cryptographic scheme in which biological information is embedded into a secret key.[3~5] Such a DNA-based approach uses, as personal information, personal differences in the number of short nucleotide sequences repeated, called micro-satellites or short tandem repeats (STRs). In this case, the personal identification information is inherently digital and fixed. Accordingly, incorporating it as is into a cryptographic key is a totally natural idea. Moreover, we have found it convenient for a number of reasons to apply the fruits of the contemporary cryptographic theory to the biometric authentication system.

Based on the products[6] of studying the personal identification system with DNA, we began with the evaluation of the conventional systems using fingerprints or irises from a new angle.

† NTT Data Technology Corporation
†† Chuo University

A variety of systems in which individuals are identified based on their biological features to enable their secret keys have been introduced up to now.[7],[8] However, the secret key is only information generated by conventional mathematical calculation and is irrelevant to the biological information used for personal authentication, which is handled independently.

This paper offers a way to upgrade our previously proposed personal authentication system, in which DNA-based information is embedded into a secret key, to a system able to use general biological information as well.

2. The Features of Various Biological Information

Biological information used for biometric personal identification includes fingerprints, iris, face, voiceprint, signature, and DNA, some of which are already in practical use and some of which are in the process of research.

Figure 1 shows the features of biological information.

All the biological information other than DNA is a pattern (analog information), which is used as an attribute for identification. Personal identification is carried out by extracting featuring points from the pattern in a certain algorithm. The algorithm is unique to each device, that is, each manufacturer, and is not disclosed. The information content of the featuring points is called a template and is relatively high, from 250 to 1,000 bytes.[8] The deviation of the pattern changes the information to be identified. The lower FAR (False Acceptance Rate which is the probability of incorrectly accepting an unauthorized person), the higher FRR (False Rejection Rate which is the probability of incorrectly rejecting a genuine person). The identifying method based on pattern information must find a compromise between FAR and FRR. At the present time, FAR needs to be from half-millionths to one millionth.[9],[10]

The DNA method uses the combination of four bases (A: Adenine, G: Guanine, C: Cytosine, and T: Thymine) to partially extract the individually different sequence from so-called genome information for conducting personal identification.[6]

A problem common to all kinds of biological information is unauthorized use. Figure 1 also shows risks of biological information being stolen. Face-to-face personal authentication provides little risk of unauthorized use, but

identifying terminals or individuals remotely over networks may cause unauthorized use of biological information to increase notably.

For biological information to be globally used as a system in society, measures against unauthorized use are mandatory.

3. Typical Personal Authentication System Based on Biometry

Figure 2 is a flowchart that illustrates a typical personal authentication system based on biometry.[2] In the registration process on the left of the figure, the system first acquires biological information from the individual to be registered. Of the resulting data, featuring points are extracted in a given algorithm. Next, the system records the featuring points as a template in a database for later personal identification. The database is called the biological information registration database (DB).
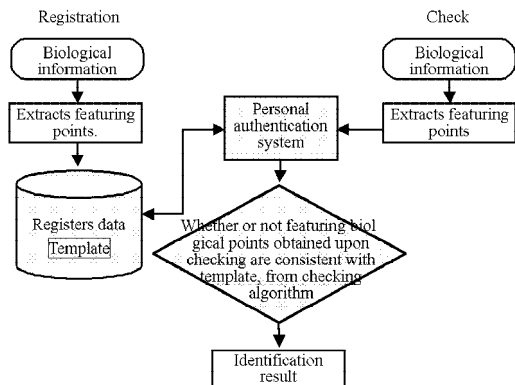
Fig. 2 Conventional Personal Authentication System Based on Biometry

In the checking process on the right of the figure, the system extracts featuring points in the same algorithm as when they were registered from the genuine biological information provided by the person (prover) who made a request for personal identification, to check whether the resulting data is consistent with the template recorded in the registration database. Finally, it compares the registered data (template) and the acquired data to calculate the degree of similarity or the distance from similarity and to make a decision based on a certain threshold.

Note that a new system has already been developed to carry out biometric personal identification by unlocking digital signatures or cryptographic communications, using a correspond-

Biological information		Fingerprint	Iris	Face	Voiceprint	Signature	DNA
Identifying principle		Personal difference in fingerprints or featuring points	Personal difference in iris patterns	Personal difference in facial features	Personal difference in vocal sounds	Personal difference in handwritten letters, pressure, and timing	Personal difference in short tandem repeats
Matching accuracy	FAR	2×10^{-6} or less	8.3×10^{-7} or less	10^{-2} or less	3×10^{-2} or less	10^{-2} or less	10^{-15} or less
	FRR	0.05% or less	0.1% or less	1% or less	3% or less	1% or less	Less than measuring error
Sensor		Image sensor	Camera	Camera	Microphone	Pressure sensor	Swab in mouse and DNA analyzer
Data size of template in bytes		250 to 500	250	1000	1000	1000	20
Feature and problem		Small-size, economic, and high precision	Small psychol stress and high precision	Small psychological stress	Small psychological stress	High precision in dynamic signature	High precision, uniqueness, and high stability with time
		Degradation of fingerprint due to dried skin	Low cost	Change due to aging, camera angle, hat, or eye glasses	Voice change in puberty or due to thirsty throat	Ease of imitation	Long analyzing time, high price, and privacy concerns
Risk of unauthorized use		Fingerprint marked	Eye captured by camcorder	Face captured by camcorder	Voice recorded by microphone	Handwriting imitated	Stolen hair with root

FAR: False Acceptance Rate, FRR: False Rejection Rate, and values are quoted from catalogs [9] and [10] as well as references [2] and [8].
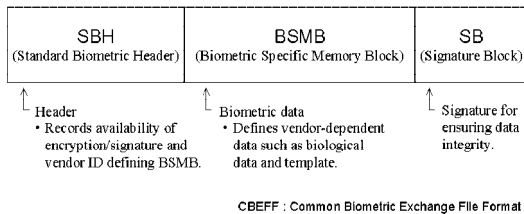
Fig. 1 Features of Various Biological Information

ing secret key. This system first checks the biological information based on the algorithm mentioned above, and after the personal identification is successful, it makes the secret key recorded in an IC card effective to make calculations for the signature.[2]

4. Problems that the Conventional System Encounters

4.1 Template Standardization Problems

CBEFF, mentioned before, is being standardized as a template along with work on BioAPI.[11),12)] Figure 3 illustrates a file format for exchanging common biometric data, this is the CBEFF standard, disclosed by the National Institute of Standards and Technology (NIST), the Department of Commerce.

SBH (Standard Biometric Header)	BSMB (Biometric Specific Memory Block)	SB (Signature Block)

Header
• Records availability of encryption/signature and vendor ID defining BSMB.

Biometric data
• Defines vendor-dependent data such as biological data and template.

Signature for ensuring data integrity.

CBEFF : Common Biometric Exchange File Format

Fig. 3 Structure of CBEFF

This format consists of three blocks. The first block is a header including a biometric type, the availability of encryption and signature, and a vendor ID that has defined a template written in the data block.

The second block is a data field called a biometric specific memory block (BSMB) and can be defined arbitrarily by vendors. For example, biological information itself or a template unique to a vendor may be stored.

The third part is a signature block (SB) which is used to ensure data integrity.

As mentioned above, the template standardization activity is taking place but, in addition to the file format of the template itself not yet being standardized, the following important challenges still persist: the compatibility of related products, the data exchange between biological databases, and the encryption of biological data in a unified concept.

4.2 Problem in the Relationship between Biological Information and Cryptographic Keys

Since biological information is personal identification data, it should be positioned as a personal secret key or the equivalent to precisely define the relationship between both. However, no report about this problem has been submitted in the activities for CBEFF and BioAPI. Moreover, the existing personal authentication systems with biological information handle the biological information and personal secret key as an independent element. The systems use the biological information to identify the genuine person and send a signal to enable the secret key only when the authentication is successful. This means that the biological informa-

tion is not incorporated into the secret or pubic key as information for generating the key.

Letting an external organization like PKI issue the biological certificate is a way to ensure the validity of biological information. Even in this case, the secret key is generated mathematically and is not related to the biological information.

Addressing this problem is not only a goal in the mathematical algorithm field but also an important challenge for giving the overall system humanity by intuitively linking cryptographic theory and biological information, that is, embedding the biological information into a key used by a cryptographic system having the function of authenticating the information.[13]

### 4.3 Problem in Constructing the Biological Database

Assuming that a template has the data size shown in Figure 1, when the biological database is built up on a global scale huge facilities are necessary as a new social infrastructure. Since data stored in the template is not standardized completely and is left to vendors, as mentioned before, the interface between the databases may be various and complicated. If it takes a long time to standardize the template, it is necessary to study a simple biological database during that time.

5. Personal Authentication System in which Biological Information is Incorporated into Cryptographic Keys

5.1 Significance of Incorporating Biological Information into Cryptographic Keys

To address the problems mentioned in the previous section, this section examines how to embed biological information into cryptographic keys. The cryptographic keys consist of secret and public keys used in the public-key cryptographic scheme. Incorporating the biological information into the secret key means that the same data is automatically embedded into the public key generated in an algorithm defined by the system. After this, the cryptographic keys consisting of the secret and public keys into which the biological information is embedded are referred to as the biometrics-cryptographic keys.

As shown in Figure 4, embedding biological information into the cryptographic keys has the following advantages.
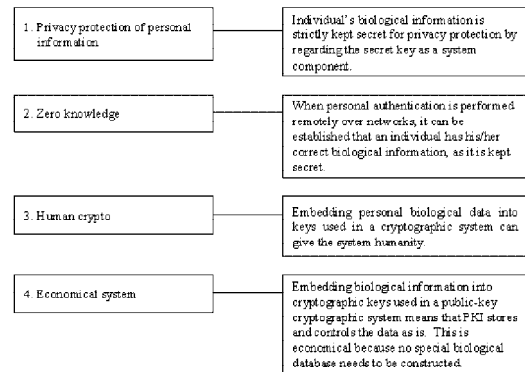
Fig. 4 Advantages Given by Embedding Biological Information into Cryptographic Keys

(1) Privacy protection of personal information

(2) Zero knowledge, which means that no biological information is given directly to an inspector

(3) Humanity resulting from embedding biological information into the cryptographic keys

(4) Economical system which doesn't need to build up its own biological database

If personal authentication is performed according to featuring points generated from a fingerprint or iris, the identifier has never been incorporated into any cryptographic key because of its ambiguity. However, the system that we will propose later in this paper can embed biological information into the cryptographic keys.

Since there is the risk of unauthorized use of the biological information, as shown in Figure 1, measures, such as preparing the original secret key separately, are necessary to guarantee the security of the overall personal authentication system even if the biological information itself is deciphered.

### 5.2 How to Generate the Biometrics-Cryptographic Keys

This subsection, along with figures 5 and 6, describes how to generate the biometrics-cryptographic keys, into which biological information is embedded.

(1) Extracting featuring points to generate a template

An algorithm for extracting featuring points from raw biological information to generate a template is vendor-specific. For example, from a fingerprint pattern, the positional and relational information of branch and end points is acquired as
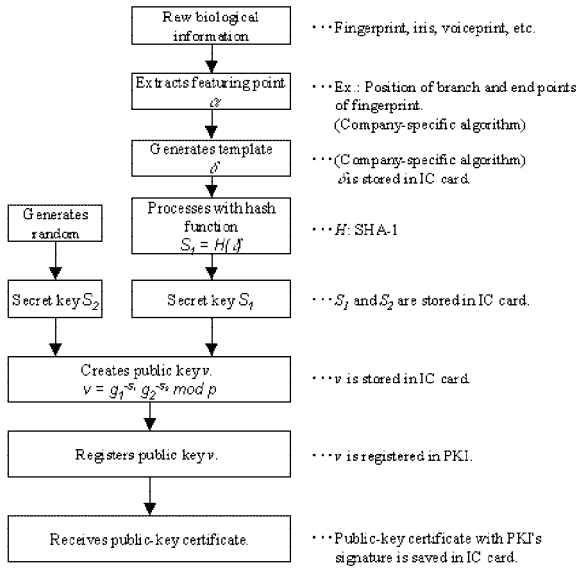
Fig. 5 How to Generate Cryptographic Keys Having Biological Information
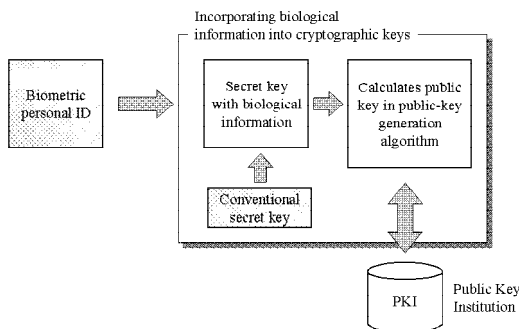
Fig. 6 How to Incorporate Biological Information into a Secret Key

a featuring point to create a template. The practical compressed data size of the template ranges from 250 bytes to 500 bytes.

Letting the featuring points extracted and the template formatted be α and δ respectively, the latter is stored in the BSMB (Biometric Specific Memory Block) of the data file format defined by the CBEFF standard.

Since the data format of the template δ is not yet standardized and still vendor-specific, this paper studies a way to store δ in an IC card that each individual has and to first compare the genuine person and the IC card that he/she has in a client system. This system is called a client authentication model be-

cause identification is carried out by the client having the biological information and matching algorithm.

Since the client authentication model gives the client authority and responsibility for authentication, the 3-communication protocol to be described later in this paper, for example, is necessary to prevent anyone from impersonating the client system itself. This protocol registers a generated public key in PKI and uses a challenge bit when personal authentication is performed over networks.

(2) Generating a secret key

Though the template δ consists of bit strings written in various formats defined on a vendor basis, it can be regarded as a personal identifier showing that the template corresponds uniquely to the genuine person with a certain matching probability. To minimize the ambiguity of δ as a personal identifier, a fingerprint is taken repeatedly upon initial registration if the same δ is stored in the client system. When the same δ is generated from two individuals even if their fingerprints are repeatedly taken many times, it shows the accuracy limitation of this system. In this case, it is necessary to study a new way, such as combination with other biological information.

In this paper, assuming that δ is generated uniquely as a personal identifier, s_1 is given by processing δ through a hash function. Namely, $s_1 = H(\delta)$ is defined.

The left side s_1 is a secret key having an individual's biological information. A candidate for the hash function H is SHA-1. In this case, the input data size is smaller than 2^{64} bits, which is enough for the bit string size of the template. The data processed and provided by the hash function is 160 bits long, that is a compressed key length of 20 bytes.

If the biological information and template generating algorithm are stolen, the risk of forging s_1 occurs. Therefore, the second secret key s_2 is defined as a random number generated by conventional mathematical calculation and the two secret keys s_1 and s_2 guarantee the secu-

rity.

In order to conduct personal authentication in this condition, this paper employs the 3-communication authentication model proposed in the paper[14] presented by Dr. Tatsuaki Okamoto.

In this model, the public key v is derived from a pair of the secret keys s_1 and s_2. The v generating algorithm and 3-communication authentication model will be described later in this paper.

The created key v is registered in Public Key Infrastructure (PKI).

A public-key certificate with PKI's signature issued after the registration is stored in the personal ID card with δ, s_1, s_2, and v. The IC card must be tamperproof and be sufficiently protected from unauthorized use.

5.3 Algorithm Used in the 3- Communication Authentication Model

This subsection describes the 3-communication authentication model and protocol.

[Notation]

Secret key : (s_1, s_2)

s_1 : Biological information-based secret key (given by processing the template through the hash function). The data is, for example, 160 bits long.

s_2 : Conventional secret key (for example, 160 to $1,024$ bits long)

Both s_1 and s_2 belong to the Z_p group.

Public key : (v, p, q, g_1, g_2, t)

$$v = g_1^{-s_1} g_2^{-s_2} \bmod p \qquad (1)$$

where

p and q are prime numbers and $q \mid p - 1$.

g_1 and g_2 are integers of the order q belonging to the multiplicative group $Z_p{}^*$.

g_2 is given by α chosen arbitrarily and the equation $g_2 = g_1^{\alpha} \bmod p$.

α is discarded after the generation of g_2.

t is an integer and $t = 0(\mid p \mid)$.

[Personal identification protocol]

Step 1: Mr. A (prover) selects the random numbers r_1 and r_2 $(r_1, r_2 \in Z_q)$ arbitrarily to give x with the following equation:
$$x = g_1^{r_1} g_2^{r_2} \bmod p \qquad (2)$$
Mr. A sends the resulting value x to Mr. B (verifier).

Step 2: Mr. B selects the arbitrary random number $e(e \in Z_{2^t})$ and sends it to Mr. A.

Step 3: Mr. A derives y_1 and y_2 from the following equations:

$$y_1 = r_1 + es_1 \bmod q \qquad (3)$$
$$y_2 = r_2 + es_2 \bmod q \qquad (4)$$

Mr. A sends (y_1, y_2) to Mr. B.

Step 4: Mr. B checks whether the following equation is satisfied.

$$x \equiv g_1^{y_1} g_2^{y_2} v^e \bmod p \qquad (5)$$

The personal identification is successful if satisfied, otherwise unsuccessful.

The above proves that even if the secret key s_1 derived from the biological information is stolen, the system security can still be guaranteed unless the conventional secret key s_2 is leaked.

5.4 Personal Authentication System Using Biological Information

This subsection describes an embodiment of the personal authentication system in which biological information is embedded into the cryptographic keys.

Figure 7 is a flowchart that illustrates the process of the generation and registration of the biometrics-cryptographic keys shown in Figure 5. Since this system employs a client authentication model, various data and programs for identification are stored in an individual's IC card and the client terminal.

Figure 8 is a flowchart that illustrates the process of actual personal authentication with biological information after the preparation is complete.

This system performs personal authentication from the remote application server via the network.

For the personal authentication, the three communications of steps 1, 2, and 3 are conducted among the application server, client terminal, and individual's IC card.

Explaining more precisely, the prover, first of all, inserts his/her ID card having his/her biological information template and cryptographic keys into the client terminal, and his/her fingerprint is taken by the sensor. The client terminal compares the template stored in the tamperproof IC card and the featuring points of the biological information acquired, and enables the secret key if the authentication is successful based on a given threshold and matching algorithm. The application server communicates with the IC card and client terminal over the network, that is challenge and response.
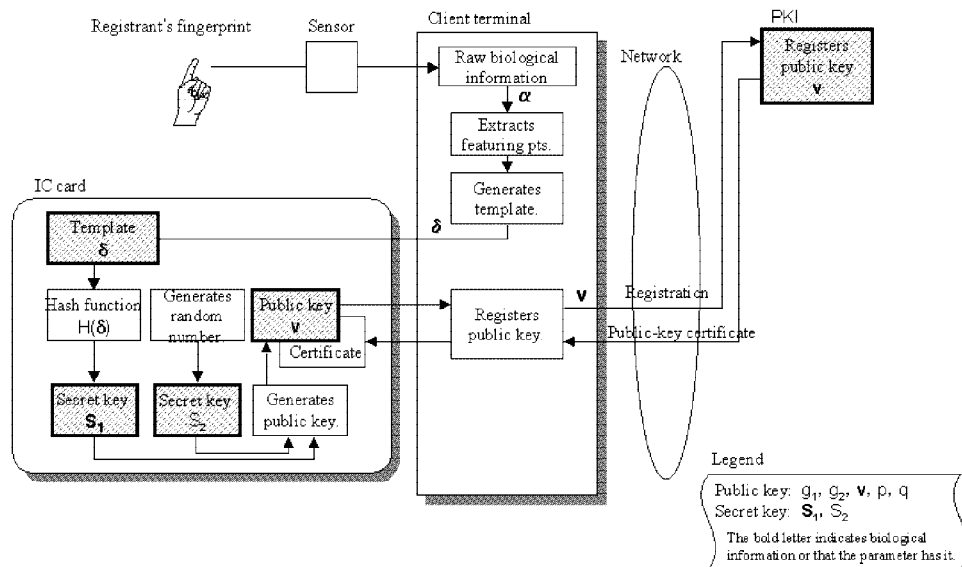
Fig. 7 Generation and Registration of Biometrics-Cryptographic Keys
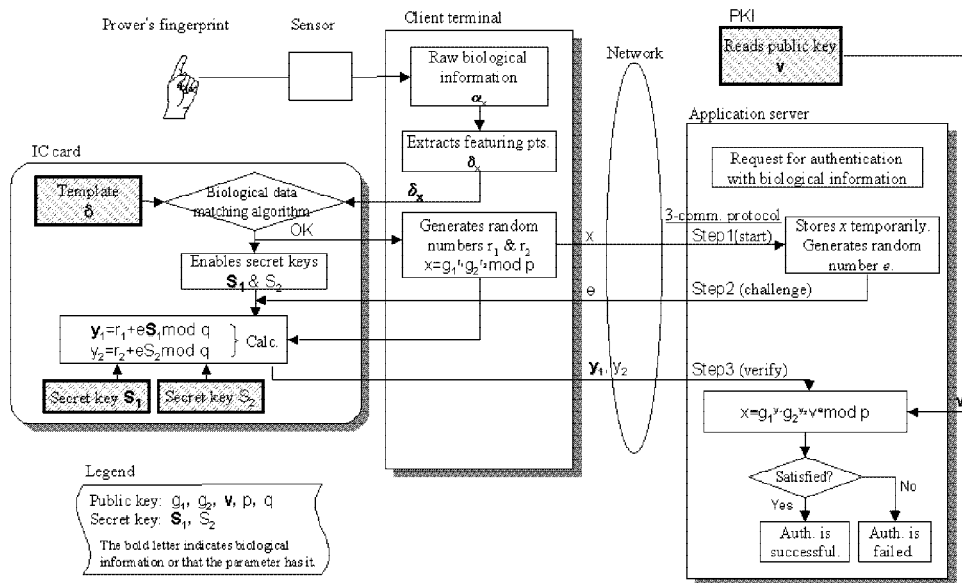

Fig. 8 Personal Authentication System with Biometrics-Cryptographic Keys

Since these three communications are conducted via the cryptographic keys into which the raw biological data and template are embedded, rather than directly via them, the privacy is never leaked to the outside via the network. Even if s_1 can be derived from stolen personal biological information and template generating algorithm, the security of the authentication system can still be guaranteed, unless the secret key s_2 is leaked.

To impersonate another person, the impersonator needs to steal the person's IC card and biological information to reproduce an artificial finger. Even if the impersonation is successful, it is exposed easily because the matching of the impersonator's fingerprint taken by an inspector is unsuccessful in court.

However, the impersonation would be successful if the public key could be registered in the system under disguised ownership. This is a

problem common to the other systems. Accordingly, the best effort is necessary to avoid the impersonation by precisely comparing an individual with his/her passport and driver's license upon registration.

5.5 Security Verification

Reference [14] shows that the difficulty of solving a discrete logarithm problem is the same as the safety factor of the authentication system mentioned in Section 5.4.

It is assumed that another person, Mr. A', steals the biometric personal ID s_1 of Mr. A and pretends to be Mr. A.

Moreover, Mr. A' can derive $s_2{}^*$ that satisfies the following equation from the stolen s_1 and his biometric personal ID $s_1{}^*$.

$$v = g_1^{-s_1} g_2^{-s_2} \bmod p = g_1^{-s_1{}^*} g_2^{-s_2{}^*} \bmod p \quad (6)$$

Mr. A' uses his $s_1{}^*$ and $s_2{}^*$ given by the equation above to derive $y_1{}^*$ and $y_2{}^*$ in the same fashion as the calculation in Step 3.

$$y_1{}^* = r_1{}^* + es_1{}^* \bmod q \quad (7)$$
$$y_2{}^* = r_2{}^* + es_2{}^* \bmod q \quad (8)$$

Mr. A' sends the resulting $y_1{}^*$ and $y_2{}^*$ to Mr. B.

Provided $r_1{}^*$ and $r_2{}^*$ have been chosen by Mr. A' who pretends to be Mr. A in Step 1, e has been generated by Mr. B (verifier) in Step 2, and Mr. A' has sent x given by the equation $x = g_1^{r_1{}^*} g_2^{r_2{}^*} \bmod p$ to Mr. B.

In Step 4, Mr. B makes the following calculation:

$$
\begin{aligned}
&g_1^{y_1{}^*} g_2^{y_2{}^*} v^e \bmod p \\
&= (g_1^{r_1{}^* + es_1{}^* \bmod q} g_2^{r_2{}^* + es_2{}^* \bmod q} g_1^{-es_1{}^*} g_2^{-es_2{}^*}) \\
&\quad \bmod p \\
&= g_1^{r_1{}^*} g_2^{r_2{}^*} \bmod p \\
&\equiv x
\end{aligned}
$$
$$(9)$$

Since the equation above holds, the authentication is passed. Accordingly, Mr. A' can pretend to be Mr. A successfully.

Letting an algorithm for deriving $s_2{}^*$ be ALG, ALG can be defined as follows:

[Definition 1]

ALG is a stochastic algorithm having the input parameters g_1, g_2, v, s_1, and $s_1{}^*$ to provide $s_2{}^*$ that satisfies Equation (6) with a non-negligible probability.

First of all, g_1 and g_2 are specified so that $g_1 = g_2^\beta$ holds. In Equation (1) for generating a public key, $g_2 = g_1^\alpha$ is provided, so $\beta = 1/\alpha$ holds. The algorithm DLP_{g_2} that solves a discrete logarithm problem against g_2 is defined as follows:

[Definition 2]

DLP_{g_2} is a stochastic algorithm having the input parameters g_2, k, and p to provide w that satisfies $k = g_2^w$ with a non-negligible probability.

Assuming that ALG exists, DLP_{g_2} can be solved with a non-negligible probability. Actually, the following theorem is established.

[Theorem 1]

If ALG is used, DLP_{g_2} can be constructed in polynomial time.

[Demonstration]

It is assumed that the input parameters g_2, k, and p of DLP_{g_2} are given. A random value is substituted into β and g_1 is given by $g_1 = g_2^\beta$. Next, v is derived from $v = g_1^{-s_1} k$ and then $g_1, g_2, v, s_1, s_1{}^*$ are substituted into ALG, which provides $s_2{}^*$ that satisfies Equation (6) with a non-negligible probability. Since $g_1^{-s_1} g_2^w = g_1^{-s_1{}^*} g_2^{-s_2{}^*}$, using $g_1 = g_2^\beta$ gives the following equation:

$$-s_1 \beta + w = -s_1{}^* \beta - s_2{}^* \quad (10)$$

The equation above can be rewritten as $w = \beta(s_1 - s_1{}^*) - s_2{}^*$, resulting in the output w of DLP_{g_2}.

However, DLP_{g_2} cannot be solved with a non-negligible probability, which denies the assumption that ALG exists. Accordingly, the following theorem holds.

[Theorem 2]

A stochastic algorithm ALG that has the input parameters g_1, g_2, v, s_1, and $s_1{}^*$ to provide $s_2{}^*$ that satisfies Equation (6) with a non-negligible probability does not exist.

This theorem proves that unauthorized use by means of calculating $s_2{}^*$ is very difficult.

References

1) Tomoyuki Kan, et al. : *Cutting-edge Biometric Personal Authentication System*, Information Processing, Vol. 40, No. 11, pp. 1072-1103 (1999).

2) Yoichi Seto : *Activities for Standardizing Biometric Technologies*, Text of Symposium, IEICE's Tokyo Branch (May 2002).

3) Shigeo Tsuji : *Private letter to Kaoru Kurosawa, Professor of Tokyo Institute of Technology*, (January 1999).

4) Shigeo Tsuji : *Document for Third FAIT Meeting*, (August 1999).

5) Shigeo Tsujii, Yukio Itakura, Hiroshi Yamaguchi, Atsushi Kitazawa, Shinya Saito, Masao Kasahara : *Public-key Cryptographic Scheme Having a Structure in which Biological Information is Embedded into a Secret Key*, IEICE Symposium, SCIS2000, D07 (January 2000).

6) Yukio Itakura, Masaki Hashiyada, Toshio Nagashima, and Shigeo Tsujii : *Statistical Verification of DNA-based Personal ID*, IEICE Transactions, Vol. 101, No. 214, ISEC2001-19, pp. 1-7 (July 2001).

7) Yoichi Seto : *Personal Authentication System*, Japanese Examined Patent Application Publication No. 2000-215280 (August 2000).

8) Yoichi Seto : *Biological Authentication Technologies*, Kyoritsu Shuppan Co., Ltd., pp. 13 (2002).

9) NEC Solutions: The fingerprint identification system : *SecureFinger*.
   http://www.sw.nec.co.jp/pid/about_pid.html

10) Oki's Product: IRISPASS-WG Gate Control System.
   http://www.oki.com/jp/SSG/JIS/Prod/iris/iriswg.html

11) *CBEFF, Common Biometric Exchange File Format*, NIST6529 (2001.1)
   http://www.itl.nist.gov/div895/isis/cbeff/

12) *BioAPI Specification Version 1.1*, The BioAPI Consortium (2001.3).
   http://www.bioapi.org/

13) Hideki Imai, Kazukuni Kobara, Yodai Watanabe : *About Human-Crypto*, IEICE Transactions, Vol. 100, No. 77, ISEC2000-17, pp. 57-64 (May 2000).

14) Okamoto.T : *Probably Secure and Practical Identification Schemes and Corresponding Signature Schemes*, Proc. of Crypto '92, LNCS740, Springer-Verlag, pp. 31-53 (1995).