Forgotten Secret Recovering Scheme and Fuzzy Vault Scheme Constructed Based on Systematic Error-Correcting Codes

Masao KASAHARA
Faculty of Informatics, Osaka Gakuin University, Suita-shi, 564-8511 Japan.
kasahara@ogu.ac.jp

Abstruct

In this paper, we revisit the Forgotten Secret Recovering Scheme for k users, referred to as FSRS(k) previously proposed by the present author. FSRS(k) takes advantage of the fact that Reed-Solomon code used in FSRS(k) is constructed in a form of systematic code. We show that a particular class of FSRS(k), FSRS(1), can be successfully applied to the various cryptoscheme including Fuzzy Vault Scheme (FVS).

Keywords

Forgotten Secret Recovering Scheme, Fuzzy Vault Scheme, Reed-Solomon code, Movie Lover's Problem

1 Introduction

As is well known, error-correcting codes are used in many areas of cryptography such as McEliece PKC [1] and secret sharing scheme due to McEliece and Sarwate [2].

From 2000 to 2009, the present author proposed the "Forgotten Secret Recovering Problem (FSRP)" [3-5], McEliece-Type PKC[6] and Multivariate PKC[7-8] all based on error-correcting codes.

In 1999, A. Juels and M. Wattenberg presented an interesting "Fuzzy Commitment Scheme" [9], where they combined the techniques from the fields of error-correcting code and cryptography. In 2002, A.Juels and M.Sudan proposed an interesting problem referred to as "Movie Lover's Problem (MLP)" [10]. They presented a Fuzzy Vault Scheme (FVS) based on Reed-Solomon (RS) code in a non-systematic form, on this matter.

In Aug. 2000, the present author submitted a challenge problem "How to prevent a user forgetting about password?" in the news-letter of FECCS Society, IEICE Japan having a circulation of about 10,000 [3]. Unfortunately we got no mathematical solution from the readers. Accordingly in 2001 the present author published the solution in Ref. 4. The scheme given in Ref. 4 is referred to as Forgotten Secret Recovering Scheme for k users, FSRS(k).

In this paper, we show that a particular class of FSRS(k), FSRS(1), using systematic Reed-Solomon code can be successfully applied to the FVS.

2 Forgotten Secrets Recovering Scheme

First we present the following remark on the using of Reed-Solomon code.

Remark 1 : The present author believes that RS code should be defined as cyclic code (more generally pseudo cyclic code) constructed by the generator polynomial $G(X)$. The constructing RS code in a systematic form would open up a wide area of applications in cryptography. □

Let us define the sysmbols :

N :	Code length, $N \leq 2^m - 1$.
k :	Number of information symbols.
(N,k) RS code :	RS code of code-length N and k information symbols over $\mathbb{F}_{2^m}$.
$G(X)$:	Generator polynomial of (N,k) RS code over $\mathbb{F}_{2^m}$.
d :	Degree of $G(X)$, $2e$.
D :	Minimum distance of RS code, $D = d+1$, $2e+1$.
s_i :	User I's secret.
$\{s_i\}$:	Set of secrets, $s_1, s_2, \cdots, s_k$ over $\mathbb{F}_{2^m}$.
$\boldsymbol{s}$:	Vector of secrets, $(s_1, s_2, \cdots, s_k)$ over $\mathbb{F}_{2^m}$.
$\{u_i\}$:	Set of check symbols, u_1, u_2, $\cdots$, u_d over $\mathbb{F}_{2^m}$.
$\boldsymbol{u}$:	Vector of check symbols, $(u_1, u_2, \cdots, u_d)$ over $\mathbb{F}_{2^m}$.
$\sharp A$:	Order of set A.
$H(x)$:	Entropy of x (in bit).
$H(x\|y)$:	Conditional entropy of x when y is given (in bit).

From the above definition it is easy to see that $\sharp\{s_i\}$ and $\sharp\{u_i\}$ are given by k and d respectively.

Letting $s(X)$ be

$$s(X) = s_1 + s_2X + \cdots + s_kX^{k-1}, \quad (1)$$

the check symbols are given by

$$s(X)X^d \equiv u_1 + u_2X + \cdots + u_dX^{d-1} \mod G(X). \quad (2)$$

Let us assume the followings :

Assumption 1 : Entropy of $s_i \in \mathbb{F}_{2^m}$, $H(s_i)$, takes on the same value of $H(s_i) = m$ (bits), $i = 1, \cdots, k$. □

Assumption 2 : Conditional entropy, $H(s_i|s_j)$, satisfies

$$H(s_i|s_j) = H(s_i), \text{ for } i \neq j. \quad (3)$$

□

An interesting scheme referred to as "Forgotten Secret Recovering Scheme for k users (FSRS(k))" can be summarized as follows :

FSRS(k)(Forgotten Secret Recovering Scheme for k Users):

Premiss 1	:	$m(k-d) \gtrsim 128$.
Public Data	:	$\boldsymbol{u} = (u_1, u_2, \cdots, u_d)$.
Secret Data	:	$\boldsymbol{s} = (s_1, s_2, \cdots, s_k)$.
s_i	:	User i's secret.

In FSRS(k) it is assumed that, at k users, $s_1, s_2, \cdots, s_k$ are completely deleted and learned only by heart, while $u_1, u_2, \cdots, u_d$ are publicized. When λ secrets among $\{s_i\}$ are completely forgotten, the forgotten secrets can be recovered correctly under the cooperations of the remaining $(k - \lambda)$ users without disclosing any amount of secrets each other, provided that

$$\lambda < d \quad (4)$$

holds [4,5]. [1] □

It should be noted that, in the recovering process, no secrets is disclosed. We believe that the secret recovering algorithm, KRA presented in [4] and [5] would be of very interest.

3 Movie Lover's Problem

3.1 Movie Lover's Problem

In 2002, Juels and Sudan presented a very interesting problem referred to as "Movie Lover's Problem (MLP)[10]". They also presented an interesting scheme for solving MLP. We shall refer to their scheme as JS scheme.

MLP(Movie Lover's Problem) :

For simplicity we assume that Alice plans to invite the people to her movie club, keeping the title of her favorite movie secret. Bob receives the application form for admission to Alice's movie club. To be a member of Alice's movie club, a great part of Bob's favorite movies is required to be coincident with Alice's favorites. □

3.2 Solution of MLP Based on JS Scheme

We shall refer to the solution presented due to Juels and Sudan in this subsection as Solution JS [10]. Let us define the symbols :

$\boldsymbol{s}$:	Alice's secret-key vector, $\boldsymbol{s} = (s_1, s_2, \cdots, s_\mu)$ over $\mathbb{F}_{2^m}$ that allows an applicant to join her movie club.
$s(X)$:	Alice's secret-key polynomial, $s_1 + s_2X + \cdots + s_\mu X^{\mu-1}$ over $\mathbb{F}_{2^m}$.
A :	Set of Alice's favorites, $\{a_i\}$ over $\mathbb{F}_{2^m}$.
$\boldsymbol{v}_A$:	Vector representing Alice's favorites, $(a_1, a_2, \cdots, a_k)$, where $k > \mu$.
B :	Set of Bob's favorites, $\{b_i\}$ over $\mathbb{F}_{2^m}$.
$\boldsymbol{v}_B$:	Vector representing Bob's favorites, $(b_1, b_2, \cdots, b_k)$.
S_r :	Set of random elements, $\{r_i\}$ over $\mathbb{F}_{2^m}$, $S_r \bigcap A = \phi$.
S_A :	Set of secret pairs $\{(a_i, s(a_i))\}$, $i = 1, \cdots, k$.
λ_i :	Element of S_A, i=1, 2, $\cdots$, k.
$\boldsymbol{w}_A$:	Vector, $(\lambda_1, \lambda_2, \cdots, \lambda_k)$.
S_R :	Set of random pairs $\{(r_i, v_i)\}$, $i = k+1, \cdots, 2^m - 1$, $v_i \in \mathbb{F}_{2^m} - \{s(a_i)\}$.
$S_{A,R}$:	$S_A \bigcup S_R$.

In Ref [10], the vector($s(a_1), s(a_2), \cdots, s(a_k)$) is considered to be the codeword of (k, μ) shortened RS code over $\mathbb{F}_{2^m}$ in a non-systematic form.

Let the elements of $S_{A,R}$ be denoted by $\eta_1, \eta_2, \cdots, \eta_N$, where N is $2^m - 1$. Letting $\boldsymbol{P}_R$ be an $N \times N$ random permutation matrix, we obtain

$$(\eta_1, \eta_2, \cdots, \eta_N)\boldsymbol{P}_R = (M_1, M_2, \cdots, M_N) \quad (5)$$

The vector $(M_1, M_2, \cdots, M_N)$ is now publicized. We assume that the set $S_{A,R}$ is constructed in a sufficiently random manner. Thus the conditional entropy is given by

$$H(M_i \in S_A | M_1, M_2, \cdots, M_N) = -\log_2 \frac{k}{N} \text{ (bit)}. \quad (6)$$

Let us define a distance between $\boldsymbol{w}_A$ and $\boldsymbol{w}_B$ referred to as "permutation distance" and denoted by $D(\boldsymbol{w}_A \backsim \boldsymbol{w}_B)$ in the following manner :

$$D(\boldsymbol{w}_A \backsim \boldsymbol{w}_B) = \min_{i,j} H(\boldsymbol{w}_A\boldsymbol{P}_i - \boldsymbol{w}_B\boldsymbol{P}_j), \quad (7)$$

where $\boldsymbol{P}_i$ and $\boldsymbol{P}_j$ are the permutation matrices and $H(\boldsymbol{x})$ is the Hamming distance.

[1] In FSRS(k), forgotten secrets can be regarded as erasure errors.

For example, when $\boldsymbol{w}_A$ and $\boldsymbol{w}_B$ are given by

$$\boldsymbol{w}_A = (7\ 1\ 3\ 4\ 2) \tag{8}$$

and

$$\boldsymbol{w}_B = (6\ 2\ 1\ 8\ 3), \tag{9}$$

then by letting

$$\boldsymbol{w}_A \boldsymbol{P}_A = (1\ 2\ 3\ 4\ 7) \tag{10}$$

and

$$\boldsymbol{w}_B \boldsymbol{P}_B = (1\ 2\ 3\ 6\ 8), \tag{11}$$

the permutation distance between $\boldsymbol{w}_A$ and $\boldsymbol{w}_B$ is given by

$$D(\boldsymbol{w}_A \backsim \boldsymbol{w}_B) = 2. \tag{12}$$

It is easy to see that Bob is able to recover the set of Alice's secret keys $s_1, s_2, \cdots, s_k$ provided that

$$D(\boldsymbol{w}_A \backsim \boldsymbol{w}_B) \leqq \frac{k-\mu}{2}. \tag{13}$$

When Eq.(16) holds, Bob is granted admission to join Alice's movie club.

4 Solutions of MLP Based on FSRS(1)

4.1 Preliminaries

We shall present a solution for MLP referred to as Solution A, based on FSRS(1). The followings are published.

Public Data

T_{MOV} :	Table listing all kinds of movie titles over $\mathbb{F}_{2^m}$.
$\sharp T_{\mathrm{MOV}}$:	$2^m - 1$.
M_i :	Movie title, $i = 1, 2, \cdots, 2^m - 1$.
k :	Total number of any person's favorite movies.

It should be noted that any movie title is represented by an element of $\mathbb{F}_{2^m}$. Thus the order of T_{MOV} is given by $2^m - 1$. The vectors representing Alice's favorite movies and Bob's favorite movies, $\boldsymbol{v}_A$ and $\boldsymbol{v}_B$ are given by

$$\boldsymbol{v}_A = (a_1, a_2, \cdots, a_k) \tag{14}$$

and

$$\boldsymbol{v}_B = (b_1, b_2, \cdots, b_k) \tag{15}$$

respectively. In 4.2 we shall present Solution A for MLP based on FSRS(1).

4.2 Solution A for MLP

We assume that Alice selects the set of her k favorite movies $\{a_i\}$ over $\mathbb{F}_{2^m}$ under the condition that $H(a_i) = -\log_2 \frac{1}{N_T}$ holds. Recalling that $\{a_i\} = S_A$ and $\{b_i\} = S_B$, let $S_A \bigcup S_B$ be denoted by $S_{A,B}$. When the order of $S_A \bigcap S_B$, $\sharp(S_A \bigcap S_B)$ satisfies

$$\sharp\Big(S_A \bigcap S_B\Big) \leqq \frac{d}{2} = e, \tag{16}$$

$S_{A,B}$ will be referred to as favorite sets of Alice and Bob, and denoted by $S_{A,B}^{(F)}$.

Let us define (a_i) be an integer when a_i is read as the conventional binary number. For example, for $\alpha = (101) \in \mathbb{F}_{2^3}$, (α) is read as $5 \in \mathbb{Z}$. We then have

$$\begin{aligned} a_i X^{(a_i)} \equiv\ & u_{Ai,1} + u_{Ai,2} X \\ & + \cdots + u_{Ai,d} X^{d-1} \mod G(X). \end{aligned} \tag{17}$$

Let $w_A(X)$ be given as

$$\begin{aligned} a_1 X^{(a_1)} + a_2 X^{(a_2)} + \cdots + a_k X^{(a_k)} \\ \equiv w_A(X) \mod G(X). \end{aligned} \tag{18}$$

Let us denote $w_A(X)$ as

$$w_A(X) = w_{A1} + w_{A2} X + \cdots + w_{Ad} X^{d-1}. \tag{19}$$

Alice then publishes ($w_{A1}, w_{A2}, \cdots, w_{Ad}$) along with her generater polynomial $G_A(X)$ of RS code over $\mathbb{F}_{2^m}$.

In a similar manner, referring to T_{MOV}, Bob constructs the set of his favorite movie, $\{b_i\}$ over $\mathbb{F}_{2^m}$ and obtains $w_B(X)$ using Alice's generator polynomial $G_A(X)$ as

$$w_B(X) = w_{B1} + w_{B2} X + \cdots + w_{Bd} X^{d-1}, \tag{20}$$

where $\sharp\{b_i\}$ is the same as $\sharp\{a_i\}$, namely k. Bob then calculates the "syndrome", $w_{A,B(X)}$, as

$$w_{A,B(X)} = w_A(X) + w_B(X). \tag{21}$$

It is easy to see that Bob can decode $S_{A,B}^{(F)}$ and recover the Alice's favorite set, A, based on the syndrome $w_{A,B(X)}$ provided that the relation :

$$D(\boldsymbol{v}_A \backsim \boldsymbol{v}_B) \leqq \frac{d}{2} = e \tag{22}$$

holds. Namely, when Eq.(22) holds, Bob is allowed to join Alice's movie club.

4.3 Size of public data

Let the size of the required public data for any person who wishes to organize a movie club be denoted by S_{pd}. The S_{pd} for Solution JS and Solution A are given by

$$S_{pd} = \begin{cases} 2mN_T = 2m(2^m - 1), & \text{for Solution JS} \\ 2md, & \text{for Solution A} \end{cases}. \tag{23}$$

For example, for $m = 20$, the size of the public data are given by

$$S_{pd} = \begin{cases} 41.9\text{M bit}, & \text{for Solution JS, for any } k \text{ and } d \\ 4\text{K bit}, & \text{for Solution A, for } k = 1000, d = 100 \end{cases}. \tag{24}$$

4.4 Applicability of FSRS(1) for Biometric System

FSRS(1) can be successfully applied to biometric system. In the biometric system applied by FSRS(1), k minutiae $s_1, s_2, \cdots, s_k$ lifted from User A are transformed to the check symbols $u_1, u_2, \cdots, u_d$. The k minutiae are completely deleted from the system and only $u_1, u_2, \cdots, u_d$ are kept in the system. It should be reminded that $H(s_1, s_2, \cdots, s_k | u_1, u_2, \cdots, u_d)$ can be made sufficiently large, yielding a sufficiently secure biometric system.

If k minutiae $\tilde{s}_1, \tilde{s}_2, \cdots, \tilde{s}_k$ are lifted from User A who wishes to be authenticated by the biometric system, the corresponding check symbols $\tilde{u}_1, \tilde{u}_2, \cdots, \tilde{u}_d$ are calculated.

If the Hamming weight of $(s_1 + \tilde{s}_1, s_2 + \tilde{s}_2, \cdots, s_k + \tilde{s}_k)$ satisfies

$$H(s_1 + \tilde{s}_1, s_2 + \tilde{s}_2, \cdots, s_k + \tilde{s}_k) \leqq \frac{d}{2} = e, \tag{25}$$

then $(s_1 + \tilde{s}_1, s_2 + \tilde{s}_2, \cdots, s_k + \tilde{s}_k)$ is completely decoded. User A is then recognized to be the same person who had manutiae $s_1, s_2, \cdots, s_k$ at the registration time.

5 Conclusion

We have discussed that the secret recovering algorithm found by the present author in 2000 can be successfully applied to the solving of the fuzzy vault scheme including MLP. Further works on applying our scheme to biometric system will be reported soon.

References

[1] R.McEliece : "A Public-Key Cryptosystem Based on Algebraic Coding Theory", DSN Progress Report, 42-44, 1978.

[2] R.McEliece and D.V.Sarwate : "On secret sharing and Reed-Solomon codes", Communications of the ACM, 24, 583-584, 1981.

[3] M.Kasahara : "How to recover forgotten password (Challenge Problem)", Newsletter, IEICE, A, 29, p29, Aug. 2000.

[4] M.Kasahara : "A New Class of Product-Sum Cryptosystem – Appending a Solution of Problems Related to Passwords –", SCIS2001, 535-540, Oiso, Japan, Jan. 2001.

[5] M.Kasahara : "How to recover the forgotten secrets information", IEEE, ISIT, 2004.

[6] M.Kasahara : "A Construction of Public-Key Cryptosystem Using Algebraic Coding on the Basis of Superimposition and Randomness", IEICE Trans. Fundamentals, vol. E89-A, No.1, pp.47-54, Jan. 2006.

[7] M.Kasahara, "New Classes of Public Key Cryptosystem Constructed on the Basis of Multivariate Polynomials and Error Control Coding", Technical Report of IEICE, ISEC. 2008-13, 108, 38 (2008-05).

[8] M.Kasahara, "A Construction of Public Key Cryptosystem Using Algebraic Coding on the Basis of Suprtimposition and Randomness", IEICE Trans. Vol. E89-A, 1, pp47-54, (2006-01).

[9] A.Juels and M.Wattenberg : "A Fuzzy Commitment Scheme", ACM CCS, 1999.

[10] A.Juels and M.Sudan : "A Fuzzy Vault Scheme", IEEE ISIT, 2002.

[11] M.Kasahara : "Construction of Error Correcting Codes Based on Multivariate Polynomials of Degree 1", Memorandum for file at Kasahara Lab., Osaka, (2008-12).

[12] T. Hada, M. Morii and M. Kasahara : "On error detecting capability of Reed-Solomon codes", Technical Report of IEICE, IT89-12, pp7-10, 1989.

Appendix I : Ideal Code

Definition A1 : The code over $\mathbb{F}_{2^m}$ that satisfies the following relations will be referred to as Ideal Code.

$$H(\boldsymbol{s}|\boldsymbol{u}) = (k - d)H(s), \tag{26}$$

and

$$H(s_i|\boldsymbol{u}) = H(s). \tag{27}$$

□

Remark A1 : When $m(k - d)$ takes on sufficiently large value, $\boldsymbol{u}$ can be publicized without deteriorating any security level on an individual secret s_i. □

Appendix II : Construction of Error Correcting Codes Based on Multivariate Polynomials of Degree 1

Let us construct the following simultanious equations over $\mathbb{F}_p$:

$$\begin{aligned} f_1 m_1 + f_2 m_2 + \cdots + f_k m_k &= c_1 \mod p, \\ f_1^2 m_1 + f_2^2 m_2 + \cdots + f_k^2 m_k &= c_2 \mod p, \\ &\vdots \\ f_1^d m_1 + f_2^d m_2 + \cdots + f_k^d m_k &= c_d \mod p, \end{aligned} \tag{28}$$

where we assume that f_i's are all distinct. Let the vector $\boldsymbol{v}$ be given as

$$\boldsymbol{v} = (m_1, m_2, \cdots, m_k, c_1, c_2, \cdots, c_d). \tag{29}$$

It is easy to see that the set of all the possible $\boldsymbol{v}$, $\{\boldsymbol{v}\}$, constitutes a linear code. The linear code will be referred to as $\boldsymbol{v}$ code. Let us denote $(m_1, m_2, \cdots, m_k)$ and $(c_1, c_2, \cdots, c_d)$ as $\boldsymbol{m}$ and $\boldsymbol{c}$, respectively.

Theorem A1 : The $\boldsymbol{v}$ code is a linear code of minimum distance $d+1$.
Proof : Let the Hamming weight of vector $\boldsymbol{u}$ be denoted as $w(\boldsymbol{u})$. Assume that $w(\boldsymbol{m}) = h$ and that $w(\boldsymbol{c}) = d - h - a$ holds for $a \geq 0$. Without loss of generality, we assume that $a = 0$ and $m_1, m_2, \cdots, m_k$ are all nonzeros. We also assume that, without loss of generality, $c_1, c_2, \cdots, c_h$ are all zeros. If $c_1 = c_2 = \cdots = c_h = 0$, then $m_1, m_2, \cdots, m_h$ are all required to be zeros, which is a contradiction, yielding the proof. □

Remark A2 : Although the details of doing so are omitted, we can show that $\boldsymbol{v}$ code over $\mathbb{F}_p$ where $p = 2^m - 1$ (Mersenne Prime) is an ideal code and in addition it well suits the practical use. Besides it is the maximum distance separable codes.

Remark A3 : The RS code is ideal in a sense that it is the maximum distance separable code. Besides it is extensively used in the various storage and transmission systems. Unfortunately the RS codes do not exactly satisfy Eqs.(28) and (29) [12]. However as m increases the RS code asymptotically meet Eqs. (28) and (29) [12].

It should be noted that the RS code yields theoretically almost the best performance for a large m ($m \gtrsim 20$) [12].