Secure Secret Recovery by using Weighted Personal Entropy

Leau Y. B., Dinna Nina M. N., Habeeb S. A. H., and Jetol B.

Abstract—Authentication plays a vital role in many secure systems. Most of these systems require user to log in with his or her secret password or pass phrase before entering it. This is to ensure all the valuables information is kept confidential guaranteeing also its integrity and availability. However, to achieve this goal, users are required to memorize high entropy passwords or pass phrases. Unfortunately, this sometimes causes difficulty for user to remember meaningless strings of data. This paper presents a new scheme which assigns a weight to each personal question given to the user in revealing the encrypted secrets or password. Concentration of this scheme is to offer fault tolerance to users by allowing them to forget the specific password to a subset of questions and still recover the secret and achieve successful authentication. Comparison on level of security for weight-based and weightless secret recovery scheme is also discussed. The paper concludes with the few areas that requires more investigation in this research.

Keywords—Secret Recovery, Personal Entropy, Cryptography, Secret Sharing and Key Management.

I. INTRODUCTION

AUTHENTICATION is an important element in many secure systems. Users are not allowed to enter the system before their identities have been verified. Due to this, many encryption systems require users to memorize high entropy password or pass phrase and remember them when it gets lost. As a result, users either prefer to select easy-to-remember passwords that are easy to crack or even use an insecure method of remembering the password such as writing it down or storing it in obvious places. To solve this dilemma, Ellison and Bruce suggest replacing a single, long pass phrase with multiple small ones, tied to life experiences, free association and etc.

Unfortunately, in their scheme, the larger the user chooses the number of correct answers required to reconstruct the secret, the more attempt is required to recover the secret. The scheme assigns a weight to each entrenched-knowledge parameters. A weight-based scheme for sharing and reconstructing the secret and achieve successful authentication. The main objective for this weight-based scheme is to enhance the level of security in personal entropy-based scheme and minimize the workloads on secret recovery. The notions of the proposed scheme are as below:-

1) A weight will be assigned to each question based on its category.
2) A weight-based scheme for sharing and reconstructing the secret with personal entropy will be created.
3) The scheme provides higher security by three conditions as mentioned below:
   a) A proper set of questions must be chosen.
   b) The total of weight in this proper set of questions must be equal or bigger than the ideal weight, \( W_i \geq W_t \).
   c) The given answer for each question must exactly match with their corresponding answer during registration.
4) The scheme is flexible to be used in different applications which require diverse security level by adjusting the information rate in calculating the ideal weight.

II. RELATED WORK

Secret key management is an essential point to ensure data confidentiality. Kerckhoff’s law states that “only secrecy of the key provides security”. This principle clearly explains that proper management of cryptographic keys is essential to the effectiveness of cryptography in security. Loss or corruption of the only copy of secret keys can deny users access to information. Hence, key escrow is very important in this context. According to Hal Abelson [10], “key escrow” is also called “key recovery” and generally referred to any system for assuring third-party access to encrypted data. This term was first introduced by the United States Government in a program called “clipper” in April 1993. This mechanism requires individuals to split their secret encryption keys into two
pieces, and hand them over to escrow agents which are normally trusted third parties chosen by the government. This idea was soon prolonged by Shamir and Blakey by designing the secret sharing scheme.

A. Secret Share Scheme

Shamir [1] proposed a secret share scheme called \((t, n)\) threshold scheme which is based on polynomial interpolation. His goal is to divide data \(D\) into \(n\) pieces \(D_0, D_1, \ldots, D_n\) in such a way that:

1) Knowledge of any \(t\) or more \(D_j\) pieces makes \(D\) easily computable;
2) Knowledge of any \(t-1\) or fewer \(D_j\) pieces leaves \(D\) completely undetermined in the sense that all its possible values are equally likely.

Alternatively, Blakey [9] has proposed another scheme based on probabilistic approach on linear projective geometry over finite field. The dealer will pick a one-dimensional flat, based on probabilistic approach on linear projective geometry and a \(\mathbb{F}_q^2\)

His goal is to divide data \(D\) into \(n\) pieces \(D_i, D_2, \ldots, D_n\) in such a way that:

1) Knowledge of any \(t\) or more \(D_i\) pieces makes \(D\) easily computable;
2) Knowledge of any \(t-1\) or fewer \(D_i\) pieces leaves \(D\) completely undetermined in the sense that all its possible values are equally likely.

B. Techniques of Authentication

Basicly, there are four techniques to authenticate the legitimate participants in recovering a secret key [5]. First is the personal identification technique, where participants physically present themselves to the secret key administrator. Second fax documentation technique, in which a copy of an official identification such as an ID card or passport transmitted via facsimile machine to the secret key administrator. Third, encrypted email recovery techniques where participants normally provide a public key during registration and send their recovery emails encrypted with that key. Finally, question and answer technique where participant’s personal information were recorded during account registration. Compared with a password, personal knowledge questions and answers are widely used because memorizing passwords are easily forgotten while entrenched knowledge is not [14]. Moreover, these questions could be very personal to enhance security level.

C. Secret Recovery Scheme

By applying the technique of question and answer, Ellison et al and Frykholm proposed their secret recovery schemes which are based on error-correcting and fuzzy commitment respectively. In 1999, Ellison et al [4] proposed a method where a user can protect a secret key using personal entropy of his own life by encrypting pass phrase of answers to several entrenched-knowledge questions. This method is based on Shamir’s Secret Sharing Scheme [1], \((t, n)\)-threshold scheme where a secret is distributed into \(n\) shares of which at least \(t\) of these are required to reconstruct the secret. By using the hash functions of the entrenched-knowledge questions and answers, \(n\) secret will be encrypted and decrypted. The scheme emphasizes on user fault tolerance where users are still able to achieve successful authentication and thus recover the secret even though they had forgotten the answers to some number of questions. Frykholm and Juels [17] had presented another approach which is almost similar to the one proposed by Ellison et al, deriving a strong secret key from a sequence of answers to entrenched-knowledge questions. The main difference is their method is based on the fuzzy commitment technique [3] which accepts a witness that is close to the original encrypting witness in a suitable metric but not necessarily identical. Both of their approaches are dispenses with the use of trusted third parties. In principle, the users can safely store their ciphertext even in a public directory. Users are not likely to forget answers that are more connected to personal fact, therefore, the chance of reconstructing secret or secret key is dreadfully high.

Even Frykholm and Juels [17] have stated that the Ellison et al. scheme strikes an attractive balance between convenience and security but there are some limitations in Ellison’s methods. Bleichenbacher and Nguyen [6] shown that multiple choices of parameters in Ellison’s method is insecure. From Frykholm’s view [17], the question and answer system is as secure as the entropy of the answer. But Ellison’s method is lack of rigorous security analysis because the method employs an ad hoc secret sharing scheme in the system that undermines conventional security guarantees. In addition, Ellison also mentioned that more workflows are required in recovering the secret when a larger number of required correct answers, \(t\) are chosen [4].

As a consequence, the objective of this new scheme is to apply weighted personal entropy in Ellison’s concept in the aim to solve existing problems and thus enhance security level in secret sharing and recovery.

D. Weighted Threshold Functions

Weighted threshold secret sharing was introduced by Shamir in his seminal work on secret sharing [1]. In this scheme, users are not of the same status. There are a set of users where each user is assigned with a positive weight. A dealer wishes to distribute a secret among those users so that a subset of users may reconstruct the secret if and only if the sum of weight of its users exceeds a certain threshold.

A secret sharing scheme is ideal if the domain size of shares for each user is at least as large as the domain of possible secrets [19]. In this case, Shamir’s threshold secret sharing scheme could be considered ideal in the sense that the domain of shares for each user correspond with the domain of possible secrets.

The concept of weighted threshold functions was also applied by Ito, Saito and Nishizeki [16]. They generalized the concept of secret sharing as an arbitrary monotone collection of authorized sets called access structure. Only set requirements in the access structure are allowed to reconstruct the secret, while sets that are not in the access structure should gain no information of the secret.

Additionally Beimel, Tassa and Weinreb [2] stated that a weighted threshold access structures is ideal if and only if it is
a hierarchical threshold access structure as introduced by Simmons [8] or a tripartite access structure or a composition of two ideal weighted threshold access structures that are defined on smaller sets of users.

III. WEIGHT-BASED SECRET RECOVERY SCHEME

The new scheme is a weight-based scheme for recovering secret with personal entropy. It is similar to previous papers mentioned but is slightly different in focusing on weighted personal entropy and calculated ideal weight.

A. Weight in Various Categories of Question

Challenging questions are always used as hints for secret recovery. Generally, there are nine types of questions which are Bipolar question (i.e., Yes/No, True/False, Agree/Disagree), Multiple-choices question, Fill in the blank questions, Which, What, Who, When, Why and How questions. Each type of question can be differentiated from the length of possible given answer and also on difficulties in answering the question correctly. For authentication reason, users are required to provide a correct answer for each question before successfully recover the secret.

Basically, questions can be open-ended or close-ended. Some questions have answers which fall along an implied range (rating scales), others have answers in no particular order (lists). Other questions have multiple-choices (check all that apply) options and others provide relevant answer choices but respondents are free to add different answer [20]. In this context, all of these questions can be categorized into three, open-ended question, close-ended question and bipolar/multiple choices question. Open-ended questions are those where respondents provide their own answers to the questions without any previous options. These questions give the breadth and depth of reply and are difficult to analyze. The closed-ended questions restricted the respondent’s options. This is because the respondents can only reply with a finite number such as “None” or “One”. The bipolar and multiple choices questions are special kinds of closed question. These types of questions limits the respondents even more further by only allowing a choice on either pole such as “yes/no”, “true/false”, “agree/disagree” or multiple-choices such as “Monday/Tuesday/Wednesday/Thursday/Friday/Saturday”. The respondent are not allowed to write down their response but is still considered correct in answering the question [11].

In this new scheme, it is assumed that the system will provide these three types of questions which can be compared based on their characteristic such as precision of data, breath and depth data, and ease of analysis of data. Precision of data refers to the ability of respondents in answering to that particular question. The breadth and depth data refers to the level of difficulties in answering that particular question, and ease of analysis of data refers to the ability of attackers or hackers to analyze the correct answer of that particular question. These three types of questions can then be further categorized into weighted-based questions regarding to the level of entropy. Classifications of questions and their weight values are listed in Table I [11] and II.

Table II shows that open-ended question is assigned with the weight value 3 because of having the lowest precision of data, the most breath and depth and is the most difficult to analyze. For bipolar and multiple choices question, its weight value is 1 as it is high precision of data, limited breath and depth and is the easiest to analyze. For close-ended question, which is ranked between open-ended question and bipolar/multiple choices question, has weight value of 2 each.

In this paper, it is assumed that each question is assigned with a weight value (example: 1, 2 and 3). These assumed values are used since they are the most common divisors for any number, which thus will provide some significant advantages for the scheme. From Table II, weighting the questions according to their characteristic does not mean it provides another alternative for attackers to hack into the system, but it is believed that it will lead to a noteworthy confusion effect to attackers who are trying to guess and analysis the proper set of questions based on the value of ideal weight.

B. Weight-Based Secret Recovery Scheme

Encryption and decryption process are important in secret sharing scheme to protect the secret between two entities (i.e., sender and recipient). To enhance security level, McCurley [15] has introduced an encryption scheme which is based on Diffie-Hellman problem and RSA. This scheme is similar to ElGamal scheme but it works in a subgroup of $Z_N$ where $N$ is a special-form of composite number. The user needs to produce a module $N = pq$, where $p$ and $q$ are two large primes having criteria as below [13]:

1) $p = 3 \pmod{8}$ and $q = 7 \pmod{8}$
2) $(p - 1)/2$ and $(q - 1)/2$ are primes
3) $(p + 1)/4$ and $(q + 1)/8$ have large prime factors.

Briefly, in the McCurley scheme a composite modulus $N$ is formed as the product of two primes $p$ and $q$ of rather

<table>
<thead>
<tr>
<th>Types of Question</th>
<th>Precision of Data</th>
<th>Breath and Depth</th>
<th>Ease of Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bipolar/Multiple Choices</td>
<td>High</td>
<td>Little</td>
<td>Easy</td>
</tr>
<tr>
<td>Close-ended</td>
<td>Moderate</td>
<td>Moderate</td>
<td>Moderate</td>
</tr>
<tr>
<td>Open-ended</td>
<td>Low</td>
<td>Much</td>
<td>Difficult</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Types of Question</th>
<th>Weight Value (Assumption)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bipolar/Multiple Choices</td>
<td>1</td>
</tr>
<tr>
<td>Close-ended</td>
<td>2</td>
</tr>
<tr>
<td>Open-ended</td>
<td>3</td>
</tr>
</tbody>
</table>
particular construction. Basically both are being safe primes, one congruent to 3 mod 8 and the other congruent to 7 mod 8. A safe prime \( p \) is defined to be one for which \( (p - 1)/2 \) is also a prime. Therefore, Diffie-Hellman [18], scheme was provably secure since it is based on intractability of factoring.

This Weight-based Secret Recovery Scheme applies the factoring features of McCurley scheme to generate \( P_i \) and \( Q_i \) in order to encrypt and decrypt the secret which involves a set of weighted entrenched-knowledge questions and answers.

The following is a description of this weight-based secret recovery scheme for sharing and reconstructing the secret message. It is assumed that a legitimate set of questions and proper value for \( n \) (number of questions), has been chosen.

The scheme for sharing the secret message \( M \) is:

1) The user is asked to choose a random number, \( s \) and \( n \) questions \( q_1, \ldots, q_n \) to generate answers \( a_1, \ldots, a_n \).
2) Then using McCurley encryption scheme to compute \( y_i = 16^s \mod N_i \) where \( N_i = P_i Q_i \) for \( i \in \{1, 2, \ldots, n\} \). \( P_i \) and \( Q_i \) will be randomly generated by the system to each question, \( q_i \) and its corresponding answer \( a_i \). Value \( P_i \) and \( Q_i \) must fulfill the criteria as mentioned in McCurley encryption scheme.
3) Simultaneously, weight value, \( w \) will be assigned to each question chosen by the user and the system will accumulate the value of \( W = w_1 + w_2 + \ldots + w_n \).
4) Then, value \( W_t \) ideal weight will be calculated by using the information rate formula that will be discussed in section 3.3.
5) A \((t, n)\)-threshold scheme is used to split the secret message \( M \) into \( n \) secret shares \( M_1, \ldots, M_n \).
6) Encrypt each share \( C_i = M_i y_i^{W_i} \mod N_i \), \ldots, \( C_n = M_n y_n^{W_n} \mod N_n \) and send \( C_n \), ciphertext to \( n \) user.
7) User will remember \( q_1, \ldots, q_n, s \), and \( C_1, \ldots, C_n \).

The scheme for reconstructing the secret message \( M \) is:

1) The user is asked to enter his or her random number, \( s \) and choose proper \( t \) questions \( q_1, \ldots, q_t \) to answer in order to generate the \( a'_1, \ldots, a'_t \).
2) Only accurate \( s \) and proper question-answer will give the exact value \( y_1, \ldots, y_t \) for decrypting the \( C_1, \ldots, C_t \).
3) The total of weight for \( t \) questions, \( W_t \) will only be calculated from those matched answer given by the users during registration.
4) For any value of \( W_t \geq W_i \), the scheme will use \( W_t \) in calculating \( M_1, \ldots, M_n \)

\[
W_t = \begin{cases} \text{W}_i, & \text{if and only if } W_t \geq W_i \\ W_i, & \text{for any } W_t < W_i \\
\end{cases}
\]

(This is a significant security feature when user needs to answer the proper questions in order to get the exact value of \( M_1 \), resulting from value of \( W_t \geq W_i \).
5) Decrypt each ciphertext \( C_i \) using the corresponding \( N_i \) and will get the \( M_1 = C_i y_i^{-W_i} \mod N_i \), \ldots, \( M_n = C_n y_n^{-W_n} \mod N_n \).

There are some differences in calculating the value of \( W_t \) in sharing and reconstructing secret respectively. In a sharing scheme, \( W_t \) is accumulated from questions and answers that are given by the user. But in the reconstructing process, \( W_t \) will only be accumulated from each pair of correct question and matched answer given during registration. By using the \((t, n)\)-threshold scheme, select subsets of \( t \) shares til the secret message \( M \) is correctly reconstructed by the exact value of \( W_t \).

In other words, to successfully recover the secret, at least \( t \) of the questions are needed to be answered correctly in the condition of \( W_t \geq W_i \), where \( W_t \) is the total weight of \( t \) questions which user had already answered correctly to a set of proper questions, and \( W_i \) is the ideal weight for the scheme.

Fail to answer more than \( t \) questions correctly will unable the user to recover the secret. Basically, each piece of share \( M_i \) will not be recovered by a correct answer if \( W_t < W_i \).

C. Calculating \( W_t \)

The value of \( W_t \) is calculated from the value of \( W_t \). To calculate the proper \( W_t \), a formula for calculating information rate will be applied. By a given secret sharing scheme in which \( S \) is the set of possible secrets and \( T \) is the set of possible shares, then information rate, \( \rho \) of the scheme from [7] is as below:

\[
\rho = \log |T| / \log |S|
\]

\( Simmon [8] \) describes a secret scheme extrinsic if the set of possible shares is the same for all participants. It is confirmed that a secret sharing scheme is ideal if it is perfect and has information rate 1.

As a result, to calculate \( W_t \) in this new scheme, the information rate needs to be set regarding to the requirement of the application. The formula is as below:

\[
\rho = \log |W_t| / \log |W_t| , \text{ where } 0 \leq \rho \leq 1
\]

Hence, the value \( W_t \) can be differentiated in various applications by determining different information rates. This feature makes this new scheme more flexible in meeting the security level requirements in different situations. In this scheme, users can only succeed in reconstructing their secret if and only if \( W_t \geq W_i \), but they will fail to do so if \( W_t < W_i \). For any value of \( W_t \geq W_i \), only the value of \( W_t \) will be used in calculation for recovery this is because the above equation (2) is only true when \( W_t = W_i \).

IV. ANALYSIS ON OTHER WORKS AND THE NEW SCHEME

Previously, Ellison had proposed a secret recovery scheme called personal entropy to encrypt secrets or passwords through answers to several entrenched-knowledge and personal questions [4]. In his scheme, every question is considered the same attributes, regardless of its difficulties in answering. In other words, no weight value assigned to each question and is all considered the same. Therefore, the larger number user chooses \( t \), the more work is required to recover the secret for both user and attacker. The users need to do extra work in the form of extra typing where it is approximately linear to \( n \) which always increase as \( t \) increase.

For more secrecy, \( t \) will be increased so will \( n \) since \( n \neq t \). This eventually increases the keystrokes (workloads) [4]. Dissimilarity, the new scheme assigns a weight value to each
personal question based on their attributes to minimize workloads on secret recovery and simultaneously enhances its security level.

Fig. 1 shows the obvious differences in level of security for Ellison’s scheme and this new scheme. Assume $p = 0.95$ and each question has the same weight value which is 2, the result are as below.

![Fig. 1 Comparison between this Scheme and Ellison’s Scheme](image)

The tremendous feature of this graph is the number of correct questions required in this new scheme that outnumbers Ellison’s scheme in all numbers of questions, $n$. As for example in Ellison’s scheme, assuming $n = 22$, $t = 14$, a user could answer all 22 prompts but might make some errors. Then the algorithm must search for a correct subset of 14 shares to use in order to recover the password [4]. In contrast, for the new scheme, when $n = 22$, users need to answer correctly the subsets of 18 shares in order to reconstruct their password. The gap between the number of correct questions in Ellison’s scheme and this new scheme for each set of questions, $n$ is roughly 4 to 5 questions. By comparing with the result from Ellison’s scheme, this significant difference indicates that users must answer more correct questions to recover their password using this new scheme. Meanwhile, an attacker must go through a lot of answers to a large subset of questions in order to recover the secret. Straightforwardly, the result shows that by assigning a weight value to personal entropy and calculating the ideal weight by applying a formulation of information rate, this new weight-based scheme could enhance the level of security of previous personal entropy-based schemes. Therefore, this finding has achieved the main objective of this research.

Simultaneously, to enhance the level of security, this new scheme also minimizes the workloads on secret recovery. It focuses on the weight value of each personal question. Every question is assigned with a desired weight based on their attributes as discussed in Section 3. Therefore, although the number of correct questions, $t$ required in this scheme is higher than Ellison’s scheme, it doesn’t mean it will burden the user in recovering their password. This is because, assumptions of value for each question were made. However in the real situation, some questions might have different values. In this scheme the sum of weight value for this particular set of questions will absolutely reduce the number of questions that needs to be answered correctly when reconstructing the secret. Unfortunately, in Ellison’s scheme it does not differentiate the question’s level of breadth and depth as well as their precision of data and ease of analysis, the value of $t$ is fixed and unchangeable, depending on value $n$.

V. FUTURE WORK

This scheme’s aim is to provide more secure weight-based secret recovery scheme. This paper only serves to demonstrate the weight-based scheme concept and a number of findings, but there are a few areas that require more investigation.  
1) Designing a set of challenging entrenched-knowledge questions is a difficult task and has nothing to do with the system’s cryptography. It is an interesting area of psychological research.  
2) Research needs to be done on actual entropy from the attacker’s point of view in answering the entrenched-knowledge questions.  
3) Additional empirical research is required for the actual value weight for each type of question. In this paper, it has been assumed that each question is assigned with weight values of 1, 2 and 3 but no more than a little empirical basis.

VI. CONCLUSION

In this paper, a more secure and flexible weight-based scheme in recovering the secret with personal entropy is presented. The objective of this scheme is to enhance the security level and minimize the workloads in secret recovery. This scheme assigns a weight to each personal entropy and as long as the total of weight from all the questions which user had answered correctly is equal or bigger than the ideal weight, then the secret recovery is successful. It avoids the attacker from suspecting the question-answer based on its entropy and thus brute force them. Although this scheme is far from being perfect and has a few issues to consider such as designing sets of challenging entrenched-knowledge questions and finding the actual value of weight for the particular question, it has provide higher security level with minimum workloads and flexibility for different applications.

REFERENCES


