High-Entropy 2-Dimensional Key Input Method for Symmetric and Asymmetric Key Cryptosystems

Kok-Wah Lee @ Xpree Li, *Graduate Student Member, IEEE*

***Abstract* — conventionally, single-line key field is used to input a key. The selection of a key depends on the factors of memorizability and security. The minimum key sizes for symmetric and asymmetric key cryptosystems are 80 and 160 bits, respectively. Advanced Encryption Standard (AES) sets the key sizes at 128, 192, and 256 bits. The asymmetric key cryptosystems that require a minimum key size of the private key are finite field cryptography (FFC) and elliptic curve cryptography (ECC). The corresponding sizes of private keys to the AES are 256, 384, and 512 bits, respectively. Here, we propose a 2-dimensional (2D) key input method to create high-entropy keys. The 2D key has the styles of multiline passphrase, crossword, ASCII art, colorful text, or spatial variation. It can resist dictionary attack and fulfill the demands of human-memorizable key sizes even until 256 bits, which is impossible by using the single-line passphrase. Special Chinese-character-encoded passphrase generates 13 ASCII characters for every Chinese character. 2D key also allows the creation of memorizable public-key cryptosystem (MePKC), where the private key is memorizable. MePKC provides mobility, lower cost and higher efficiency, which is better than the encrypted private key, split private key, and roaming private key.**

***Keywords* — User interfaces, human factors, memorizability, key/password security, Chinese-character-encoded passphrase, big memorizable secret, memorizable public-key cryptosystem (MePKC) (aka MoPKC - mobile public-key cryptosystem).**

I. Introduction

There are three main types of authentication methods: What you know, what you have, and what you are. Their representatives are secret, token, and biometrics, respectively. Among them, due to the factors of cost and compatibility, secret in the form of key is the most popular method. A key is a piece of information to control the operations of cryptographic algorithm. Short key is called password and long key is called passphrase. The selection of a key is always the balance of the factors of memorizability and security.

According to Kerckhoff's Law [1], a cryptosystem shall depend 100% on the secrecy of key only. In the words of Shannon's Maxim, it means "enemy knows the system". This law makes the civilian cryptosystem to have publicly known algorithm except the classified governmental and military information. This is needed to gain the public confidence for general daily applications from the fear of possible backdoor.

For symmetric key cryptosystem, National Institute of Standards and Technology (NIST) of USA proposed security level of 80-bit key to be phased out by year 2015 and used until year 2010 [2]. US government has an export policy to control the power of cryptographic algorithm by setting the maximum key size. The current export limit of symmetric key size has been raised from 40 bits to 128 bits.

For the symmetric key cryptosystem of Advanced Encryption Standard (AES), there are three key sizes: 128, 192, and 256 bits. The asymmetric key cryptosystems, which demand for the minimum private key size at 160 bits by year 2010, are finite field cryptography (FFC) and elliptic curve cryptography (ECC). FFC and ECC are based on the mathematical hard problems of discrete logarithm problem and elliptic curve discrete logarithm problem, respectively. The corresponding sizes of private keys to the AES are 256, 384, and 512 bits, respectively [2]-[3]. The symmetric key is normally remembered by brain; whereas the asymmetric private key is encrypted using another symmetric key.

ASCII characters have an entropy of 6.57 bits per character. Therefore, the nominal bit of an ASCII character is 8 bits, but its effective bit is 6.57 bits. To cater for the different symmetric key sizes at 80, 96, 112, 128, 192 and 256 bits [2]-[3], we need 13, 15, 18, 20, 30, and 39 ASCII characters, respectively. An amount of 15 ASCII characters is perhaps still affordable and convenient for the human users. However, higher amounts may introduce two problems. Memorizability is the main problem. The difficulty to type a long passphrase into a computer will be another open problem [1].

In this paper, we propose a high-entropy key input method to solve these problems. This method uses a 2-dimensional display as the user interface to improve the human factors of memorizability and input of ASCII characters from keyboard. It is called 2-dimensional (2D) key.

In addition to fulfilling the various key sizes of symmetric key cryptosystem, 2D key has novel revolution to the private key storage of asymmetric key cryptosystem. For the prior arts, we have encrypted private key, split private key, and roaming private key. With the introduction of 2D key, there shall be no more need to store the private key in a computing system, but inside the brain as like the symmetric key. This

Manuscript received December 09, 2008; accepted March 19, 2009.

Kok-Wah Lee is with the Faculty of Engineering & Technology, Multimedia University, Jalan Ayer Keroh Lama, 75450 Bukit Beruang, Melaka, Malaysia (corresponding author to provide phone: 606-252-3320; fax: 606-231-6552; email: kwlee@mmu.edu.my; contact@xpreeli.com; E96LKW@hotmail.com).

allows the creation of memorizable public-key cryptosystem (MePKC). MePKC has the special features of mobility, lower cost and higher efficiency.

Section II discusses on the related works and prior arts. Sections III and IV focus on 2D key and its styles, respectively. Section V elaborates the application of Chinese-character-encoded passphrase for the styles of multiline passphrase. Sections VI give the applications of 2D key for symmetric and asymmetric key cryptosystems. Section VII ends the paper with a conclusion.

II. Related Works and Prior Arts

A. Single-Line Key/Password Field

Conventionally, when we use secret as an authentication method, single-line key field will be the area for a user to enter a key. For the current longest possible key, it is a single-line passphrase. For passphrase, it can be formed from acronym, sentence, diceware, and coinware [4]. Nevertheless, there is a limit due to the problems of memorizability and ASCII character input from keyboard. The first problem is due to the human factor; whereas the second is due to the user interface. These problems prohibit the applications of symmetric key sizes at higher security levels whenever a user cannot remember and/or conveniently enter a long single-line passphrase.

B. Environ Password

Good memorizability exists when it is linked to a learnt language. For English language, U.K. government introduced the case insensitive Environ password in October 2005 for short-term protection [5]. It has an 8-character key pattern as in Table I. This pronounceable password has 34.9 bits per unit.

TABLE I: Environ Password

Form	c = consonant, v = vowel, d = digit : c – v – c – c – v – c – d – d
Example	pinray34, yankan77, supjey56, kinkin99

C. Chinese Input Methods, Chinese Character Encodings, and Memorizability

Chinese character is also called Han character, where it is used by CJK languages. The Chinese input methods are either based on pronunciation, character structure, or a combination of pronunciation and character structure [6]. These methods are closely linked to Chinese character encodings [7] to allow a user to enter a Chinese character. A user normally remembers the pronunciation and/or character structure of a Chinese character to facilitate its input.

Reference [8] is a Chinese character encoding to ease Chinese input by using the combination of pronunciation and character structure. This kind of Chinese character encoding can create a key per Chinese character. The maximum size of this encoding is six characters, where there are three characters for phonetic sound, one character for tone, and two characters for character structure. The memorizability of this Chinese-character-encoded key is better than the Environ password, but its security is subject to dictionary attack.

D. Key Sizes of Symmetric Key Cryptosystems

Reference [3] provides a table in page 29 to relate the symmetric key sizes to the protection periods. This table is briefly shown again in Table II for easy reference.

TABLE II: Minimum Symmetric Key Sizes for Different Levels of Protection

Security Level	Security (bits)	Protection
1	32	Only acceptable for authentication tag size.
2	64	Very short-term protection.
3	72	Short-term protection.
4	80	Smallest general-purpose protection for ≤ 4 years.
5	96	Legacy standard level for 10-year protection.
6	112	Medium-term protection for 20 years.
7	128	Long-term protection for 30 years.
8	256	Foreseeable future. Good protection against quantum computers.

For the AES suggested by NIST to replace the Data Encryption Standard (DES), it has three types of symmetric key sizes. These key sizes are 128, 192, and 256 bits. Therefore, we have AES-128, AES-192, and AES-256 to fulfill the demands of security levels at 128, 192, and 256 bits [2]. For other security levels at 80 and 112 bits, NIST suggested two-key Triple Data Encryption Algorithm (2TDEA) and three-key Triple Data Encryption Algorithm (3TDEA), respectively [2].

E. Key Sizes of Asymmetric Key Cryptosystems

There are 3 conventional mathematical hard problems used in asymmetric key cryptosystem, which is also called public-key cryptosystem. These problems are integer factorization problem, discrete logarithm problem, and elliptic curve discrete logarithm problem. NIST categorizes the applications of these problems for public-key cryptography as integer factorization cryptography (IFC), finite field cryptography (FFC), and elliptic curve cryptography (ECC), respectively.

IFC has a long key size for public and private keys. FFC has a long public key and a short private key. ECC has a short key size for public key and private key. The minimum asymmetric key sizes for IFC, FFC, and ECC in equivalent with the security levels of symmetric key sizes are shown in Table III [2].

TABLE III: Minimum Asymmetric Key Sizes in Equivalent with the Security Levels of Symmetric Key Sizes

Security (bits)	IFC		FFC		ECC	
	Public	Private	Public	Private	Public	Private
80	1024	1024	1024	160	160	160
112	2048	2048	2048	224	224	224
128	3072	3072	3072	256	256	256
192	7680	7680	7680	384	384	384
256	15360	15360	15360	512	512	512

F. Storages of Private Key

For the current asymmetric key cryptosystem, a private key is normally encrypted using another symmetric key. The encrypted private key is stored in a local computing system or token; whereas the symmetric key is stored in the human brain. The present possible attacks for this method are guessing attack, dictionary attack, and precomputation attack.

Another method is to split the private key into two or more portions [9]-[11]. The first portion of the private key can be derived from a normal human-memorizable symmetric key. The other portions of the private key are stored as encrypted partial private key alike the normal encrypted private key. This method resists the precomputation attack.

A third method is to store the encrypted private key in a server connected to a computer communication network [12]-[13]. A user has the roaming capability where the encrypted private key can be downloaded from the server for decryption at anywhere. Proxy servers are needed for this method to avoid single point of failure. Its possible attacks are the same as encrypted private key stored in the local computing system.

G. Key Strengthening

Key strengthening is also called key stretching. It is used to make a weak key stronger. There are two forms of key strengthening. One uses password supplement [14]-[15], and another uses many rounds of hash iterations [16]. In this paper, key strengthening is applied in the MePKC to achieve larger protection period.

$$S = n * L * R / P \qquad (1)$$

S = Key space

n = Number of networked computers

L = Maximum lifetime of a key in years

R = Number of guesses per unit of time per unit of computer

P = Probability that a key can be guessed in its lifetime

Typical values:

$n = 10^9$ units = 29.9 bits

L = 4, 10, 20, 30, 300 years = 2, 28.2, 29.2, 29.8, 33.1 bits

$R = 1.5 \times 10^7$ s^{-1} = 23.8 bits (best performance in year 2005)

$R = 1$ s^{-1} = 0 bit (using key strengthening)

$P = 10^{-6}$ = -19.9 bits

Equation (1) is a password length equation. When key strengthening is used, R becomes 1 guess per second and the variety of computer is a main factor to set the number of hash iterations. The computer performance of a variety of computers varies from 1 time for the slowest computer to 20 times for the fastest computer. This contributes a factor of $\log_2 20$ = 4.3 bits to (1). Moore's Law states that the number of transistors on an integrated circuit for minimum component cost doubles every 24 months [17].

$$S = (n * L * R / P) * 2^{4.3} * 2^{L/2} \qquad (2)$$

When the variety of computers and Moore's Law are considered, it becomes (2). From (2), key strengthening can make a weak key to become 19.5 bits stronger.

III. 2-Dimensional Key Input Method

For single-line passphrase, the numbers of ASCII characters for different symmetric key sizes are shown in Table IV. An amount of 15 ASCII characters is a memorabilty limit for many human users. The difficulty of user interface to enter a key using keyboard into the single-line key field is another big problem.

TABLE IV: Numbers of ASCII Characters for Various Symmetric Key Sizes

Symmetric key size (bits)	80	96	112	128	192	256
Number of ASCII characters	13	15	18	20	30	39

The problems of human factor and user interface limit the practical application of symmetric key cryptosystem to be at the key size of 96 bits with 10 years of protection. Using key strengthening, the 96-bit key can be made 19.5 bits stronger, and 20-year protection is the maximum theoretical limit.

The 2-dimensional (2D) key input method is created to allow keys with higher entropy. It tries to solve the human factor of memorizability and user interface of key input. 2D key has a 2-dimensional display alike a 2D matrix, where each character of a key is an element of the matrix.

The font used for 2D key has to be fixed-width font [18]. Fixed-width font is also called non-proportional font and monospaced font. It is a typeface using fixed width for every glyph. Examples of fixed-width fonts are *Courier* for ASCII and *MS Mincho* for Unicode. When ASCII encoding is used, the 2D key has 6.57 bits per character. Meanwhile, when Unicode is used, it has 16 bits per character. Even though Unicode-based 2D key has higher entropy, it is inconvenient to enter a Unicode symbol for the mean time, and the fixed-width font for all the Unicode symbols has not yet been created. Hence, ASCII-based fixed-width font is used currently for the discussions as well as prototype.

Firstly, a user needs to select the row size and column size of the 2D matrix for 2D key. The current built prototype has a maximum row size or height of 10 characters, and a maximum column size or width of 13 characters. The column size is set at 13 due to the Chinese-character-encoded passphrase proposed in Section V has a maximum size of 13 per Chinese character. Alternatively, it can be a word in English language or other languages that has a size of 13 characters per word.

We can have many input styles of 2D key. There are multiline passphrase, crossword, ASCII art, Unicode art, colorful text, and spatial variation. Multiline passphrase, crossword, and ASCII art are currently implemented in the prototype; whereas Unicode art, colorful text, and spatial variation require additional supports.

After selecting the row size and column size, the user can input ASCII characters using keyboard as the elements of the 2D matrix. The input characters can have any style or a mixed style of 2D key. These styles have good memorabilty, and the 2D nature of 2D key generates more references at the user interface for key input. Single-line key field has only one reference at the first location of the only line. 2D key has a number of horizontal lines and each first location of the horizontal lines acts as references for key input. In addition, the first locations of the vertical lines can be secondary set of references for key input. This solves the problem of user interface in facilitating a user to enter a high-entropy key.

Good memorizability allows a user to repeat a high-entropy key. The elements of 2D matrix can be either fully or partially filled. The characters entered into the 2D key field will be read by a computer line by line horizontally from top to bottom, hashed, and processed as usual alike the single-line key field. The hashing process is one round if key strengthening is not used. If key strengthening is used, the hashing iteration is set according to the computer response time per access ranging from 0.05 to 1 second.

The advantages of 2D key are good memorizability, high-entropy key, more references at the user interface to facilitate key input, and resistance to dictionary attack. Its disadvantages are more time for key input and possible shoulder-surfing attack. Fig. 1 displays the pseudocode of 2D key input method.

```
1.0 User selects row size.
2.0 User selects column size.
3.0 User enters ASCII characters or Unicode symbols one by one.
4.0 User ends the key input by pressing the “Enter” key.
5.0 Computer hashes the input key.
6.0 Computer compares the hashed key with the stored hash.
6.1 If the hashes match, authentication is verified.
6.2 If the hashes mismatch, authentication is rejected.
```

Fig. 1 Pseudocode of 2D key input method and system

IV. Styles of 2D Key

A. Multiline Passphrase

For single-line key field, it is hard to input a high-entropy single-line passphrase due to the problem of user interface. A user may lose the reference of starting character of a word in a passphrase. Using 2D key, we can input multiline passphrase, where each line consists of one word of a passphrase. Each word is padded to the longest word in the passphrase. The padding character can be any ASCII character and acts as a text-based semantic noise.

```
HAPPY
*HOME
PINK*
*JOBS
```

Fig. 2 Styles of 2D key: Multiline passphrase

Fig. 2 shows a 2D key example using multiline passphrase. Its dimensions are 4 x 5, and uses character ‘*’ as the padding character. This 2D key has an entropy of 131 bits.

B. Crossword

The second style of 2D key is crossword. Instead of horizontal and vertical multiline passphrase, a user can enter a mixture of horizontal, vertical, and slanted passphrases. Fig. 3 shows a 2D key example using crossword. Its dimensions are 5 x 6, and uses character ‘*’ as the background character. This 2D key has an entropy of 197 bits.

```
SPIRIT
O*N**R
R*NOPU
R*E**T
YARRUH
```

Fig. 3 Styles of 2D key: Crossword.

C. ASCII Art / Unicode Art

The third style of 2D key is ASCII art or Unicode art. ASCII art is a graphical presentation of computer using the 95 printable ASCII characters [19]. Unicode is a variant of ASCII art, where instead of using ASCII characters, Unicode symbols are used to create artistic graphics.

```
116611
161161
116611
111611
166611
116661
166611
```

Fig. 4 Styles of 2D key: ASCII art

Fig. 4 displays a 2D key example using ASCII art. ASCII characters ‘1’ and ‘6’ are used to display a picture of “key” with the opening part pointing downwards. Its dimensions are 7 x 6. This 2D key has an entropy of 275 bits.

Fig. 5 shows a 2D key example using Unicode art. Unicode symbols ‘¥’ and ‘©’ are used to draw a Chinese character (人) meaning “human”. Unicode ‘¥’ is entered using the keyboard by pressing the keys ‘0165’ while holding the key of ‘Alt’. Unicode ‘©’ is entered using the keyboard by pressing the keys ‘0169’ while holding the key of ‘Alt’. Once the ‘Alt’ key is released, the Unicode symbol is entered. Its dimensions are 4 x 5. This 2D key has an entropy of 320 bits.

```
©©¥©©
©©¥©©
©¥©¥©
¥©©©¥
```

Fig. 5 Styles of 2D key: Unicode art

D. Colorful Text

The style of this 2D key needs some additional supports. Color encoding, special graphical user interface, and special computer processing are required. Although these supports make the user interface complicated for the computer, they can be implemented and have better memorizability for the human users. Color is definitely a main element of good memorizability. For example, by having 16 types of colors, every character in the 2D key will have an additional 4 bits. The memorizability of ASCII-based and Unicode-based 2D key will be improved by 60.9% and 25%, respectively. ASCII-based 2D key will become 10.57 bits per character; whereas Unicode-based 2D key is 20 bits per character.

E. Spatial Variation

Another possible style of 2D key is spatial variation. Spatial variation depends on the sequence when a character is entered for a specific element of a 2D matrix. If a 2D key has the dimensions of m x n, the key space is increased by (mn)!. If a 2D key of 4 x 5 as in Fig. 2 is used, the key space is increased by 20! or 61.1 bits to 192 bits, which is close to the example in Fig. 3 for a 2D key of 5 x 6.

This style requires the space encoding for the element location of 2D matrix, table-like graphical user interface of m

x *n* matrix, and human memory for the sequence of characters. In term of memorizability, there is not much improvement. However, the time to enter a 2D key of similar size is greatly reduced.

V. Chinese-Character-Encoded Passphrase

A. *Chinese Language Password*

Chinese character is also called Han character. Zhonghua Zihai in 1994 has 85, 568 Chinese characters [20]. It means a Chinese character has an entropy of 16.38 bits. For key security, this is an advantage over the 6.57-bit ASCII characters, which are used for the Latin languages. For computers with support of Chinese character encoding, Chinese language password is shorter for the similar key size of ASCII-based password. This indicates better memorizability. For computers without support of Chinese character encoding, which are general for majority of the computers, Romanization of Chinese language is needed to create the same advantage in term of memorizability.

Chinese input methods and Chinese character encodings can be used to Romanized Chinese language password. The Romanization of Chinese language is either based on pronunciation, character structure, or a combination of the both. To uniquely represent a Chinese character, [8] is a good reference, where both pronunciation and character structure are used to create a Chinese-character-encoded word with a maximum of six characters.

However, this approach requires modernization. The pronunciation system of Hanyu Pinyin (汉语拼音) [21] and character structure system of Sijiao Haoma or four-corner method (四角号码) [22]-[24] are proposed to create a Chinese language password. In Hanyu Pinyin, there are 415 unique syllables with 22 initials (or onsets) and 39 finals. This pronunciation system is illustrated in Table V.

TABLE V: Phonetic Encoding of Hanyu Pinyin (Mandarin-Based)

Initial (22)	nil	b	p	m	f	d	t	n
	l	g	k	h	j	q	x	z
	c	s	zh	ch	sh	r		
Final (39)	a	o	e	ê	ai	ei	ao	ou
	an	en	ang	eng	ong	i	ia	io
	ie	iao	iu	ian	in	iang	ing	iong
	u	ua	uo	uai	ui	uan	un	uang
	ueng	ü	üe	üan	ün	-i	er	

N.B.: For Romanization, ê and ü can be represented by [oe] and [v], respectively.

In addition to initials and finals, there are 5 tone marks. These tone marks are numbered as 1, 2, 3, 4 and 5 in corresponding with Yinping (阴平), Yangping (阳平), Shangsheng (上声), Qusheng (去声) and Qingsheng (轻声).

横一垂二三点捺
叉四插五方框六
七角八八九是小
点下有横变零头

Fig. 6 Chinese poem for easy memorization of Sijiao Haoma

TABLE VI: Character Structure Encoding of Sijiao Haoma

Stroke name (笔名)	Digit (号码)	Stroke (笔形)
Tou (头)	0	亠
Heng (横)	1	一
Chui (垂)	2	丨 丿 亅
Dian (点)	3	丶
Cha (叉)	4	十 乂
Chuan (串)	5	扌 丰
Fang (方)	6	口 囗
Jiao (角)	7	乛 厂
Ba (八)	8	八 人 入
Xiao (小)	9	小 忄

Then the 4+1-digit Sijiao Haoma is added to describe the character structure of a Chinese character. The upper left number is the first digit. The upper right number is the second digit. The lower left number is the third digit. The lower right number is the fourth digit. The fifth digit is Fuhao or attached number (附号), which represents the middle character structure on the right hand side. Fig. 6 shows a Chinese poem to easily memorize the Sijiao Haoma. Table VI shows the strokes represented by Sijiao Haoma.

TABLE VII: Forms of Romanized Chinese-Character-Encoded Words

Form	[Hanyu Pinyin] (Tone Mark) [Sijiao Haoma] (Fuhao)				
Example	han3714	han43714	han37140	3714han	3H7A1N4

Finally, the Hanyu Pinyin, tone mark, Sijiao Haoma, and Fuhao are joined to form a Romanized Chinese-character-encoded word as in Table VII for the Chinese character of Han (汉). Tone mark and Fuhao are optionally included.

This creates a Chinese-character-encoded word ranging from 5 to 12 ASCII characters. Several Chinese-character-encoded words can be used as a Chinese language password. The capitalization and permutation can slightly increase the entropy. However, it is subject to dictionary attack.

B. *Self-Created Signature-like Han Character*

The combination of 415 Hanyu Pinyin syllables, 5 tone marks, and 10, 000 Sijiao Haoma numbers are more than enough to encode all the Han characters available at present. In order to increase the randomness or entropy of Han character, the creation of new Han character is a must.

This situation happens in real life for the individual name in gaining uniqueness. The created Han character is also signature-like. For Han character creation, it may follow the six methods of Liushu (六书). The Liushu includes Xiangxing (象形) (pictograms), Zhishi (指事) (ideograph), Huiyi (会意) (logical aggregates), Xingsheng (形声) (pictophonetic compounds), Jiajie (假借) (borrowing), and Zhuanzhu (转注) (associate transformation) [20] [25]-[27].

An example of created Han character is shown in Fig. 7. The Han character of (汉) is modified from [han437140] to [han437141] by adding a horizontal stroke between the upper right corner and lower right corner.

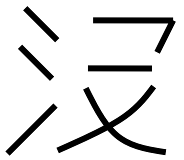

Fig. 7 Example of self-created signature-like Han character by modifying the Han character of (汉) from [Hanyu Pinyin = han4] and [Sijiao Haoma = 37140] to [Hanyu Pinyin = han4] and [Sijiao Haoma = 37141]

Self-created signature-like Han characters enlarge the key space of Chinese language password to 4, 150, 000. When tone mark and Fuhao are included, it becomes 207, 500, 000 or an entropy of 27.63 bits per Han character. The efficiency of Chinese language password is greatly increased.

C. *Self-Created Chinese Language Passphrase*

To further increase the entropy of Chinese language password, we can have self-created Chinese language passphrase. At least one non-alphanumeric character has to be included together with *capitalization*, *permutation*, *character stuffing*, and text-based semantic noise. Character stuffing is like bit stuffing in data communication to enable the syllable length at a fixed value of 6. It is 6 because the maximum syllable length is 6, excluding the tone mark.

Adding fixed syllable length, tone mark, Sijiao Haoma with Fuhao, and one non-alphanumeric character together, a string of 13 ASCII characters is obtained as a basic unit of a self-created Chinese language passphrase. The non-alphanumeric character is used as a separator and text-based semantic noise.

Table VIII shows examples of self-created Chinese language passphrases. This type of Chinese-character-encoded passphrase has an entropy of 85.41 per Han character. It has good memorizability, resistance to dictionary attack, and suitability for general password usages.

TABLE VIII: FORMS OF SELF-CREATED CHINESE LANGUAGE PASSPHRASE FOR (汉)

Form	No character stuffing	With character stuffing & noise	Capitalization & permutation
Example	han4&37140	h@n4***&37140	37140&HaN4***

D. *Cantonese Language Password Using Jyutping*

Han unification of Unicode builds Han characters database for CJK languages (Chinese, Japanese and Korean). The proposed password and passphrase generation method can be applied to any CJK languages using the Han characters by changing the pronunciation Romanization system. The character structure encoding of Sijiao Haoma remains the same for all the Han characters in any CJK languages.

Cantonese language is used by a global population of about 80 millions. Being the official language in Hong Kong SAR (Special Administrative Region) and Macau SAR of PRC (People's Republic of China), the regulation works of Cantonese language are done here. It shares majority of the Han characters with Chinese language in Mandarin except those Han characters in the HKSCS (Hong Kong Supplementary Character Set). For HKSCS-2004, it has 4941 Han characters as in year 2004 under ISO 10646 standard. Hence, it is compatible with Unicode, which implements the ISO 10646 standard.

There are many Cantonese pronunciation systems. Among them, two systems are Romanized and computer friendly. One of them is standard Cantonese pinyin or HKED （《常用字廣州話讀音表》拼音方案）（「教院式」拼音方案）. This is the only pronunciation Romanization system accepted by Education and Manpower Bureau of Hong Kong and Hong Kong Examinations and Assessment Authority. Another is jyutping proposed by LSHK (The Linguistic Society of Hong Kong) in year 1993.

Nowadays, regulation works of Cantonese pronunciation for Unicode adopt jyutping system. Han characters in Unicode are matched with jyutping, where the lists are downloadable from the URLs of [http://www.iso10646hk.net/jp/index.jsp] and [http://www.info.gov.hk/digital21/eng/structure/jyutping.html].

In jyutping system, there are 20 initials and 59 finals as in Table IX. These initials and finals construct about 629 syllables for Cantonese language as compared to 415 syllables for Chinese language in Mandarin. For tone mark, 6 distinct tone contours are used for 9 tones. For completeness, the jyutping has syllables that have no matching Han character. Nevertheless, in the application for Cantonese language password, all jyutping syllables are useful for self-created key.

Table X shows the examples of Cantonese language password. It is similar to Chinese language password in Mandarin. Sijiao Haoma is exactly encoded. For jyutping, the maximum syllable length is 6. Capitalization, permutation, and character stuffing can be used to generate self-created signature-like Cantonese language password and passphrase. The key space of self-created Han characters in Cantonese can reach 377, 400, 000 keys or 28.49 bits per Han characters.

TABLE IX: PHONETIC ENCODING OF JYUTPING IN CANTONESE LANGUAGE

Initial (20)	nil	b	p	m	f	d	t	n
	g	k	ng	h	gw	kw	w	z
	s	j						
Final (59)	i	ip	it	ik	im	in	ing	
	yu		yut			yun		
	u	up	ut	uk	um	un	ung	ui
	e	ep	et	ek	em	en	eng	ei
			eot			eon		eoi
	oe		oet	oek			oeng	
	o		ot	ok		on	ong	oi
		ap	at	ak	am	an	ang	ai
	aa	aap	aat	aak	aam	aan	aang	aai

TABLE X: FORMS OF SELF-CREATED CANTONESE LANGUAGE PASSPHRASE FOR (汉)

Form	Traditional Chinese (漢)	Simplified Chinese (汉)	With character stuffing
Example	hon3&34185	hon3&37140	hon3***&34185

E. *Japanese Language Password Using Rōmaji*

In Japanese language, there are four writing systems: Two syllabaries of Hiragana (平仮名) and katakana (片仮名), one logogram of kanji (漢字), and one Romanization of rōmaji (ローマ字). The most widely used Hepburn Romanization is adopted for rōmaji. The password generation method for Chinese language password can be used for Japanese kanji via the combination of rōmaji and Sijiao Haoma.

TABLE XI: FORMS OF JAPANESE LANGUAGE PASSWORD FOR (大), (漢) AND (山)

Form	dai (大) (だい)	kan (漢) (かん)	yama (山) (やま)
Example	dai&40800	kan&34185	yama&22770

Firstly, obtain the Sijiao Haoma with Fuhao for the Japanese word in kanji. Then, the kanji is converted to rōmaji for pronunciation Romanization. Character stuffing is longer for Japanese language password as the kanji is having variable number of syllables from a minimum of one syllable. For Hepburn Romanization, there are about 132 syllables.

Table XI shows examples of kanji passwords. To avoid dictionary attack, self-created kanji with character stuffing, capitalization, and permutation, as in Section V.B. can be used. Coinware allows random selection of Han characters [4].

VI. APPLICATIONS FOR SYMMETRIC AND ASYMMETRIC KEY CRYPTOSYSTEMS

With the emergence of 2D key with the styles of mutliline passphrase, crossword, ASCII art / Unicode art, colorful text, and spatial variation, high-entropy key as high as 256 bits is possible. Chinese-character-encoded passphrase can be efficiently used for the 2D key style of multiline passphrase. We can now overcome the human factor of memorizability and user interface problem of single-line key field, which limit the key size to 96 bits.

TABLE XII: DIMENSIONS OF 2D KEY FOR VARIOUS SYMMETRIC KEY SIZES

Symmetric key size (bits)	80	96	112	128	192	256
Number of ASCII characters	13	15	18	20	30	39
Dimensions of 2D key	3x5	3x5	3x6	4x5	5x6	5x8

Table XII shows the dimensions of ASCII-based 2D key for various key sizes of symmetric key cryptosystem. Key strengthening can boost up another 19.5 bits. If Unicode-based 2D key is used, the dimensions of 2D key can be greatly reduced.

For asymmetric key cryptosystem, memorizable public-key cryptosystem (MePKC) can be created. This is possible by using the FFC and ECC with minimum size of private key at 160 bits. The private key of MePKC is stored in the human brain, and not stored as encrypted, split, and roaming private keys as in the prior arts. This provides mobility, lower cost, higher efficiency, and resistance to dictionary and precomputation attacks.

Assuming that the maximum memorizable key size is 256 bits, 256-bit MePKC using FFC and ECC with 128-bit security strength can be realized. It has a protection period of 30 years. If key strengthening is used, 19.5 bits is added, or an increase of 10-bit security, which extends the protection to 50 years. This is very much enough for practical applications.

VII. CONCLUSION

High-entropy 2D key input method is proposed. It solves the human factor of memorizability and user interface problem of single-line key field. Chinese-character-encoded passphrase is efficient for the 2D key style of multiline passphrase. Besides, 2D key has the styles of crossword, ASCII art, etc. The memorizable limit of 96-bit key is increased to 256-bit key, where even the private key is memorizable. This creates 160-bit to 256-bit MePKC with protection period up to 50 years.

ACKNOWLEDGMENT

It is to note here that this topic-specific article for the easy reference on *2D key input method*, optionally being used together with the *Chinese-character-encoded passphrase*, is a remix and derivative of an electronically published book [28], by the same author Kok-Wah Lee at the Internet Archive server based in the USA. The copyright coverage of this e-book [28] includes the sharing, remix, derivative, commercial sale, etc. Even though sometimes a copyrighted work is free of charge for some scopes like remix, the moral or ethical right is still existent as can be recognized by the attribution to the author(s) for the collective and derivative works created by the other authors. For instance, people remember, recognize, and honour the author of Pythagorean Theorem for right triangle as Pythagoras of Samos, and not any intermediate persons who have remixed and derived anything from the Pythagorean Theorem. The e-book [28] carries the copyright scopes of free sharing and any remix for any derivatives. Attribution is required for the readers to know the originating source for author contact, reference, comment, error report, correction, improvement, refinement, and enhancement. Imperatively for the cases of radio broadcasting, television program, and press report, which involve the legal responsibilities to the public, the attribution is very crucial for any possible legal proceedings. Meanwhile for honour, all the awards are normally granted to the attributed author(s) for one's contribution to the civilization betterment of society.

REFERENCES

[1] B. Schneier, *Applied Cryptography (2nd ed.)*. New York City, NY, USA: John Wiley & Sons, 1996.

[2] E. Barker, W. Barker, W. Burr, W. Polk, and M. Smid. (2007, March 9). NIST Special Publication 800-57 Recommendation for Key Management - Part 1: General (Revised). [Online]. Available: http://csrc.nist.gov/publications/nistpubs/800-57/sp800-57-Part1-revised2_Mar08-2007.pdf

[3] C. Gehrmann and M. Näslund (Ed.). (2007, January 29). ECRYPT Yearly Report on Algorithms and Key Lengths (2006). [Online]. Available: http://www.ecrypt.eu.org/documents/D.SPA.21-1.1.pdf

[4] K. W. Lee and H. T. Ewe, "Coinware for Multilingual Passphrase Generation and Its Application for Chinese Language Password, " in *Proc. 2006 Int. Conf. Computational Intelligence and Security (CIS 2006)*, Guangzhou, Guangdong, China, 2006, pp. 1511–1514.

[5] Wikipedia Contributors (Eds.). (2007, February 10). Password policy. *Wikipedia, The Free Encyclopedia*. [Online]. Page Version ID: 106996972. Available: http://en.wikipedia.org/w/index.php?title=Password_policy&oldid=106996972

[6] Wikipedia Contributors (Eds.). (2007, February 26). Chinese input methods for computers. *Wikipedia, The Free Encyclopedia*. [Online]. Page Version ID: 111110951. Available: http://en.wikipedia.org/w/index.php?title=Chinese_input_methods_for_computers&oldid=111110951

[7] Wikipedia Contributors (Eds.). (2007, March 17). Chinese character encoding. *Wikipedia, The Free Encyclopedia*. [Online]. Page Version ID: 115862258. Available:

http://en.wikipedia.org/w/index.php?title=Chinese_character_encoding&oldid=115862258

[8] T. D. Huang, “Method for encoding Chinese characters, ” U.S. Patent 4, 500, 872, February 19, 1985.

[9] R. Sandhu, C. deSa, and K. Ganesan, “System and method for crypto-key generation and use in cryptosystem, ” U.S. Patent 6, 970, 562, November 29, 2005.

[10] R. Sandhu, C. deSa, and K. Ganesan, “System and method for generation and use of asymmetric crypto-keys each having a public portion and multiple private portions, ” U.S. Patent 7, 065, 642, June 20, 2006.

[11] R. Sandhu, C. deSa, and K. Ganesan, “Method and system for authorizing generation of asymmetric crypto-keys, ” U.S. Patent 7, 149, 310, December 12, 2006.

[12] C. A. Baltzley, “Public key cryptosystem with roaming user capability, ” U.S. Patent 6, 154, 543, November 28, 2000.

[13] C. A. Baltzley, “Public key cryptosystem with roaming user capability, ” U.S. Patent 6, 292, 895, September 18, 2001.

[14] U. Manber, “A simple scheme to make passwords based on one-way functions much harder to crack, ” *Computers and Security*, vol. 15, no. 2, pp. 171-176, 1996.

[15] M. Abadi, T. M. A. Lomas, and R. Needham, “Strengthening passwords, ” Systems Research Center (SRC), Palo Alto, CA, USA, Tech. Rep. SRC-1997-033, September 1997.

[16] J. Kelsey, B. Schneier, C. Hall, and D. Wagner, “Secure applications of low-entropy keys, “ in *Proc. Int. Workshop Information Security (ISW 1997), LNCS 1396*, Tatsunokuchi, Ishikawa, Japan, 1997, pp. 121-134.

[17] Wikipedia Contributors (Eds.). (2007, April 2). Moore's Law. *Wikipedia, The Free Encyclopedia*. [Online]. Page Version ID: 119636138. Available: http://en.wikipedia.org/w/index.php?title=Moore%27s_Law&oldid=119636138

[18] Wikipedia Contributors (Eds.). (2007, March 31). Typeface. *Wikipedia, The Free Encyclopedia*. [Online]. Page Version ID: 119171254. Available: http://en.wikipedia.org/w/index.php?title=Typeface&oldid=119171254

[19] Wikipedia Contributors (Eds.). (2007, April 5). ASCII art. *Wikipedia, The Free Encyclopedia*. [Online]. Page Version ID: 120415207. Available: http://en.wikipedia.org/w/index.php?title=ASCII_art&oldid=120415207

[20] Wikipedia Contributors (Eds.). (2007, April 5). Chinese character. *Wikipedia, The Free Encyclopedia*. [Online]. Page Version ID: 120519979. Available: http://en.wikipedia.org/w/index.php?title=Chinese_character&oldid=120519979

[21] Popular Book (大众书局), *Han Yu Pin Yin Xue Xi Ka (汉语拼音学习卡)*. Singapore: Popular Book, 2003. (in Chinese language)

[22] Wikipedia Contributors (Eds.). (2007, January 13). Four corner method. *Wikipedia, The Free Encyclopedia*. [Online]. Page Version ID: 100425004. Available: http://en.wikipedia.org/w/index.php?title=Four_corner_method&oldid=100425004

[23] United Publishing House (联营出版有限公司), *Zui Xin Han Yu Da Ci Dian (最新汉语大词典 [修订版])*.Seri Kembangan, Selangor, Malaysia: United Publishing House, 2001. (in Chinese language)

[24] United Publishing House (联营出版有限公司), *Xin Han Yu Zi Dian (新汉语字典)*. Seri Kembangan, Selangor, Malaysia: United Publishing House, 2002. (in Chinese language)

[25] S. Xu (许慎), *Shuo Wen Jie Zi (说文解字)*. Hong Kong SAR, China: Chung Hwa Book (中华书局), 2001. (in Chinese language)

[26] H. Y. Luo (罗华炎), *Jian Ming Han Yu Yu Fa (简明汉语语法)*. Cheras, Kuala Lumpur, Malaysia: Yakin (雅景), 1990. (in Chinese language)

[27] H. Y. Luo (罗华炎), *Xian Dai Han Yu Yu Fa (现代汉语语法)*. Ipoh, Perak, Malaysia: Seni Hijau (艺青), 2003. (in Chinese language)

[28] K. W. Lee, W. D. Chui, and W. J. Chui. (2009, March 14). Memorizable public-key cryptography (MePKC) & its applications. *Internet Archive*. [Online]. Available: http://www.archive.org/details/MemorizablePublic-keyCryptographymepkcItsApplications

Kok-Wah Lee (M’99–S’06–GS’07). Internet keywords: Xpree and/or Xpreeli. This author became a Member (M) of IEEE in 1999, a Student Member (S) in 2006, and a Graduate Student Member (GS) in 2007. Born in Ipoh, Perak, Malaysia on August 21, 1975. Primary, secondary, and pre-U school education at S.R.J.K. (C) Pei Yuan, S.M.J.K. (C) Pei Yuan, and S.M.K. Sri Kampar, Kampar Perak, Malaysia, respectively. First bachelor degree B.Eng.(Hons.) (Mal.) 1999 in electrical engineering from University of Malaya, Kuala Lumpur, Federal Territory, Malaysia. Master degree by research M.Eng.Sc. (MMU) 2003 in electrical engineering from Multimedia University, Cyberjaya, Selangor, Malaysia.

From May to July 1999, he was a Research Assistant at the Faculty of Engineering, Multimedia University, Bukit Beruang, Melaka, Malaysia. From July 1999 to December 2003, he became a Tutor to pursue his master’s degree at the Faculty of Engineering & Technology (FET), Multimedia University, Bukit Beruang, Melaka, Malaysia. From January 2004 till now, he is a lecturer and pursuing his Ph.D. degree at FET. His current research interests are information security, computer communications, Byzantine Generals Problem, artificial neural networks, and electrocardiogram (ECG) signal processing. At the present time, he has his PhD doctoral thesis ready by October 2008 and now pending for the thesis evaluation by the peer experts.

Besides, K.-W. Lee (Mr.) is a registered Graduate Engineer (Electrical Engineering) of the Board of Engineers, Malaysia (BEM), since 2000. He is also a Graduate Member (Electrical Engineering) of the Institution of Engineers, Malaysia (IEM), since 2003. Due to his keen interests in IP (Intellectual Property), he has become an LESM (Licensing Executives Society Malaysia) member since January 2008.