Coinware for Multilingual Passphrase Generation and Its Application for Chinese Language Password

Kok-Wah Lee
Faculty of Engineering & Technology
Multimedia University
Bukit Beruang, 75450 Melaka, Malaysia
kwlee@mmu.edu.my

Hong-Tat Ewe
Faculty of Information Technology
Multimedia University
Cyberjaya, 63100 Selangor, Malaysia
htewe@mmu.edu.my

Abstract

Civilian cryptosystem applies Kerckhoff's law to have security dependency 100% on the password secrecy. This reflects the fact that key length and key space are very important to ensure enough entropy or randomness for securing a cryptosystem. For stronger password, passphrase is suggested. Currently, there are three methods to generate passphrase: Acronym, full sentence and diceware. Here, we propose an alternate method to diceware: Coinware, by using the coin. Coinware uses four coins to generate one hexadecimal digit. The created word lists will be in hexadecimal order and can be applied for multilingual passphrase generation. Its exemplary application for Chinese language password is shown. Readily-made Chinese character word list in the CJK unified ideographs of the Unicode enables fast hexadecimal reading for random passphrase generation. Hanyu Pinyin and Sijiao Haoma are used for Chinese character romanization to uniquely represent each Han character. Jyutping is then used for Cantonese language.

1. Introduction to passphrase generation

Based on Kerckhoff's law [1], a civilian cryptosystem security relies 100% on the secrecy of password or key. This acknowledges the importance of key length and key space. Due to advancement of computing technologies, the secure key sizes have been increasing since 20 years ago. Short password is expected to be replaced by long password or passphrase. Currently, there are three passphrase generation methods: Acronym, full sentence and diceware [1] [2] [3]. Here, we introduce random passphrase generation using coins: Coinware.

1.1. Acronym

For the passphrase created using the acronym [1] [2] [3], a user has to remember one or a few sentences. Then, the first, second or last, etc. characters of each word in the sentence(s) are joined to form an acronym. Both alphanumeric and non-alphanumeric ASCII characters may become the character of the acronym. The techniques of capitalization and permutation may be used to increase the randomness. This acronym will then act as the key. It has the features of high randomness and short key length. The examples of this method are in Table 1.

1.2. Full sentence

The passphrase generation using the acronym is sufficient if the key length requirement is short. When the minimum key size demand is long, normally one full sentence or a few short sentences are entered directly as the key [1] [2] [3]. So far, it is an open problem to type the entire phrase into a computer with the echo turned off [1]. If the masked password is shown during the password entering process, then it can be under shoulder surfing attack.

Besides, since the full-sentence passphrase is having each word to be selected associatively, its randomness is magnitude-wise high but relatively low if ciphertext of password is available. Superuser of any computing system can easily obtain ciphertext of the password. By gaining access to the encrypted password, the threats of ciphertext-only attack and frequency analysis of short cryptogram [4] [5] will be possible. For instance, the unicity distance of English language is about 30 characters. Once the encrypted password is equal to or more than the unicity distance, unique decipherability of the encrypted password will be feasible.

Table 1. Passphrase generation from acronym

Sentence	Passphrase
Passwords shouldn't be impossible to remember and never written down	psbitranwd
Good or bad, you have to do it.	drd,ueoot.
It may be a few sentences: One, two or more.	Imbafs:O,tom.

Table 2. Minimum diceware words (7776 word list) for different security levels

Key Size (bit)	48	64	80	112	128	192	256
Diceware (word)	4	5	7	9	10	15	20

1.3. Diceware

Using full sentence for passphrase generation, the word frequency distribution can be under computational analysis. To get rid of the association of words, diceware [2] introduced by A. G. Reinhold will be an improved passphrase generation method.

There are many software pseudo-random number generators. Unfortunately, they are having lots of pitfalls [6] to ease any possible attack. Hence, some hardware random number generator such as coin and dice are very much better than the software random number generator.

For diceware, it uses dice to select a word from an ordered word list. The word list can be in any language and based on senary or base-6 numeral system. For the most popular diceware, it is an English word list with 7776 (= 6^5) words. Five dice values are needed to locate one word randomly. Every selected word will carry an entropy of 12.92 bits. Table 2 shows the minimum diceware words for different security levels.

2. Coinware

In addition to diceware using dice, coinware using coin is proposed here. Coin tossing is conducted to generate random passphrase. Each face of the coin is labelled as binary bit '0' or '1' respectively. Four coin values are used to derive a hexadecimal digit. Therefore, the word list will be in hexadecimal order.

2.1. Mono-, bi-, multi-lingual word lists

Having word list and random number generator, computational analysis on word frequency distribution is avoided and random passphrase generation is ensured. For the word list, one may use readily-made word list or prepare a new word list.

For readily-made word list, it is normally monolingual unless one combines two or more monolingual word lists with different languages. To prepare a new word list, one may go for monolingual, bilingual or multilingual to suit one's linguistic ability. The purpose to have word list consisting of more than one language is to increase the key space and consequently the key entropy per word.

For the words in the word list, each word has to be unique, short and memorable. Start with the shortest word. Then slowly increase the character length of the word until the key space setting of the word list is met. To be easy, one may set the key space of monolingual word list to 4096 (= 2^{12}) or 8192 (= 2^{13}) words. Two or more monolingual word lists can be joined to form bilingual or multilingual word lists.

2.2. Key length requirements

It is importance to know the key size equivalence for symmetric and asymmetric (RSA, discrete logarithm and elliptic curve) cryptosystems [1] for different security levels. This step enables a user to prepare suitable and sufficiently strong password or passphrase before opening an account and conducting an encryption. Table 3 shows this important information [7].

For the minimum coinware word, it depends on the key space of the word list. The monolingual, bilingual or multilingual word lists of 4096, 8192, 12288, 16384 and 24576 words have key entropies per word of 12.00, 13.00, 13.58, 14.00 and 14.58 bits respectively. Table 4 shows the minimum coinware words for various word list sizes.

The current common demands of security levels are 80- and 128-bit for the symmetric cryptosystem. These security levels ensure protection of 5 and 30 years respectively. From Table 4, word list size of 8192 will be suitable for monolingual and bilingual users. Monolingual users can use a monolingual word list of 8192 words. Meanwhile bilingual users can use two unique monolingual word lists of 4096 words each. For multilingual users, word list size of 24576 is suggested where three unique monolingual word lists of 8192 words each can be used.

Table 3. Key size equivalence for symmetric and asymmetric cryptosystems (bit)

Symmetric Key Size (bit)		48	64	80	112	128	192	256
RSA		480	816	1248	2432	3248	7936	15424
Discrete Logarithm	Field Size	480	816	1248	2432	3248	7936	15424
	Subfield	96	128	160	224	256	384	512
Elliptic Curve		96	128	160	224	256	384	512

Table 4. Minimum coinware words for various word list sizes (WLS)

Symmetric Key Size (bit)		48	64	80	112	128	192	256
Coinware (word)	WLS 4096	4	6	7	10	11	16	22
	WLS 8192	4	5	7	9	10	15	20
	WLS 12288	4	5	6	9	10	15	19
	WLS 16384	4	5	6	8	10	14	19
	WLS 24576	4	5	6	8	9	14	18

Table 5. Minimum coinware words for Han character combined list (70000 words)

Symmetric Key Size (bit)	48	64	80	112	128	192	256
Coinware (Chinese Language) (word)	3	4	5	7	8	12	16

3. Chinese language password generation

For the applications of coinware, we can have readily-made word list in Unicode [8] for various languages. This is because both coinware and Unicode are in the hexadecimal order for their word lists. This is especially true for the CJK languages of Chinese language, Japanese language and Korean language that use the Han characters. Word list is also a character list for CJK languages. Here, we discuss on the Chinese language password generation by using coinware.

3.1. Unicode

Unicode unifies the Han characters of CJK languages into CJK unified ideographs or Unihan under ISO 10646. There are three major blocks of Han characters or Chinese characters in the Unicode character encoding: CJK unified ideographs, CJK unified ideographs extension A and CJK unified ideographs extensions B. For the mean time, Unicode Consortium is preparing the CJK unified ideographs extension C.

For Unicode 4.1, the first block lists the Han characters from [4E00] to [9FBB] in hexadecimal value. The second block lists from [3400] to [4DB5]. The third block lists from [20000] to [2A6D6]. Hence, we have three readily-made word lists or character lists. These word lists have 20924, 6582 and 42711 words or characters respectively. For a combined word list, it is a key space of 70217 characters. After excluding the radicals, it is a net key space of about 70000 characters. This allows a Chinese language word list with high entropy of 16.10 bits per Han character to be formed.

To start coinware, first flip or toss a coin to randomly select a binary bit '0' or '1'. If bit '0', the first and second blocks of CJK unified ideographs and CJK unified ideographs extension A are chosen. If bit '1', the third CJK block of CJK unified ideographs extension B is chosen. Then toss coin again to get 4 coin values representing 4 binary bits. These 4 binary bits are converted into 1 hexadecimal digit. Repeat coin tossing to get 4 coin values for another 3 rounds. Four random hexadecimal digits locate a unique Han characters in the selected CJK block(s). These 3 blocks are available at URL [http://www.unicode.org/charts/]. If the hexadecimal digits do not hit any Han character, get another set of hexadecimal digits. Coming to here, the selected Han character will need Chinese character romanization to enable computer input.

3.2. Chinese character romanization

For Chinese character romanization, the phonetic encoding of Hanyu Pinyin (汉语拼音) and symbolic encoding of Sijiao Haoma (四角号码) are used to uniquely represent each Han character in the Unihan of the Unicode. Sijiao Haoma is used in addition to Hanyu Pinyin as there are only 415 syllables and 4+1 tone marks and hence unable to uniquely encode all Han characters. Also, Sijiao Haoma helps differentiate traditional and simplified Chinese characters.

For Hanyu Pinyin, it is taught in primary and secondary education. For Sijiao Haoma or four corner method, one can learn from a dictionary [9]. In Sijiao Haoma, there are 4 main digits representing the shape of strokes of a Han character at the upper left, upper right, lower left and lower right corners. An attached number or Fuhao (附号) can be entered as the fifth

digit. Fuhao represents the stroke between the upper right and lower right corners. To ease memorization of Sijiao Haoma, a Chinese poem is shown in Figure 1.

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

Figure 1. Chinese poem for easy memorization of Sijiao Haoma

For the computer input of Chinese language password for Han character, the Hanyu Pinyin and Sijiao Haoma can be typed side by side. If the coin tossing gives a 5-digit hexadecimal string of [06C49], then the Han character of (汉) is selected from the Unicode. The possible forms for the Chinese character romanization of (汉) [Hanyu Pinyin = han4] [Sijiao Haoma = 37140] are [han3714], [han437140], [3714HAN], [37140han4], [3H7A1N4], etc.

Referring to the created Han character combined list in Section 4.1, it is 16.10 bits per Han character. Using this key entropy, we can derive minimum coinware words at different security levels for Chinese language password as in Table 5.

Due to high key entropy of Chinese language word list, it is obvious to observe the significant drop of minimum coinware words as from Tables 4 and 5. As Japanese language and Korean language are using the Han characters as well, similar word lists with large key space can be created. For Chinese language family, it is applicable to Mandarin, Wu, Cantonese, Min, Jin, Xiang, Hakka, Gan, Hui and Ping languages/dialects with speaking population of 800, 90, 80, 50, 45, 35, 35, 20, 3 and 0.2 millions respectively.

3.3. Jyutping & Cantonese password creation

For a member of Chinese language family, it has a special position as like Mandarin language. Cantonese language is the official language of Hong Kong SAR and Macau SAR of PRC (People's Republic of China). Hence, Cantonese language is modernized by Hong Kong government to be applicable for electronic communication in the computer age.

Jyutping is proposed by LSHK (The Linguistic Society of Hong Kong) for the pronunciation romanization of Cantonese language. With the pronunciation standard of jyutping and the encoding standard of Unicode, there are readily-made word lists for Cantonese language password generation. These materials can be downloaded from the URLs of [http://www.info.gov.hk/digital21/eng/structure/jyutping.html] and [http://www.iso10646hk.net/jp/index.jsp]. From the coin tossing of coinware, as for the hexadecimal strings of [03400], [04E00] and [0E000], the respective jyutping of these Cantonese characters will be [jau1], [jat1] and [mou5].

4. Conclusions

In summary, a new passphrase generation method is proposed: Coinware. The hardware random number generator of coin is used as like dice in diceware since software random number generator is having lots of weakness. Coinware with word list in hexadecimal order has readily-made word lists. It is especially useful for the passphrase generation of CJK languages. Here, an exemplary coinware application for Chinese language password generation is presented. It can be applied to be a password generation method as in [10].

5. References

[1] Schneier, B., *Applied Cryptography*, 2nd Ed., John Wiley & Sons, New York City, New York, USA, 1996.

[2] PGP Corporation, *PGP Desktop 9.0 for Windows User's Guide*, PGP Corporation, Palo Alto, California, USA, 2006, pp. 229-232.

[3] J. Yan, A. Blackwell, R. Anderson, A. Grant, Password Memorability and Security. *IEEE Security and Privacy Magazine 2(5)*, 2004, pp. 25-31.

[4] G.W. Hart, To Decode Short Cryptograms, *Communications of the ACM 37(9)*, 1994, pp. 102-108.

[5] K.W. Lee, C.E. Teh, Y.L. Tan, Decrypting English Text Using Enhanced Frequency Analysis, *National Seminar on Science, Technology and Social Sciences (STSS 2006)*, Kuantan, Pahang, Malaysia, 2006.

[6] Eastlake 3rd, D., Crocker, S., Schiller, J., *Randomness Recommendations for Security*, Network Working Group, Request for Comments: 1750 (RFC 1750), The Internet Engineering Task Force (IETF), Sterling, VA, USA, 1994.

[7] Gehrmann, C., Näslund, M. (ed.), *ECRYPT Yearly Report on Algorithms and Keysizes*, European Network of Excellence in Cryptology, IST-2002-507932. Katholieke Universiteit Leuven, Leuven-Heverlee, Belgium, 2006.

[8] The Unicode Consortium, *The Unicode Standard 4.0*, Addison-Wesley Professional, Boston, MA, USA, 2003.

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

[10] M.R. McCulligh, Password Generation Method and System, *USPTO Issued Patent*, 6,643,784, 2003, pp. 1-11.