METHOD AND SYSTEM TO CREATE CHINESE LANGUAGE SEEDS FROM MOST FREQUENTLY USED CHARACTERS AND THEIR APPLICATIONS

Author cum Inventor: Kok-Wah Lee

© 04 January 2010. All rights reserved. (URL: www.xpreeli.com)

Field of Invention

As knowledge area for generality, the present invention relates to telecommunications and computer engineering of electrical engineering. As knowledge field for particularity, the present invention relates to cryptography, information security, and Chinese language processing of information engineering. As knowledge focus for specificity, the present invention relates to computer-free cryptography, polyalphabetic cipher or Vigenère cipher, Chinese character encoding, and Chinese input method. As knowledge strength for uniqueness, the present invention relates to method and system to create Chinese language seeds from most frequently used characters, normal distribution, and IQ (Intelligent Quotient), together with their derived applications for Chinese language processing towards cryptographic usages and specific stroke encoding.

Background to the Invention

--- Chinese Language Processing ---

There are two types of Chinese language processing here: Manual and computer modes. To learn Chinese language, which is a logographic language using morpheme as symbol representation, it is well-known that there are too many Chinese characters to be learned. For Unicode, it has amounted to more than 70,000 CJKV characters (or Han characters, Chinese characters). To ease the Chinese language learning, in recent decades, the educationists have proposed to learn firstly the most frequently used characters (Yang & Zhu, 1999). As for primary school students, it amounts to 2,500 Chinese characters; whereas for foreign learners as a second language, it amounts to 2,905 Chinese characters.

However, these unique popular Chinese characters exist individually character by character, or word by word. Yet in recent decade, there exists linguistics research (Xiao, 2004) to group Chinese characters with similar semantics, like synonyms and antonyms together, to form related phrases, sentences, and stories for easy memorization. However, the sentences and stories are not compact enough to have only the most popular Chinese characters occurred and emerged only once.

Hence, there exists a need to improve this efficient educational method to have more compact or most compact sentences formed from most popular Chinese characters. If this need can be realized, then Chinese language will have the most frequently used characters to be arranged in easily memorizable

sequence as like the English alphabets from A, B, C, …, to Z. Following from this, there are lots of useful and practical applications like Chinese input method and polyalphabetic cipher using Chinese language.

Yet there are also linguists showing that yijing (易经) (Book of Changes) has relationship with Chinese characters (Chang, 2000). It has shown the one-to-one relationships of the bagua (八卦) to its individual Chinese characters in ancient Chinese literature, such as qian (乾) (☰) for sky (天) (tian), dui (兑) (☱) for swamp (泽) (ze), li (离) (☲) for fire (火) (huo), zhen (震) (☳) for thunder (雷) (lei), xun (巽) (☴) for wind (风) (feng), kan (坎) (水) (☵) for shui (水) (shui), gen (艮) (☶) for mountain (山) (shan), and kun (坤) (☷) for earth (地) (di). However, Chang (2000) only discussed on some Chinese characters and their related thoughts based on yijing. Hence, there exists another need to link more closely or most closely the relationships between the yijing and Chinese characters.

--- Chinese Input Method ---

For Chinese language processing using computer, Chinese character encoding and Chinese input method are two of the main problems to be optimally solved. For Chinese character encoding, attempted encoding systems have once used and are still using the basic features of Chinese characters, such as stroke or character structure like radical (部首), pronunciation, and semantics. In the currently most widespread Chinese character encoding, i.e. Unicode encoding, the radical has been used to arrange the Chinese characters for further pure digit encoding without referring to any Chinese character feature.

For Chinese input method, there are three types of them till today: Pure digit encoding, pronunciation, and character structure (or stroke). For pure digit encoding like Unicode, one can enter a Chinese character by using character map software and hot key of keyboard, like pressing and holding “Alt” button together with Unicode encoded decimal value followed by a release of the “Alt” button. The advantage of this method is any Chinese character can be entered. However, its disadvantage is reference tables to the Chinese characters are required. Besides, it is very ineffective in term of input time.

For Chinese input method using pronunciation like Mandarin and Cantonese, pinyin (拼音) input method is its most famous name. For those people literate in Chinese pronunciation system, this method is very easy to be learned. However, for non-Chinese speaking community, it is a hard problem like human talking with the bull and cow. Pinyin input method supports both traditional and simplified Chinese characters. Its other weakness is slow entry bottlenecked by a character input buffer holding overlapping Chinese characters within certain conversion length. The Chinese characters sharing the same code are normally arranged in certain order, like popularity, at the input buffer screen. For pinyin method, the overlapping Chinese characters using the same coded value are a lot. Hence, the character input buffer, that holds the overlapping characters and allows only one selection per pinyin code, is really very inefficient when medium and large computing tasks based on Chinese language have to be carried out.

In year 1976, Bong-Foo Chu (朱邦复) from Taiwan (China ROC) invented the Cangjie (仓颉) input method, which is a Chinese input method using character structure embedded with semantics. This method initially works on traditional Chinese characters (正体字，又曰繁体字) only, but now it supports also simplified Chinese characters (简体字). In term of advantage, it only has about 140 decomposed components (字根) and auxiliary shapes (辅助形) to be remembered by a user. These 140 components may be and may not be alike a Chinese radical. For 300 most frequently used Chinese characters, only two keyboard alphabet buttons are required to enter a popular Chinese character. There are now five generations of Cangjie input method. For the third generation, its overlapping code rate (重码率) for the most popular 5,401 Chinese characters in Big5 encoding is about 8%, which is one among the lowest rates for Chinese input methods using character structure. The maximum number of Cangjie code for a Chinese character is five. However, users may have difficulties in applying the character decomposition rule into standard predefined shapes to get a Cangjie code. In short, the main weakest of this method is the memorization difficulty to remember all the components, auxiliary shapes, and decomposition rules.

For pinyin input method, it is popular in overseas China. For Cangjie input method, it is widely used in Taiwan (ROC) and Hong Kong (China SAR). For mainland China PRC, the most popular Chinese input method is Wubi (五笔) input method invented by Yong-Min Wang (王永民) from China PRC in year 1983. This method uses five types of stroke in Chinese characters: Horizontal (横) (一), vertical (竖) (丨), falling left (撇) (丿), falling right (捺，同点) (丶), and turn (折，同提) (乙). These five stroke groups are mixed to form another 25 higher groups by using the sequentially ordered first and second strokes of the decomposed components from a Chinese character. The decomposed component may be and may not be the same as a Chinese radical. All the 25 higher stroke groups are represented by ASCII alphabetic buttons A, B, C, …, to Y. For alphabetic button Z, it may optionally be set as a magic key or wild card key to represent an unknown stroke group.

There are two popular version of Wubi input method: Version 1986 and version 1998. Here, it has 130 and 245 components for versions 1986 and 1998, respectively. Wubi input method (version 1986) initially supports simplified Chinese characters only, but now its version 1998 has included also traditional Chinese characters. For the amount of supported Chinese characters, versions 1986 and 1998 have altogether 6,763 and 21,003 symbols, respectively. Its overlapping code rate is another lowest among the same type of Chinese input method. The maximum number of Wubi code for a Chinese character is four. However, Wubi input method has similar problems like Cangjie input method, i.e. its main weakest is the memorization difficulty to remember all the components, auxiliary shapes, and

decomposition rules. Nevertheless, Wubi input method has weakly correlated mnemonic poems to help new learners with poor memory.

To use a Chinese input method that friendly supports both the traditional and simplified Chinese characters using the Chinese character structure or stroke, ZhengMa (郑码) input method invented by Yi-Li Zheng (郑易里) and Long Zheng (郑珑) is perhaps a good option. It has scalability to include large set of Chinese characters. Moreover, this method functions alike the Wubi input method and hence it is also easy to be learnt. However, it is still new and popular only among the scholars, but not the public yet.

Coming to here, one has to know that there are too many types and variants of Chinese input methods. If Chinese language is needed to be alike English language to act as a language friendly for cryptography and information security, then there exists needs to do better and more organized Chinese language processing for Chinese character encoding and Chinese input method. Cryptography normally requires one-to-one mapping between the plaintext and ciphertext symbols, or rarely many-to-one mapping from plaintext to ciphertext. Without a common standard encoding between the message sender and receiver, then secure communications cannot be successfully conveyed. Here, the need for a common standard of Chinese character encoding can be based on Unicode encoding.

Yet, the need for a common standard of Chinese input method is still a problem, unless all the Chinese input methods share the same Chinese character encoding. Hence, for computing purposes, there exists a specific need to have a special Chinese input method locally standard to the books and application software using the Chinese language as another language tool of cryptography and information security. In addition, there exists a general need to further simplify and improve optimally the encoding and input method of Chinese language processing. The rules of thumb are simple to learn, easy to memorize, less overlapping code, and fast enough speed.

Moreover, there exist concerns for proprietary application software of Chinese input method and user's independence to use the proposed applications in this article. Hence, in view of software independence of Chinese input method for imperative cases, a further need exists for the author cum inventor Kok-Wah Lee to develop a specific Chinese input method for its specific applications suggested here.

--- Computer-Free Cryptography ---

Before discussing on using the Chinese language as a cryptographic tool like English language, let us survey the possibility of computer-free cryptography. In the ancient time, information security began with body language, sign language, and spoken language. Then, when more messages had to be conveyed, the written language on written media, especially paper, was used to form ciphertext.

Without any mechanical machine, although the encryption algorithm was simple, paper cryptography, like substitution and transposition ciphers, had worked for a few centuries. Coming to the end of nineteenth century, mechanical machines, like rotor machines, running at fast speed had been able to crack the paper cryptography, whenever the secret entropy of the key and encryption algorithm was low enough. Furthermore, mechanical machine did multi-stage encryption and decryption functions at a very promising speed for quite large amount of messages.

Nevertheless, the prosperity of mechanical machine cryptography had not lasted long. It ended in the years of 1940's when the electronic computers were invented to fight against and crack them. So, computer cryptography has begun since years 1940's during the WW2 (World War II) and lasts till today.

However, one would not have known when the electricity power failure will occur. Also, there may be time one has computer absence from convenient access. More seriously, will the computer technology disappear when certain human generations cannot cope with the various engineering knowledge complexities? Hence, there exists a need to let us prepare for computer-free cryptography age. During the prevailing period of computer technology, this computer-free cryptography shall be able to resist the brute-force attacks from computers. Meanwhile, in the era of total computer absence, this computer-free cryptography shall still be able to operate without direct and indirect computer assistance.

In the modern day, computers have helped to keep various multimedia formats as electronic records and evidence. Be it bitstream, text, image, audio, and video for the multimedia. From criminology, it is well-known that bad people doing crime tend not to leave anything behind as record. Any possible left record may be a proof to bring them to jail as punishment, or a threatening tool to force them to become dolls or slaves of the evidence holder. In short, bad people are also afraid of losing their freedom. Please keep in mind that bad people like to extort among themselves to get dolls and slaves

In view of this possibility, computers are great anti-crime warriors. However, there may be one day, in case bad people have turned worse to become evil or even devil, then some or all the computer technologies may not function properly or totally lose in actions. Why? This is because bad people tend to convey messages vocally without a record as archival, or with malicious record to confuse investigators. On the other hand, for good people, they like to leave record as evidence to show truth.

Subsequently, bad people tend to keep data related to bad deeds, like lies and matched truth, in the brain memory. To bad people, other humans, papers, computers, and so on can only keep partial truth that will not affect their future actions. To good people, other humans, papers, computers, and so on can help them to keep most of the truth, as long as feasible in time and space. So, there are more memory spaces of own brain and other secondary media for good people to keep data, as compared with the bad people. Coming to here, recordable cryptography like computer cryptography tends to be used by good people in high

frequency and huge amount of messages. To bad people, the more they have used the recordable cryptography, the more proofs are left behind, and the more dangerous they are in future. Bad people like unrecorded and transient cryptography to get rid of “faulty leg” (痛脚) from being caught, like body language, sign language, and spoken language, which cannot carry medium and huge amount of messages. In fact, the correct measure is to settle the “faulty leg” by public account clearing (找数).

Moreover, recordable cryptography has higher accuracy rate and wider broadcasting coverage due to its replaying ability. Assisted by digital communications and cryptography, added features like signal transmission, error detection and correction, access control, authentication, data confidentiality, data integrity, and non-repudiation can boost up the good people’s capabilities from using recordable media.

Consequently, in case destructive weapons are in the hands of evils and devils planning to destroy any evidence and actions against them, computers may then be totally destroyed one day. Hence, there exist critical and ultimate needs to prepare ourselves for computer-free cryptography. This step can be a tool to frighten the organized crime syndicates to retreat, because a conspired plot to destroy all the computers would have not been a total success to break down the communications networks of good people.

--- Polyalphabetic Cipher or Vigenère Cipher ---

After expressing the needs for computer-free cryptography, let us rewind to the age just before the computer cryptography. So, does mechanical machine cryptography using a mixture of multi-stage substitution and transposition ciphers is feasible to be a good option? The answer is not a definitely YES. The more components of a mechanical machine, the more people involved in manufacturing it, the higher the chances to get the manufacturing secrets disclosed to public. Furthermore, complicated mechanical machines are normally having medium size and heavy. There may also be expensive cost and high incompatibility problems. So, it may be not affordable, as well as cannot be easily portable and transported everywhere. In short, it is inconvenient.

Subsequently, let us resort to polyalphabetic cipher or Vigenère Cipher. In term of paper cryptography, polyalphabetic cipher is perhaps the peak performer. When a conventional polyalphabetic cipher operates over the 26 alphabets in English language, and in case the cipher table is a secret algorithm, then its secret entropy is 4.70 (= $\log_2 26$) bits added with the secret key size.

For human memory, a secret key size is normally less than 100 bits. Therefore, polyalphabetic cipher using English language at about 100 bits the secret entropy is not strong enough to resist the brute-force attack of computers. Of course, this is the main reason why polyalphabetic cipher has been claimed to be obsolete. Modern cryptography requires secret entropy at 128 bits and above. Hence, there exists a need to research and develop further, in case if polyalphabetic cipher can still be possibly improved to resist brute-force attacks from modern computers and potential mechanical machines.

--- Cryptography and Information Security in Chinese Language ---

To make the polyalphabetic cipher practically secure to resist the brute-force attacks from computers and mechanical machines, its secret entropy has to be lengthened. Subsequently, it is either or both the secret key size and unique symbol set of cipher table to be enlarged.

Firstly from Kok-Wah Lee's novel copyrighted article (2008, 2009) filed for patent application, the human memorizable key size can be normally as high as 256 bits using method like 2D key (two-dimensional key) together with ASCII encoding. It can also be even higher at 512 bits using method like 2D key together with Unicode encoding. Nevertheless, one may think that there is proprietary concerns to explain and use human memorizable secret key (Lee, 2008, 2009). Here, it is to note that the copyright license is free to lots of human community group type. After that, for huge amount of one-lump-sum copyright licensing annually, a copyright license may be as low as one cent per novel edge dependent idea and ten cents per novel core independent idea in the local currency unit per head or computer count. In short, this copyrighted big memorizable key size shall have an entry cost affordable to all. Further added value providers, like application software developers, may charge further costs for their products and services. However, the rules of thumb are to keep it free for certain human groups and affordable to all people.

Secondly to get larger unique symbol set size of cipher table, Chinese language and Japanese language using the CJKV characters (or Han characters, Chinese characters) (CJKV: Chinese, Japanese, Korean, Vietnamese) are two potential candidates. Due to the reasons that China has an expected peak population at 1.4 billion soon and Chinese language has larger CJKV character set than the Japanese language, then Chinese language is a more suitable language candidate for polyalphabetic cipher working on papers.

As info, there are about 3,000 Chinese characters to be learnt by primary school students at advanced level and learners taking Chinese language as a second language. For a current small Chinese language dictionary suitable for general usages, it has about 10,000 Chinese characters. However, for a Japanese language dictionary, the most frequently used kanji character list as announced in year 1981 (昭和 56 年) has only about 1,945 kanji characters to be learnt for normal usages. For CJKV characters (or Chinese characters, Han characters), it is called hanzi in Chinese language, kanji in Japanese language, hanja in Korean language, and Hán Tự in Vietnamese language. In Unicode, it has over 70,000 Han characters.

However, Chinese characters have problem to be sequentially arranged in an easily memorizable order. This problem can be observed from the index tables of Chinese language dictionary to search for a Chinese character. So far, radical (or character structure) and pinyin pronunciation are used to locate a Chinese character. This issue has indirectly reflected how cryptography using Chinese language is retarded and may work.

For polyalphabetic cipher using English language, there are two basic forms of encryption: Substitution and transposition. These encryption forms only perform functions on 26 alphabets and optionally an added space character. Cipher operates on alphabetic character, and code operates on word or character stream. The English language alphabets basically carry only the information of character structure, phonetic sound, sequential order, and very limited semantic meaning.

Meanwhile for Chinese language characters, they only carry the information of character structure, phonetic sound, and broad semantic meaning. However, for the present time, it is literally without clear and easily memorizable sequential order. The absence of clear sequential order in Chinese language has retarded its research and development in the field of cryptography and information security.

So far, to apply cryptography into spoken Chinese language, one way is to use the method so called as "snake dialect" (蛇话) (Hong, 2009). For example, for the plaintext of "ni hao ma?" (你好吗？), the ciphertext may be pronounced as "ni si hao sao ma sa?", by adding one or more specific immediate phonetic units for every plaintext character. Of course, the speaker and listener have to firstly achieve consensus on the phonetic encryption algorithm and secret key.

For cryptography using written Chinese language, the features of character structure, phonetics, and semantics can be used. Let us take some possible Chinese language ciphertext examples from "Shan Hai Jing" (山海经) （Book of Mountains and Oceans） (Liu & Liu, 2002). For example using character structure, a ciphertext "shi jiu" (始鸠) may have a plaintext "nv tai jiu niao" (女台九鸟), having decomposed components to mean a table hosted by a female has another nine birds as participants. If all of these 10 members around the table are of individually unique feature each for the yinyang five elements totaling to 10, then it is of fair game for one to win alike around a gambling table. For example using phonetics, a ciphertext "yin yin hu" (因因乎) may have a plaintext "yin yin fu" (印阴符), having adjacent pronunciation to mean printing yin charm. For example using semantics, a ciphertext "han yan" (韩雁) may have a plaintext "han xin" (韩信), having similar semantic Chinese characters to mean a general name during the late Qin Dynasty (秦朝) and early Han Dynasty (汉朝) in China.

In a fast scan, it is clear that these cryptographic techniques using the Chinese language involve only simple substitution and transposition ciphers. So in order to convey more complicated hidden meaning in Chinese language, the Chinese poets have done their jobs very well to create lots of fantastic poems (诗词). However, the problem is normally another human receiver requires similar Chinese literature strength or higher, or hints and explanation from the poet composing the poems. From here, another subsequent problem has occurred. Normally if a poem has offended a person in power, then there will be no clear hint and explanation, or even misleading messages to get rid of big troubles.

Consequently, the question is “Could polyalphabetic cipher be used in Chinese language?” A reduced question form is “Could Chinese language characters have clear and easily memorizable sequential order?” These questions can be answered by using the hint that Chinese characters have sequential ordered list of most frequently used Chinese characters. Hence, there exists a need to process the list of most frequently used Chinese characters into clear and easily memorizable sequential order.

--- Big Memorizable Secret Key: 2D Key Using Sensitive Input Sequence ---

For the current security threshold of symmetric secret key, it has to be at least 128 bits now. When the symmetric secret key is more than 256 bits, it is considered as to be able to resist even the future potential quantum computer attacks. Here, in case the Chinese polyalphabetic cipher can be and has been realized, then the minimum secret key size shall be 128 bits. For maximum security, then it has to be at least 256 bits.

From Kok-Wah Lee (2008, 2009), out of the various generation methods of big memorizable secret key, 2D key is the most suitable one for polyalphabetic cipher. When 2D key is based on ASCII and Unicode encodings in the computer systems, its respective key sizes are 6.57 (= $\log_2 95$ for all keyboard printable characters) and 16.59 (= $\log_2 98{,}884$ for Unicode 5.0) bits.

So for paper cryptography using the polyalphabetic cipher in Chinese language, its 2D key size may range up to 16.10 (= $\log_2 70{,}000$) bits depending on the size of unique symbol set. Nevertheless, human has memory limitation to keep long sequence of sorted characters. Furthermore, books made up of papers have weight and size limitations to keep long list of cipher table. Till here, there exists a need to decide an optimum size of unique Chinese character set for the applications of Chinese polyalphabeitc cipher.

To help ease the large symbol set requirement to fulfill the minimum security threshold of secret key size, a key style called sensitive input sequence of 2D key can be considered. One may use general spatial input sequence. However, there will be a second secret burden of hidden character reading sequence of the 2D matrix of 2D key. Here, there exists a need to reduce this second secret burden by whatever possible means.

--- Conclusions ---

In a nutshell, due to computer tendency to keep multimedia data as records with replaying capability, the current computer age may have problematic time periods that there are electricity power failure and computer absence. This is possible when organized crime syndicates launch their conspired plot attacks to destroy the computers and their networks. Hence, there exist critical, ultimate, and imperative needs to prepare for computer-free cryptography.

Till here, paper cryptography using polyalphabetic cipher in Chinese language is a good option. Therefore, the sequential ordering problem of Chinese characters has to be solved. The hint is to use most frequently used Chinese character list. Consequently, Chinese language processing in the manual and computer modes has to be improved. For computing purposes, the Chinese character encoding and Chinese input method have to be further studied, researched, and developed to enable polyalphabetic cipher using Chinese language on the platforms of papers and computers. To reduce the symbol set size requirement of unique Chinese characters for polyalphabetic cipher applications as computer-free cryptography, key style of sensitive input sequence of 2D key has to be improved before it can be friendly applied.

Summary of the Invention

The present invention provides novel method and system to create Chinese language seeds as quantized sentences by IQ levels, from most frequently used Chinese characters, to get sequential order list of about one thousand Chinese characters, as well as their applications for cryptographic usages like computer-free cryptography and Chinese language processing like Chinese input method.

The first (main) independent embodiment of the present invention is the method and system to create Chinese language seeds as mnemonic sequential order of about 1000 most popular Chinese characters using most frequently used Chinese characters, normal distribution, and IQ (Intelligent Quotient). These Chinese language seeds are in the form of mnemonic sentences quantized by IQ levels and strongly correlated into a few storylines.

The second (main) independent invention embodiment is 2D key (two-dimensional key) the big memorizable secret using general spatial input sequence and specific sūdoku (数独) input sequence to further increase the secret entropy.

The third (side) independent invention embodiment is, in particular, an independent and a complementary Chinese input method to smoothen the applications of created Chinese language seeds for Chinese language processing under the computer mode. This side independent invention embodiment uses special stroke encoding of Chinese character structure based on bagua (八卦), yijing (易经), and Chinese character yong (永) (meaning forever). Its given name is “Yongzi Guaxiang” (永字卦象) input method.

Other than these, there is a first group consisting of two dependent invention embodiments, from the extended applications of the first (main) independent invention embodiment, used for Chinese language processing. Then, there is a second group consisting of two dependent invention embodiments, from the extended applications of the first (main) and second (main) independent invention embodiments, used for cryptographic usages.

Yet, there is a third group consisting of another three dependent invention embodiments, from the extended applications of the first (main), second (main), and third (side) independent invention embodiments, used for computer cryptography using polyalphabetic cipher in Chinese language. Lastly, there is a fourth group consisting of only one dependent invention embodiment, from the extended applications of the seventh application using key strengthening and multihash key, to create pseudo-longer semantic unicity distance and stronger encryption algorithm.

--- Method and System to Create Chinese Language Seeds ---

In a first preferred embodiment of the present independent invention to create Chinese language seeds, one has to know that there are many lists of most frequently used Chinese characters prepared by various people and legal entities. Basically, there are three types of list: Pure traditional Chinese characters, pure simplified Chinese characters, as well as mixed traditional and simplified Chinese characters.

Nevertheless, since the relationship between the traditional and simplified Chinese characters is a mapping, either one to one or many to one, so there is normally no problem to convert from traditional Chinese characters to simplified Chinese characters. Inversely to convert from simplified Chinese characters to traditional Chinese characters, the contextual semantics of a particular Chinese character has to be considered.

In this article, the adopted list of most frequently used Chinese characters is of the type pure traditional Chinese characters prepared on Internet for archival by Chih-Hao Tsai (蔡志浩) (1996-2006) from Taiwan (China ROC). The corpus is based on Big5 encoding and copied from Usenet newsgroup during years 1993 to 1994. There are 13,060 unique Chinese characters in this corpus forming a total of 171,882,493 Chinese characters. The average frequency per Chinese character is 13,161 times. The frequency statistics was firstly reported by Shih-Kun Huang (黃世昆) from Taiwan (China ROC) twice in two continuous years 1993 and 1994. This corpus prepared and maintained by Shih-Kun Huang is perhaps the largest Chinese corpus on Earth then.

For the first look from the top at these 13,060 Chinese characters sorted according to descending usage frequency as the first tool, they are still too random for easy memorization. So, it is not enough and other tools are needed. Kok-Wah Lee the author cum inventor has applied also IQ levels as the second tool and normal distribution as the third tool. For IQ levels with mean ($\mu = 100$) and standard deviation (S.D.) ($\delta = 15$), the sorted Chinese character list is further quantized according to IQ levels beginning from (IQ = 70) to IQ levels of normal people and excellent geniuses, according to different language education needs.

According to Global Language Monitor (2009), for English language, it has about 2,000 to 3,000 core defining vocabulary words to explain any word entry in English language dictionary. For Cambridge International Dictionary of English, it has a strictly controlled 2,000-word defining vocabulary to ensure

fast accessibility of root word. On 10 June 2009, English vocabulary has reached its 1,000,000-th word, by using the new English word selection criteria of minimum 25,000 citations are needed for the necessary breadth of geographic distribution and depth of citations. For average newspaper and King James Bible, there are about 8,000 different English words. Meanwhile, an average English speaking person can recognize and recall about 50,000 and 14,000 English words, respectively. A linguistically gifted person in English language will only use about 70,000 English words.

Lian-Zhi Zhuang (1997) reported that for the first 120 Chinese characters having the strongest phrase formation power, over 26,000 phrases could be built. For the first 1,000 and 2,500 most popular Chinese characters, 90% and 99% of the common newspaper corpus have been covered, respectively. Nowadays, by knowing about 90% and 99% of most frequently used Chinese characters, a Chinese speaking person will need only a Chinese language dictionary explaining about 10,000 root Chinese characters for intensive and mild raw character search, respectively, to be able to read an average newspaper.

Therefore, from the top of the most popular Chinese character list, the more the Chinese characters can be arranged into easily memorizable sequential order, the more the Chinese characters can be used for cryptographic applications and Chinese language processing here, as proposed by the author cum inventor Kok-Wah Lee from Kampar, Perak, Malaysia. Table 1 shows the usages of IQ levels (x) together with normal distribution $N(\mu, \delta)$ and $[z = (x - \mu)/\delta]$ (Miller & Powell, 2003) to quantize the list of most frequently used Chinese characters. The currently quantized IQ levels run from the first to 986-th Chinese character. If punctuation marks of the Chinese language seed sentences are considered, then the total symbol set is about 1,000.

For a living creature's thought, it develops from subconscious mind to conscious mind. A person's character decides a habit. The habit decides a behavior. The behavior decides an action. The action decides from internal cause to internal effect (and external cause) to external effect. Likewise, many cause-effect cycles are triggered. In short, one's wishes are factors of one's future undertakings, or one's wish one's future. Here, let us see if the most popular Chinese character list can form quantized sentences under the Chinese speaking people's subconscious minds, be it weakly or strongly correlated storylines.

The author cum inventor Kok-Wah Lee born in year 1975 is fond of Chinese comics since seven years old and Chinese literature since 10 years old. He prefers to read Chinese storybooks including the translated foreign language storybooks outside the school classrooms from primary school times till today. As appreciation, Kok-Wah Lee hereby expresses his gratitude to his parents, who are Hew-Fong Lee and Ah-Mooi Choi, for financially sponsoring him to purchase books during his child times. They cannot be forgotten because they are the founders of Kok-Wah Lee's initial data crystal (or hyperspace knowledge library) in his brain mind for further development till today.

Based on Kok-Wah Lee's capable reach of Chinese literacy as one of his various data crystals till year 2009, Figures 1 and 2 show the quantized Chinese language sentences that can be made by him into the so called Chinese language seeds later (Lee, 2009b, 2009c, 2009d, 2009e). Similar approaches may be applied on other Chinese language popularity lists and other languages to derive their language seeds.

Figure 1 illustrates three Chinese language seeds that have triggered his interests to investigate more. However, these Chinese language seeds have not been fully used for further applications in this article. Figures 2(a) to 2(n) illustrate another 26 Chinese language seeds up to the 986-th Chinese character in the sorted descending list of most popular Chinese characters. These 26 Chinese language seeds are fully used for various proposed applications in this article.

For these seeds, to keep symbol uniqueness in the future for cipher table of polyalphabetic cipher in Chinese language, normally a traditional Chinese character is only converted to its matched simplified Chinese character when there is no ambiguity of uniqueness. To satisfy the entry demand of fast speed for encryption, decryption, and Chinese input method, simplified Chinese character has been chosen as default settings.

--- Better Chinese Language Learning for Better Chinese Language Processing ---

In a first dependent application from the first preferred embodiment of the present independent invention, Chinese language learning has been improved for better Chinese language processing. From the 3+26 Chinese language seeds up to 986 most frequently used unique Chinese characters, they can be observed that these seeds can be weakly or strongly correlated sentences, depending on how good is a storyteller. Comics, fables, novels, music, movies, Chinese language scrabble, etc., are some possible forms to link and deliver the literal and figurative meanings of these Chinese language seeds, so as to build strong memory bases in one's brain.

After forming understanding fundamental, then it is time saving to learn Chinese language, rather than taking abundant time to explain character by character. Knowing these 986 Chinese characters will mean knowing at least 90% of the unique Chinese characters in an average newspaper. Till here, only an average Chinese language dictionary is needed to fully understand and literate in reading the daily news for information collection, opinion exchange, and self-improvement. In short, the Chinese language seeds can help better Chinese language learning directly, and Chinese language processing indirectly.

--- 2D Key Using General Spatial Input Sequence and Specific Sūdoku Input Sequence ---

In a second preferred embodiment of the present independent invention, 2D key using sensitive input sequence has been further improved for computer-free cryptography or paper cryptography. For key style of sensitive input sequence of 2D key, Figure 3 illustrates the 2D key sizes under normal mode and sensitive input sequence mode. It can be observed that transformed 2D matrix at minimum size (3 * 4) is

enough to secure 128-bit symmetric key security. For 256-bit security, the minimum size of transformed 2D matrix is (4 * 5). Comparatively, under normal mode of 2D key without transformation, matrix sizes of (3 * 5) and (3 * 9) are needed for 128- and 256-bit security, respectively. Here, the adopted symbol set of 2D key consists of 986 most popular Chinese characters.

Figures 4(a) to 4(c) illustrate the 2D key transformation from normal to final forms using general spatial input sequence and specific sūdoku input sequence. The various common magic squares of (3 * 3) and (4 * 4), as well as other square and rectangular settings, are hard to be remembered by brain memory. If they are jotted down in note form and hidden somewhere, then its weirdness is too obvious that it may be discovered as a secret key once it is found somewhere.

As better alternative, sūdoku is a better option. This game has been and is currently public favorite for decades. There are lots of books specifically written for people to solve the sūdoku puzzles. Moreover, sūdoku game can sometimes be found in popular daily newspapers as well. Hence, if one of the sūdoku puzzle solution is a transformation key to another normal 2D key, then its normality is hard to be discovered that it is in fact holding double identities as another second secret key to a first normal 2D key. As keynote, both 2D key the secret key using cryptography and sūdoku puzzle the hidden input sequence using steganography have been applied for this proposed secret writing. Of course, the memory burden of hidden sūdoku puzzle solution can be reduced from recall type memory to recognition type memory by searching and picking up a (9 * 9) square in a page of a book.

Nevertheless, all the proposed sensitive input sequences here are still unfriendly to the computer cryptography, because the input time of secret keys to the computer is simply too slow. For paper cryptography, it is different because writing, numbering, counting, and indexing the Chinese characters as a polyalphabetic cipher using pens and papers are simply still tolerable for its various processing times.

--- Computer-Free Cryptography Using Polyalphabetic Cipher in Chinese Language ---

In a second dependent application from the first and second preferred embodiments of the present independent inventions, computer-free cryptography using polyalphabetic cipher in Chinese language has been proposed. In view of electricity power failure and computer absence, computer-free cryptography may be critically, importantly, and imperatively needed in some specific time-space coordinates.

In Figures 5(a) and 5(b), the first Chinese language seed has been used to show a simplified polyalphabetic cipher in Chinese language. Figures 5(a) and 5(b) shows key-plaintext and key-ciphertext cipher tables, respectively. For instance, the key (中) and plaintext (是) give ciphertext (了); whereas the key (中) and ciphertext (了) give plaintext (是).

In short, if the plaintext-to-ciphertext mapping under functions of key symbols is one to one, then key-plaintext cipher table can be used conveniently as key-ciphertext cipher table as well, or vice versa. Hence, it hints that the symbol set sizes of key, plaintext, and ciphertext can be adjusted as the same or different, as long as the plaintext-to-ciphertext mapping under any key symbol function is maintained as one to one. This also indicates that the conversion process between the traditional and simplified Chinese characters has to be careful to ensure one-to-one mapping relationship between the symbol sets of plaintext and ciphertext under key symbol functions.

For larger polyalphabetic cipher in Chinese language, other Chinese language seeds can be added by following the sorted order. Till today, up to 986 unique Chinese characters in 26 Chinese language seeds can be used in the cipher table of polyalphabetic cipher in Chinese language. These 986-character cipher table has entropy at 9.95 (= $\log_2 986$) bits. Therefore, for open cipher table, its secret entropy is purely the secret key size. Meanwhile for hidden cipher table, its secret entropy is the addendum of secret key size and unique symbol set size of cipher table. When 2D key with or without sensitive input sequence, like sūdoku order, has been applied to generate the secret key for Chinese polyalphabetic cipher, then as known from the previous section, a minimum 2D key size of 12 (= 3 * 4) squares or Chinese characters will be enough to satisfy the demand of 128-bit security threshold.

Yet for even larger Chinese polyalphabetic cipher, more Chinese characters in more Chinese language seeds from 987-th sorted Chinese characters and onwards can be included. These can be done as long as the message sender and receiver have common agreement on the cipher table of polyalphabetic cipher to be used. Coming to here, a question is “How about the punctuation marks in Chinese language?”

For short cryptogram, the punctuation marks of Chinese language can be neglected. However, for special short cryptogram, medium cryptogram, and long cryptogram, these Chinese punctuation marks carry critical decisive meanings. Hence, for these cases, the unique symbol set of cipher table of Chinese polyalphabetic cipher has to embed all the possible Chinese punctuation marks. The more unique symbol in a cipher table, the longer is the unicity distance of short cryptogram for impossible decipherability.

In concise context, Chinese polyalphabetic cipher is a paper cryptography having some main advantages like equipment-free operation, cheap cost, space convenience, and time efficiency.

--- Ten Variants of Chinese Polyalphabetic Cipher ---

In a third dependent application from the first and second preferred embodiments of the present independent inventions, eleven variants of Chinese polyalphabetic cipher can be derived to provide higher secret entropy for safer computer-free cryptography. For the first, and fourth to ninth variants, the optionally added game rule here is that the cipher table(s) of Chinese polyalphabetic cipher has to be only

shared by message sender and receiver. In case if the cipher table(s) is open to the public, then its secret entropy will only rely on the secret key.

The first variant is to add weird symbols, like rarely used Chinese characters, to enable flexible size of cipher table of Chinese polyalphabetic cipher. These weird symbols can be added anywhere in sequence, like being appended at the end of the symbol sets of plaintext, ciphertext, and key. Likewise, the second variant is to upgrade the Chinese polyalphabetic cipher into bilingual polyalphabetic cipher using a second Latin language like English language. This can be done by including the English language characters like ASCII printable symbols into the cipher table of Chinese polyalphabetic cipher. Similarly, a third variant exists, where the rarely used Chinese characters may undergo pronunciation Romanization like pinyin for matched sounds, as an indirect inclusion into the plaintext, ciphertext, and key files of bilingual polyalphabetic cipher.

For the fourth, fifth, and sixth variants, the set sizes of plaintext symbols, ciphertext symbols, and key symbols, respectively, can be different. Figures 6(a) and 6(b) illustrate an operational example that the set size of ciphertext symbols is larger, while the set sizes of plaintext symbols and key symbols are the same. Likewise, set size of plaintext symbols or key symbols can be alternatively larger, while the other two set sizes remain the same or different. The rule of thumb to keep these three variants working is that the plaintext-to-ciphertext mapping under any key symbol function has to be always uniquely one to one.

For the seventh, eighth, and ninth variants, they are derived by having variations in the cipher table arrangement for the ciphertext or plaintext symbol set. Figure 7(a) is a normal cipher table shown for comparisons with other variants of Chinese polyalphabetic cipher. The seventh variant shown in Figure 7(b) uses different starting point of ciphertext or plaintext symbol set. Meanwhile the eighth variant shown in Figure 7(c) adopts left-right (or up-down) directing of cipher table. For the ninth variant in Figure 7(d), it applies random order of symbol set lines in cipher table. For the case here in Figure 7(d), it is a vertical symbol set line positioning of the normal case in Figure 7(a) by a 2-step increment to the right and cycled back to the left to repeat.

From the normal, and first to ninth variants of Chinese polyalphabetic cipher, abundantly unique cipher tables can be constructed. Let say all these unique cipher tables are compiled into book form as a single book only or more book volumes. Then we can have the tenth and eleventh variants. For the tenth variant, it is called skipping number of cipher tables in book form. Every consecutive plaintext or ciphertext symbol is referred at different cipher table, gapped by a skipping number jumping from the previous cipher table to the current cipher table in a book.

For the eleventh variant, it is modified from the tenth variant and called sequential scrolling number of cipher tables in book form. Before there is a skipping action for continuous plaintext or ciphertext

symbols, due to skipping number from the current cipher table to the next cipher table in a book, sequential scrolling number delays the skipping action by a few number of continuous cipher tables matched with a continuous plaintext or ciphertext symbol series, before skipping number is effective to take action. Both the skipping number and sequential scrolling number may also be a number series.

In a nutshell, all these variants of Chinese polyalphabetic cipher help to build stronger secret writing systems for computer-free cryptography. To make use of steganography or hidden secret, the cipher tables of these variants have to be kept as secret and open only to message sender and receiver. In case the cipher tables have been disclosed, then it is either to change the cipher tables for future communications, or to rely merely on cryptography or scrambled secret controlled by a secret key like 2D key. 2D key with or without sūdoku input sequence shall have enough entropy secret size to resist not just the human manual attacks, but even the computer brute force attacks. For normal case and some variants of Chinese polyalphabetic cipher, the cipher tables can be easily replaced by mathematical equations to save paper space.

--- Chinese Input Method Using Most Frequently Used Characters ---

In a fourth dependent application from the first preferred embodiment of the present independent invention, a Chinese input method for computing purposes by using most frequently used Chinese characters in Chinese language seeds can be proposed.

From Figure 2(a) to 2(n), 26 Chinese language seeds cover 986 most popular Chinese characters in mnemonic order. These 986 Chinese characters have been coded in these figures for an easy Chinese input method by using the keyboard buttons. For instance, the Chinese character (要) is encoded by code “b7”; whereas another Chinese character (信) is encoded as “dw0” for the first case of punctuation mark included (O+P) and “ds” for the second case of punctuation mark excluded (O-P).

All the 26 Chinese language seeds may appear partially in the input buffer screen once the first coded symbol has been entered. Whenever an input code is unique, then there will be an immediate automatic entry into the computer. All these 986 most popular characters are encoded by two to four alphanumeric symbols at an average of 2.32 symbols. They cover more than 90% of the article context in an average Chinese daily newspaper. So, for the other remaining 10%, the current prior arts of Chinese input method like pinyin input method, Cangjie input method, Wubi input method, etc., may be considered. Alternatively, one may attempt to use another complementary Chinese input method called Yongzi Guaxiang (永字卦象) input method as proposed in this article.

--- Special Stroke Encoding and Yongzi Guaxiang Input Method ---

In a third preferred embodiment of the present independent invention, a special stroke encoding called Yongzi Guaxiang input method (永字卦象输入法) is proposed. This input method may function as an

independent Chinese input method and a complementary Chinese input method to entry using most frequently used Chinese characters in Chinese language seeds. Working under independent mode, this input method may perform ineffectively as compared with Cangjie and Wubi input methods. However, when working under complementary mode, its performance may match with other popular Chinese input methods by having short codes for popular Chinese characters.

From Hua-Yan Luo (罗华炎) (1990, 2003) and Lian-Zhi Zhuang (庄连枝) (1997), it is well-known that Chinese characters have eight basic strokes out of 26 possible stroke types. There is a special Chinese character yong (永) (forever) to have included all these eight basic strokes into a single symbol. From Liang-Bao Chen (陈良保) (1994) and Chih-Hao Tsai (蔡志浩) (1996-2006), it is known that most popularly used Chinese characters have an average nine strokes per Chinese character. Yet from Tie-Kun Wang (王铁琨) (2008), it is known also that one or more Chinese strokes of a Chinese character may have changed from old times to present time, as like the variety of computer fonts for Chinese language text. For example, the Chinese character of ling (零) (zero) has different Chinese strokes as the time passing by.

After pondering over the grouping relationships of Chinese strokes and yijing bagua (易经八卦) since 31 August 2009, the author cum inventor Kok-Wah Lee has applied the categorization power of yijing bagua on the eight basic strokes of Chinese character yong (永). August the 31st is the national holiday of Malaysia to remember the independence day of Malaya Peninsula (previously called Gold Peninsula in ancient times) on 31 August 1957. Therefore, the author has time to spend a one-day trip from Melaka (Malaysia) to Singapore. It is during the leisure time of this trip that the idea on the possible linkage between the Chinese character yong (永) and yijing bagua (易经八卦) started sparkling from a little fire.

Before presenting the updated context, let us look at a new mnemonic way to link the forms (卦象) and binaries (卦爻) of yijing bagua by referring to Figure 8. Below is a list of explanation for their mnemonic linkages called “story of bagua binary” (八卦卦爻的故事):

天 (Sky) (☰): Multiple layers of sky on Earth, like atmosphere, stratosphere, ionosphere, etc.

地 (Earth) (☷): The Earth surface is mainly covered by easily separable ocean waters.

风 (Wind) (☴): The evaporated water becomes gently moving air currents.

山 (Mountain) (☶): When being blocked, the moving air currents move upwards vertically.

火 (Fire) (☲): Fire keeps nothing in the center and burns available surrounding substances.

水 (Water) (☵): Water can be easily separable and can contain substances inside.

雷 (Thunder) (☳): The air currents are so high to have become clouds with different charges.

泽 (Swamp) (☱): A low Earth area can keep abundant water shallowly.

Besides, Figure 8 shows a summary of Kok-Wah Lee's brief but important conclusions. His created Chinese poem "大白山乂成危小子" ("Big white mountain manages to safeguard the dangerous little guy.") has well-represented the intrinsic mnemonic meanings of respective Chinese strokes. Below are the summarized meanings for this poem called "mnemonic poem for bagua" (八卦的易记诗词):

大 (big): A human (人) is always below the sky (天) (representing a horizontal line (横)).

白 (white): The Sun (日) star emits spherical waves as wind (风) (representing falling left (撇)).

山 (mountain): Three vertical cones (representing vertical line (竖)) form mountain (山) range.

乂 (manage): Fengshui (风水) (representing falling right (捺)) peace can pacify lots of situations.

成 (safeguard): Yang Earth (戊，曰阳土) needs a marker milestone (地) (representing dot (点)).

危 (danger): Tragedy (厄) is normally under the wind and thunder (雷) (representing turn (折)).

小 (little): Mountain and swamp can depart the fire (火) (representing uplift (挑)) and water.

子 (guy): A crying (雷) woman giving natural (天) birth is bloody (泽) (representing hook (钩)).

There is no Unicode symbol to solely represent the basic Chinese strokes of falling right (捺) and uplift (挑). Therefore, "'八'字去左取右" (ba) has been used to temporarily represent the Chinese stroke of falling right (捺), by removing the left component; whereas "丫" (by removing the top component) and "灬" have been used to temporarily represent the Chinese stroke of uplift (挑).

For Chinese radical "丫", one can imagine its intrinsic mnemonic meaning by looking for its formable Chinese characters, like 冰 (ice), 泳 (to swim), 冶 (to cast), etc. The Chinese character 冶 (to cast) means a blacksmith casting melted (氵) metallic liquid to form tools on an equipment platform (台). In mnemonic sentence, the Chinese character 冰 (ice) has Chinese radical "丫", having water component (水) (representing falling right (捺) ("八"字去左取右)) at the top part and fire component (火) (representing uplift (挑) (丫 or 灬)) at the bottom part.

When water and fire are in balance state, then it means a substance is at melting point between the physical states of solid and liquid, or at evaporating and boiling points between the physical states of liquid and gas. Looking backwards to the Chinese character 冶, it means an equipment platform to melt metal to form tools. Hence, the Chinese radical "丫" shall have meant closely to the physical states around a melting point temperature. The above-mentioned story can be called "story of Chinese radical at melting point" (汉字部首熔点的故事).

Then, for Chinese character 泳, the Chinese radical "氵" has two water components at the top part and one fire component at the bottom part, which means it is in liquid form. Whenever there is enough water, then one can swim (泳) on the water forever (永). By linking to the story of Great Flood and Noah's ark, then one can see that a family knowing swimming skills can have its offspring survive and sustain forever. This is perhaps a reason why the Jewish families must teach the swimming skills to their children. The description in this paragraph can be called "story of Chinese character yong (to swim)" (泳字的故事).

Coming to here, it is time to present another mnemonic story written by Kok-Wah Lee in this article on the Chinese character yong (永) (forever). This story can be called "story of Chinese character yong (the forever)" (永字的故事). For this Chinese character, it is of left-middle-right (左中右) Chinese character structure (结构). The left part has wind stroke at the bottom and sky stroke at the top, meaning air currents move up. The right part has water stroke at the bottom and fire stroke at the top, meaning water evaporates under the Sun heat. For the middle part, the top portion has thunder clouds mixing to form anions like nitrate ion (NO_3^-), carbonate ion (CO_3^{2-}), and sulfate ion (SO_4^{2-}) to form soil later. The thunder clouds normally exist at the top of mountains. The formed anions precipitate together with the water vapor to drop to the mountain ground as rain. The rain water gathers as it travels down the mountain to form river and swamp at the foothill.

This explains the middle part of Chinese character yong (永) to have earth stroke, thunder stroke, mountain stroke, and swamp stroke from the top to the bottom. Looking backwards at the left and right parts of this Chinese character, it means the wind currents and water vapor move to the mountain and travel upwards before they fall as rain again. In short, these natural phenomena are natural cycles of water, hydrogen, oxygen, carbon, nitrogen, and sulfur, forming the basic elements of living beings. Subsequently, if these natural cycles can be maintained forever, then living beings like human beings can achieve sustainability on Earth planet. Hence, Chinese character yong (永) has pictorially represented the intrinsic meaning to sustain forever the living beings on Earth planet.

Yet, there is another mnemonic story written by Kok-Wah Lee here on the yinyang five elements (金水木火土) (Metal, Water, Wood, Fire, Earth). The metal element sometimes is also called as Wind element. The story has close relationship with the numeric values normally assigned to these elements in Chinese astrology (中华命理学). Here is the story called "story of yinyang 5 elements" (阴阳五行的故事):

金六局 (Metal Six Round): Metal number can be 1 and 6, with 1 wind stroke and other 6 strokes.

水二局 (Water Two Round): Chinese radical of water (氵) has 2 water strokes.

木三局 (Wood Three Round): First and second poem halves in Figure 8 have 3 unique elements.

火四局 (Fire Four Round): Chinese radical of fire (灬) has 4 fire strokes.

土五局 (Earth Five Round): Character 土 has 1 mountain stroke and 4 horizontal strokes.

Till here, it can be commonly observed that every Chinese character can be mnemonically illustrated by a pictorial story for detailed description to hit its intrinsic meaning. Consequently, the Yongzi Guaxiang input method (永字卦象输入法) attempts to use the bagua categorization of basic Chinese strokes to form a special stroke encoding method.

Figure 9 illustrates a system of ASCII keyboard buttons proposed by Kok-Wah Lee to input Chinese characters into a computer system by using the Yongzi Guaxiang input method. All the eight basic Chinese strokes are represented by the most friendly button line of ASCII keyboard. Going up a level for a group of two basic Chinese strokes, if the sequential order of Chinese stroke writing is considered, then it amounts to 64, in which has exceeded the available mnemonically friendly ASCII alphanumeric buttons. Therefore, a question here is "What will happen if the stroke writing order is neglected?"

Subsequently to a possible answer, there exists a need to obtain the mathematics of combination possibility using repeatable set elements. From Kok-Wah Lee's stand-alone calculation and its application, if m is number of members in a combination from a set, and n is number of unique members in an element set, then number of repeatable combination possibilities (P) is in Equation 1 [Eq. 1].

$$P = f_m(n) = {}^{n+m-1}C_m = (n + m - 1)! / [\,(n - 1)! * m!\,] \qquad \text{[Eq. 1]}$$

When n = 8 and m = 2 for bagua and two Chinese strokes, $P = f_2(8) = {}^{9}C_2 = 36$.

This amount is friendly to ASCII keyboard. Consequently, the other partial design of Yongzi Guaxiang input method is formed as in Figure 9. Counting lowercase and uppercase English alphabets, it amounts to 52. Adding up Chinese stroke groups of one and two strokes, they are 44 (= 8 + 36). Hence, there is an empty room consisting of another 8 ASCII alphabetic buttons. Here, let the Chinese character structure groups (结构组合) are also included into the Yongzi Guaxiang method. These included seven Chinese character structures are sole structure (独体结构), left-right structure (左右结构), left-middle-right structure (左中右结构), up-down structure (上下结构), up-middle-down structure (上中下结构), semi-enclosed structure (半包围结构), and fully-enclosed structure (全包围结构). Finally, the remaining ASCII alphabetic button is assigned to represent the Chinese character yong (永), meaning all the eight unique types of basic Chinese strokes are present.

For the same unique Chinese stroke more than two, it is represented by the continuous entry of alphabet representing the Chinese stroke followed by a number. The number of Chinese strokes represented by an alphabetic button is multiplied with the number button value to get the exact number of strokes. After all, space bar button is to represent a wild card of Chinese stroke or stroke amount. When there is uniqueness

in Chinese character encoding of this Chinese input method, then there will be automatic entry. In case there is ambiguity, then one of the listed items in the input buffer screen has to be selected, by pressing the Shift button together with a number key, or a preset ASCII punctuation mark as shown in Figure 9.

So, to enter the Chinese character yong (永), simply enter the ASCII keyboard button 'l', that is lowercase of letter 'L'. For another example like Chinese character 直 (straight), one can imagine there are one sky stroke each at the bottom and top, four sky strokes between two mountain strokes, and a wind stroke on top. The metal element has representative color as "white"; whereas the white color represents "law". For the area between mountains, it is normally a valley having flowing rivers from the mountains, which is geographically suitable for civilization. However, population in an area will only have positive growth rate, when the local residents are majority-wise good, honest, and straight. As in the Bible story, when there are too many bad and evil residents in a town, then there will be signals from God asking the good residents to migrate away from this town. So, when a valley has law represented by one wind stroke and six sky strokes, it is then both geographically and human-wise strategic for civilization. As extra notes, both sky and wind forms of bagua are of metal element. This is perhaps a mnemonic story for 直, which can be called "story of Chinese character zhi (straight)" (直字的故事).

Therefore, possible codes for 直 are [a6], [a6d2], [a6q], [a6d2s], [a6qs], [sqa6], [A3qs], [sqA3], etc. Yet for a third example like Chinese character 真 (true), one can imagine that once straightness has fire element for imagination and earth element for reality, then there will be practical realization. Those possible codes for 真 are [a6qs], [sqa6], [a6qsjg], [sqa6jg], [a6jg], [A3qsjg], [sqA3jg], etc.

In summary, this independent invention requires a user to know the character structure of a Chinese character. Indirectly, it means that this Chinese input method relies on the user's imagination power to link the Chinese character and its representative meaning in pictogram. In other words, it works on the principle that every Chinese character carries a pictorial story.

--- Computer Cryptography Using Polyalphabetic Cipher in Chinese Language ---

In a fifth dependent application from the first, second, and third preferred embodiments of the present independent inventions, as well as the fourth dependent application, one can derive the computer cryptography using Chinese polyalphabetic cipher, and even the bilingual polyalphabetic cipher.

Using the Chinese input methods of most frequently used Chinese characters in Chinese language seeds and Yongzi Guaxiang input method for operation independence, one can have flexible sizes of cipher tables of Chinese polyalphabetic cipher. If cryptography of 2D key the secret is to use together with the steganography of hidden cipher table, then the Chinese input methods proposed in this article, especially the Yongzi Guaxiang input method, can help to contribute high degrees of cipher table flexibility.

Cryptography using 2D key has enough secret entropy to resist even the future quantum computer attacks. Steganography needs the secret cipher tables to be known by the message sender and receiver only. This is convenient in view of the abundant resources of Chinese characters, and today communications technologies for accuracy, speed, and coverage. Therefore, computer cryptography using Chinese polyalphabetic cipher and proposed Chinese input methods here is self-operationally independent, and may become an alternative option to other symmetric key encryption methods and systems.

--- Chinese Polyalphabetic Cipher to Work in Hybrid Communications Network ---

In a sixth dependent application from the first, second, and third preferred embodiments of the present independent inventions, the Chinese polyalphabetic cipher is proposed to work in hybrid communications network. Here, the first added secret writing component is the steganography of one or more hidden communication channels to transmit one or more partial parts of secret. The second added secret writing component is the steganography of Chinese language ambiguity in the formation of phrase, clause, sentence, and paragraph from various individual Chinese characters.

For Chinese language, its specialty is the strong phrase formation of various individual Chinese characters into clause, sentence, and then paragraph. Let assume there is a Chinese language sentence to be transmitted secretly, where this sentence has 5 ordered phrases formed from 10 Chinese characters. Assume again at the first communication channel network, like computer communications network the Internet via email, only the 10 loosely separated individual Chinese characters are transmitted secretly in disorder form. Hence, even if a message intruder manages to steal and decode the ciphertext, the steganography of the Chinese character ambiguity can have multiple possible deciphered plaintexts. In other words, the decipherability problem can be induced by using Chinese polyalphabetic cipher in hybrid communications network.

Then let the second to sixth communication channel networks to carry one or more tag marks, to tell that there exist other pieces of partial secrets to construct a full picture of secrecy. These other communication channels may be radio broadcasting, newspaper articles like advertisement, magazine articles like advertisement as well, television broadcasting, door-to-door brochures and pamphlets, sales points like shops promoting some special products, mobile phone, postal letter, presentation of movie, drama, and music concert, etc. Once the special tag mark in a communicated message is recognized, then a message receiver will try to retrieve other parts of secrets by using the pre-arranged secret extraction algorithms.

Those other possible partial secrets for this 10-character Chinese language sentence may be the first to fifth phrases, and the sequential order of these five phrases. Putting into mathematical calculation, there are 120 (= ${}^5P_5$ in permutation) possible sequences, even if the partial secrets of five formable phrases have been stolen by a message intruder. Out of these 120 possibilities, there shall be some cases at strong trust

levels. If there were 12, then an enemy intruder will need to consume 12 times of effort to launch successful attacks.

In summary, the success percentage of Chinese polyalphabetic cipher in hybrid communications network depends on the ambiguity levels of delivered Chinese characters to form higher semantic levels into phrase, clause, sentence, and paragraph. To increase the ambiguity level of Chinese characters in Chinese language context, one can try to use negative sentence having negative words, especially negative adverbs, like not (不是), no (不), never (从没), nothing (没有), none (无), excluding (除去), without (不含), etc.

--- Chinese Polyalphabetic Cipher to Use Key Strengthening and Multihash Key ---

In a seventh dependent application from the first, second, and third preferred embodiments of the present independent inventions, Chinese polyalphabetic cipher may use key strengthening and multihash key in the computer cryptography, to increase the randomness and to resist the request for larger key size due to computer technology advancement.

Key strengthening in the form of multiple rounds of hash iteration has been part of multihash key (Lee, 2009). Meanwhile, multihash key has a second significant process called hash truncation. Therefore, if 512-bit hash function is used for processing, then the resultant key of multihash key is halved to 256 bits only. The input to multihash key shall be a 2D key and the likes to ensure large enough key length up to 128-bit and 256-bit security strength.

Figure 10 illustrates the encoding process of Chinese polyalphabetic cipher using multihash key. Every resultant key of multihash key cycle in partitioned pieces is used to select one of the cipher tables of Chinese polyalphabetic cipher, and acts as a key to encrypt a plaintext character or letter into a corresponding ciphertext character or letter.

Oppositely, Figure 11 illustrates the decoding process of Chinese polyalphabetic cipher using multihash key. If the set sizes of unique plaintext, ciphertext, and key symbols are the same, as well as there exists one-to-one mapping relationship from plaintext set to ciphertext set under any key symbol function, then the encoding and decoding process of Chinese polyalphabetic cipher using multihash key can share the same cipher table. To note here again, one can derive, prepare, and make multiple cipher tables of Chinese polyalphabetic cipher by using the abundant variety of its variants.

If the richness of cipher tables can be turned into hidden files of steganography known only to message sender and receiver, then higher security strength can be achieved. On the other hand if the cipher tables have to be in public domain, then the secret selection of a particular cipher table and its consecutive sequence series by using cycles of multihash key can remain as another inner defense shield. Multihash key in multiple cycles emerges to be alike having multiple pseudo-keys from an original big secret key.

Here, a big secret key is most possibly created by methods like 2D key. From the pseudo-key transformation of multihash key cycles, Chinese polyalphabetic cipher using this technique appears and seems to have pseudo-longer semantic unicity distance, to become a stronger polyalphabetic cipher. The unicity distance becomes pseudo-longer because the secret key entropy has become pseudo-larger.

--- Other Ciphers to Use Multihash Key and Key Strengthening for Sub-Key Generation ---

In an eighth dependent application from the second and third preferred embodiments of the present independent inventions, as well as the seventh dependent application, other ciphers can replace the cipher table of Chinese polyalphabetic cipher, but keeps on using the key strengthening and multihash key, to inherit its main advantages like pseudo-longer unicity distance.

For the generation cycles of multihash key in Figures 10 and 11, its technique is very independent of the Chinese polyalphabetic cipher. Therefore for other ciphers, like AES (Advanced Encryption Standard) cipher (aka Rijndael cipher), this technique may and can be easily transformed to provide alternative sub-key generation techniques. Subsequently, the secret key entropy will be pseudo-bigger, followed by pseudo-longer unicity distance and harder decipherability.

--- Conclusions ---

Till here in a nutshell, three independent inventions and eight dependent applications have been presented. These inventions may be under the types of independent-dependent, core-edge, and main-side.

Brief Description Using Tables and Drawings

The present invention will now be described in brief details, with references to the accompanying tables and drawings, in which:

--- Brief Description Using Tables ---

Table 1 shows the quantized list of most frequently used Chinese characters using IQ levels and normal distribution up to the 986-th character; and

Table 2 shows the matched values of keyboard input code and ordered digit value excluding punctuation mark.

--- Brief Description Using Figures ---

Figure 1 illustrates the first three formable Chinese language seeds, wherein they are not fully applied in this article;

Figures 2(a) to 2(n) illustrate another 26 formable Chinese language seeds, wherein they are fully applied in this article;

Figure 3 illustrates the key sizes of pure 2D key and 2D key using sūdoku sequence, for polyalphabetic cipher in Chinese language;

Figures 4(a) to 4(c) illustrate the 2D key transformation from normal to final forms using different matrix sizes, as general spatial input sequence and specific sūdoku input sequence;

Figures 5(a) and 5(b) illustrate a simplified polyalphabetic cipher in Chinese language, where only the first Chinese language seed has been used to show key-plaintext cipher table in Figure 5(a) and key-ciphertext cipher table in Figure 5(b);

Figures 6(a) and 6(b) illustrate the key-plaintext and key-ciphertext cipher tables of Chinese polyalphabetic cipher having bigger unique ciphertext symbol set;

Figures 7(a) to 7(d) illustrate normal cipher table and its variants of Chinese polyalphabetic cipher. Figure 7(a) is a normal case. Figure 7(b) shows different starting point of ciphertext or plaintext symbol set. Figure 7(c) shows left-right (or up-down) directing of cipher table. Figure 7(d) shows random order of symbol set lines in cipher table;

Figure 8 illustrates the special stroke encoding of Yongzi Guaxiang input method over the eight basic Chinese language strokes;

Figure 9 illustrates a possible representation of common ASCII keyboard buttons for Yongzi Guaxiang input method;

Figure 10 illustrates the encoding process of Chinese polyalphabetic cipher using multihash key; and

Figure 11 illustrates the decoding process of Chinese polyalphabetic cipher using multihash key.

Detailed Description of the Invention Embodiments Using Tables, Drawings, Mind Mapping Points, and References

The present invention will now be described in greater details, with references to the accompanying tables, drawings, mind mapping points, and references, in which:

--- Detailed Description of the Invention Embodiments Using Tables ---

Table 1 shows the quantized list of most frequently used Chinese characters using IQ levels and normal distribution up to the 986-th character. More sorted members in the most frequently used Chinese character list can be included whenever further Chinese language seeds can be made. The IQ level quantization here may also use different mean and standard deviation. Since language education for working purposes is normally completed at the ordinary level (aka O level) of secondary school, so the suitable upper threshold of IQ level for common Chinese language seeds shall range from 100 to 120, at an average IQ level 110 points.

Table 2 shows the matched values of keyboard input code and ordered digit value excluding punctuation mark. The amounts of keyboard input code at 2, 3, and 4 symbols are 673, 310, and 3 units. Therefore, for the Chinese input method using popular Chinese characters in Chinese language seeds (中文常用种子句输入法), the average code symbol per popular Chinese character is 2.32 (= (2*673 + 3*310 + 4*3) / 986 = 2288/986) symbols.

--- Detailed Description of the Invention Embodiments Using Drawings ---

Figure 1 depicts three formable Chinese language seeds, wherein they are not fully applied in this article. However, one has to know that these three Chinese language seeds are deep in the subconscious mind of Chinese language speaking people, which may influence their habits, behaviors, and actions. Also, it indicates that sorted popular Chinese characters may be used repeatedly.

Figures 2(a) to 2(n) depict another 26 formable Chinese language seeds, wherein they are fully applied in this article. Here, one has to know that the Chinese language seeds may change slightly from times to times, especially the sorted popular Chinese characters in low descending ranking. Moreover, Chinese language seeds may differ among different geo-political regions. Different languages may have their own unique language seeds because of their own unique climates, geography, culture, unwritten rules, written laws, etc.

Figure 3 depicts the key sizes of pure 2D key and 2D key using sūdoku sequence, for polyalphabetic cipher in Chinese language. They are based on a unique character set of 986 most popular Chinese characters, having 9.95 (= $\log_2 968$) bits. Sharp foci have to be seen at 128- and 256-bit security thresholds, as represented by two horizontal lines in the graph. This is because this Chinese polyalphabetic cipher is a symmetric key encryption. 256-bit security strength has already been able to resist even the future possible quantum computer attacks.

Figures 4(a) to 4(c) depict the 2D key transformation from normal to final forms using different matrix sizes, as general spatial input sequence and specific sūdoku input sequence. One has to be aware that the reading style of 2D matrix elements for its rows and columns may be in variety. It can be horizontally left or right, vertically up or down, diagonally clockwise or anti-clock-wise, mixed, etc. For specific sūdoku

input sequence, the 10-digit filling may start at arbitrary numeric digit at different location. The step changes may be static increment or decrement, dynamic increment or decrement, or mixed. In short, there are many other possibilities to contribute for higher randomness in favor of stronger security. For Box **407** in Figure 4(c), add base-10 digit at 10^1 position from digit '0' by a 1-step increment from top left (3 * 3) square, to top right (3 * 3) square, to middle left, to middle right, to bottom left, to bottom right, and cycled back to top left, until all the nine (3 * 3) squares have all the single digits transformed into double digits as in Box **408**. For Box **409**, sort the numbers in (4 * 4) matrix from (9 * 9) sūdoku square into ascending order. Arrange these sorted numbers into another (4 * 4) matrix starting from top left, to top right, to leftmost of one vertical line down, to rightmost of one vertical line down, and so on, till bottom left, and bottom right. Match this arranged ascending number series with the 2D key in Box **406**. Then, re-arranged the number indexed 2D key in Box **406** using number order in Box **409** to generate a final 2D key form as in Box **410**.

Figures 5(a) and 5(b) depict a simplified polyalphabetic cipher in Chinese language, where only the first Chinese language seed has been used to show key-plaintext cipher table in Figure 5(a) and key-ciphertext cipher table in Figure 5(b). For larger cipher table of Chinese polyalphabetic cipher, simply append more Chinese language seeds in sorted order of numeric values. In addition, these Chinese language seeds may also be in different order, as long as the individual sentences are kept as a whole for mnemonic purposes. For bilingual polyalphabetic cipher using Chinese language and a Latin language like English language, a straight approach is to put the English alphabets at the beginning of the cipher table first before other Chinese characters.

Figures 6(a) and 6(b) depict the key-plaintext and key-ciphertext cipher tables of Chinese polyalphabetic cipher having bigger unique ciphertext symbol set. An important rule of thumb here is to ensure one-to-one mapping relationship under any key symbol function from unique plaintext symbol set to unique ciphertext symbol set.

Figures 7(a) to 7(d) depict normal cipher table and its variants of Chinese polyalphabetic cipher. Figure 7(a) is a normal case. Figure 7(b) shows different starting point of ciphertext or plaintext symbol set. Figure 7(c) shows left-right (or up-down) directing of cipher table. Figure 7(d) shows random order of symbol set lines in cipher table. As shown in the figures here, there can be abundant varieties to form lots of variants for Chinese polyalphabetic cipher.

Figure 8 depicts the special stroke encoding of Yongzi Guaxiang input method over the eight basic Chinese language strokes. This figure is a mnemonic chart to use Yongzi Guaxiang input method. Do print a copy if one wants to use this method.

Figure 9 depicts a possible representation of common ASCII keyboard buttons for Yongzi Guaxiang input method. This figure is an important chart to use Yongzi Guaxiang input method via ASCII keyboard. Do print a copy as well if one wants to use this method.

Figure 10 depicts the encoding process of Chinese polyalphabetic cipher using multihash key. Box **1000** is the initialization settings. Boxes **1001** and **1002** are the encoding operations on plaintext P character by character. Box **1003** transmits the readily encoded ciphertext C.

Figure 11 depicts the decoding process of Chinese polyalphabetic cipher using multihash key. Box **1100** receives the transmitted ciphertext C, which may have errors to be detected and corrected during the message and signal transmissions. Box **1101** is the initialization settings alike the encoding process. Boxes **1102** and **1103** are the decoding operations on ciphertext C character by character, where everything is the same as encoding process except the interchange of plaintext and ciphertext parameters. Box **1104** has the readily decoded plaintext P to be read.

--- Detailed Description of the Invention Embodiments Using Mind Mapping Points ---

There are three independent inventions (I.1-I.3) and eight dependent applications (D.1-D.8) presented in this article.

(I.1) Core 1, Main 1: Method and system to create Chinese language seeds from most frequently used Chinese characters, normal distribution, and IQ (Intelligent Quotient).

(I.2) Core 2, Main 2: 2D key the big memorizable secret using sensitive input sequence like:

(I.2.1) general spatial input sequence

(I.2.2) specific sūdoku (数独) input sequence

(I.3) Core 3, Side 1: Yongzi Guaxiang input method (永字卦象输入法) as independent and complementary Chinese input methods.

(D.1) Edge 1: Faster learning of Chinese language and Chinese language processing.

(D.2) Edge 2: Chinese polyalphabetic cipher as a secure computer-free cryptography.

(D.3) Edge 3: Variants of Chinese polyalphabetic cipher using the following features:

(D.3.1) weirdly added symbols

(D.3.2) bilingual languages using a second Latin language like English language

(D.3.3) pronunciation Romanization

(D.3.4) different set size of plaintext symbols

(D.3.5) different set size of ciphertext symbols

(D.3.6) different set size of key symbols

(D.3.7) starting point of ciphertext or plaintext symbol set

(D.3.8) left-right (or up-down) directing of cipher table

(D.3.9) random order of symbol set lines in cipher table

(D.3.10) skipping number of cipher tables in book form

(D.3.11) sequential scrolling number of cipher tables in book form

(D.4) Edge 4: Chinese input method using popular Chinese characters in Chinese language seeds (中文常用种子句输入法).

(D.5) Edge 5: Computer cryptography using Chinese polyalphabetic cipher.

(D.6) Edge 6: Chinese polyalphabetic cipher working in hybrid communications network.

(D.7) Edge 7: Chinese polyalphabetic cipher using key strengthening and multihash key.

(D.8) Edge 8: Other ciphers using multihash key for sub-key generation.

Based merely on independent invention (I.1), there are dependent applications (D.1, D.4).

Based on independent inventions (I.1-I.2), there are dependent applications (D.2-D.3).

Based on independent inventions (I.1-I.3) and dependent application (D.4), there is only dependent application (D.5).

Based on independent inventions (I.1-I.3), there are dependent applications (D.6-D.7).

Based on independent inventions (I.2-I.3) and dependent application (D.7), there is only dependent application (D.8).

A note here is that dependent application (D.1) is a mental act for efficient Chinese language learning, critically important for almost all the application derivatives in this article.

--- Detailed Description of the Invention Embodiments Using References ---

[1] Hong-Qing Yang (杨洪清)，Xin-Lan Zhu (朱新兰). (1999, July). Modern Dictionary of Language Speaking and Character Explanation (现代说文解字字典). ISBN: 750-141647-8. Beijing (北京), China (中国): Public Press (群众出版社). ISBN: 750-141647-8. (in Chinese language).

[2] Qi-Hong Xiao (萧启宏). (2004, March 1). Han Character Book of China (Vol. 1 & 2) (中国汉字经（上下册）). ISBN-13: 978-780-187026-1. Beijing (北京), China (中国): New World Press (新世界出版社). ISBN: 978-780-187026-1. (in Chinese language).

[3] Bing-Yi Chang (常秉义). (2000, October). Zhouyi and Han Characters (周易与汉字). ISBN: 722-806089-X. Ürümqi (乌鲁木齐), Xinjiang (新疆), China (中国): Xinjiang People Press (新疆人民出版社). ISBN: 722-806089-X. (in Chinese language).

[4] Kok-Wah Lee. (2008, December 18). Methods and Systems to Create Big Memorizable Secrets and Their Applications in Information Engineering. PCT Patent Application [No.: PCT/IB2008/055432]. Geneva, Switzerland: WIPO (World Intellectual Property Organization). Filing Date: 18 December 2008.

[5] Kok-Wah Lee. (2009, March 14). Memorizable Public-Key Cryptography (MePKC) & Its Applications. Internet Archive. URL: http://www.archive.org/details/MemorizablePublic-keyCryptographymepkcItsApplications [2009, December 22].

[6] Zu-Qiu Hong (洪祖秋). (2009, September 13). Snake Dialect and Black Society Language (蛇话与黑社会切口). Petaling Jaya, Selangor, Malaysia: Sin Chew Daily (星洲日报). (in Chinese language).

[7] Xiang Liu (刘向), and Xin Liu (刘歆) (Eds.). (2002, January). Book of Mountains and Oceans: Pictorial Text Version (山海经：图文本). Beijing (北京), China (中国): Religious Culture Press (宗教文化出版社). ISBN: 780-123412-X. (in Chinese language).

[8] Chih-Hao Tsai (蔡志浩). (1996-2006). Frequency and Stroke Counts of Chinese Characters. Chih-Hao Tsai's Technology Page. URL: http://technology.chtsai.org/charfreq/ [2006, March 06]. (in English and Chinese languages).

[9] Global Language Monitor. (2009, June 10). URL: http://www.languagemonitor.com [2009, June 15].

[10] Lian-Zhi Zhuang (庄连枝). (1997, September 07 & 14). Chinese Language is a Scientific, Graceful, and Wise Language (中文是科学、优美、智慧的文字). Petaling Jaya, Selangor, Malaysia: Nanyang Siang Pau (南洋商报). (in Chinese language).

[11] J. C. P. Miller, and F. C. Powell (Eds.). (2003). The Cambridge Elementary Mathematical Tables (3rd ed.). Shah Alam, Selangor, Malaysia: Federal Publications Sdn. Bhd. (A member of the Times Publishing Group). ISBN: 967-914477-1.

[12] Kok-Wah Lee. (2009b, January 16). Analysis Seed to Learn Chinese Language More Efficiently. Internet Archive. URL: http://www.archive.org/details/AnalysisSeedToLearnChineseLanguageMoreEfficiently [2009, December 22].

[13] Kok-Wah Lee. (2009c, February 20). Learning Seed of Chinese Language (version 2). Internet Archive. URL: http://www.archive.org/details/LearningSeedOfChineseLanguageversion2 [2009, December 22].

[14] Kok-Wah Lee. (2009d, May 01). Chinese Lexeme Prose. Internet Archive. URL: http://www.archive.org/details/ChineseLexemeProse [2009, December 22].

[15] Kok-Wah Lee. (2009e, May 30). Learning Seed of Chinese Language (version 3). Internet Archive. URL: http://www.archive.org/details/LearningSeedOfChineseLanguage1ed-ver3 [2009, December 22].

[16] Hua-Yan Luo (罗华炎). (1990). Concise Chinese Grammar (简明汉语语法). Cheras, Kuala Lumpur, Malaysia: Yakin (雅景). ISBN: 967-770088-X. (in Chinese language).

[17] Hua-Yan Luo (罗华炎). (2003). Modern Chinese Grammar (revised ed.) (现代汉语语法：增订版). Ipoh, Perak, Malaysia: Penerbitan Seni Hijau (艺青出版社). ISBN: 983-962656-6. (in Chinese language).

[18] Liang-Bao Chen (陈良保). (1994). Chinese Stroke Sequence Dictionary of Primary School (小学笔顺字典). Kuala Lumpur, Malaysia: Penerbitan Ta Lian (大联出版社). ISBN: 983-914900-8. (in Chinese language).

[19] Tie-Kun Wang (王铁琨) (Ed.). (2008, June). General Dictionary of Shape Uprighting of Chinese Characters (通用汉字正形字典). United Publishing House (M) Sdn. Bhd. (联营出版（马）有限公司), Seri Kembangan, Selangor, Malaysia. ISBN: 978-983-010334-1. (in Chinese language).

Objectives of the Present Invention

It is an object of the present invention to provide Chinese language seeds for Chinese polyalphabetic cipher and Chinese language process, which overcome the deficiencies of existing information security systems for computer-free cryptography, as well as existing Chinese input methods for simple and yet straight encoding.

Additional objects, advantages, novel features of the present invention will become apparent to those skilled in the art from this disclosure, including the previous and following detailed descriptions, as well as by practice of the invention. While the invention is described in this article with reference to preferred embodiment(s), it should be understood that the invention is not limited thereto. It will also be appreciated that the preferred embodiment is illustrative only and that various changes may be made by those skilled in the art without departing from the spirit and scope of the invention.

Yet it will also be recognized by those skilled in the art that, while the invention has been described above in terms of one or more preferred embodiments, it is not limited thereto. Various features and aspects of the above described invention may be used individually or jointly. Further, although the invention has been described in the context of its implementation in a particular environment and for particular purposes, e.g. in providing computer-free cryptography and computer security for local and networked Internet communications, those skilled in the art will recognize that its usefulness is not limited thereto and that the present invention can be beneficially utilized in any number of environments and implementations.

Those of ordinary skill in the art, having access to the teachings herein, will recognize additional implementations, modifications, and embodiments, as well as other fields of use, in which are within the full breath, spirit, and scope of the invention as disclosed and claimed herein, and with respect to which the invention could be of significant utility.

Claims:

1. Method and system to create Chinese language seeds from most frequently used Chinese characters, normal distribution, and IQ (Intelligent Quotient), in which:

(a) lower descending ranking Chinese characters may be added for more Chinese language seeds;

(b) there is list variety of most frequently used Chinese characters;

(c) Chinese language seeds may differ among different geo-political regions;

(d) IQ level using normal distribution may have different mean and standard deviation;

(e) IQ levels may range in a slightly different mode for different aims of language education;

(f) other languages may have their own language seeds due to differently unique climates, geography, culture, unwritten rules, written laws, etc.;

(g) Chinese language seeds help improve the efficiency to learn Chinese language;

(h) Chinese language seeds can be applied for computer-free cryptography, Chinese language processing, and computer security;

(i) Chinese language seeds can be in traditional and simplified Chinese characters, or mixed;

(j) Chinese language seeds may change slightly from times to times; and

(k) mnemonic order exists in Chinese language seeds to realize Chinese polyalphabetic cipher and Chinese input method using most frequently used Chinese characters.

2. Method and system to create 2D key (aka two-dimensional key) the big memorizable secret using sensitive input sequence, like general spatial input sequence and specific sūdoku input sequence, in which:

(a) other key styles of 2D key may be used;

(b) sūdoku square may be of sizes (9 * 9), (12 * 12), etc.;

(c) there may be other abundant variety to manipulate the sūdoku square;

(d) sensitive input sequence may be worked on papers and in computer systems;

(e) 2D key may use different monolingual languages on papers and in computer systems;

(f) 2D key may use bilingual and multilingual languages; and

(g) 2D key may use different character encodings other than ASCII and Unicode.

3. As in Claim 1, there are applications of Chinese language seeds from Claim 1, for Chinese input method of Chinese language processing using most frequently used Chinese characters in Chinese language seeds, wherein:

(a) there are 26 main Chinese language seeds holding up to 986 most popular Chinese characters;

(b) more popular Chinese characters can be added beyond the 986-th Chinese character;

(c) upon entry of unique symbol code, there is automatic entry into computer;

(d) there is input buffer screen for Chinese character selection; and

(e) similar language seeds in other languages can also be a friendly character input method into computer.

4. As in Claims 1 and 2, there are applications of Chinese language seeds from Claim 1, and 2D key the big memorizable secret using sensitive input sequence from Claim 2, for Chinese polyalphabetic cipher as a secure computer-free cryptography, wherein:

(a) other big memorizable secret generation methods may be used;

(b) computer-free cryptography may work on paper or any other convenient written media;

(c) cipher table of Chinese polyalphabetic cipher can be enlarged by adding more Chinese language seeds;

(d) sequence of Chinese language seeds in cipher table of Chinese polyalphabetic cipher may be random, as long as every individual Chinese language seed arranged as a whole connected unit;

(e) Chinese polyalphabetic cipher is of critically imperative applications for the cases of electricity power failure and computer absence;

(f) Chinese polyalphabetic cipher using 2D key and the likes can achieve beyond 256-bit security threshold to resist even the future possible quantum computer attacks;

(g) it is a paper cryptography having some main advantages, like equipment-free operation, cheap cost, space convenience, and time efficiency; and

(h) for normal case and some variants of Chinese polyalphabetic cipher, the cipher tables can be easily replaced by mathematical equations to save paper space.

5. As in Claims 1 and 2, there are applications of Chinese language seeds from Claim 1, and 2D key the big memorizable secret using sensitive input sequence from Claim 2, for variants of Chinese polyalphabetic cipher using various possible features, wherein:

(a) weirdly added symbols may be used in the cipher table for flexible size;

(b) it can be bilingual polyalphabetic cipher using a second Latin language like English language;

(c) pronunciation Romanization may be used to add Romanized Chinese character in pinyin;

(d) different set size of plaintext symbols may induce variety of cipher table;

(e) different set size of ciphertext symbols may induce variety of cipher table;

(f) different set size of key symbols may induce variety of cipher table;

(g) different starting point of ciphertext or plaintext symbol set may induce cipher table variety;

(h) left-right (or up-down) directing of cipher table may induce variety as well;

(i) random order of symbol set lines in cipher table may be used;

(j) skipping number of cipher tables in book form is also practical;

(k) sequential scrolling number of cipher tables in book form can be additionally applied;

(l) hybrid mixing of all the previous possible features; and

(m) the more variety of cipher table amount, the longer the unicity distance, the harder the decipherability problem.

6. Method and system to create special stroke encoding of Chinese characters using the mnemonic relationship of basic Chinese strokes and yijing bagua for an application as a Chinese input method called Yongzi Guaxiang input method, in which:

(a) each basic Chinese stroke is matched only with a gua sign of yijing bagua;

(b) categorization power of yijing bagua is needed;

(c) mnemonic story for the relationship between the Chinese strokes of a Chinese character and categorization power of yijing bagua is needed;

(d) this Chinese input method may work independently, and act as a complementary Chinese input method to Claim 3;

(e) ASCII keyboard can be and is used to implement this Chinese input method;

(f) keyboards using other character encodings may implement this Chinese input method as well;

(g) ASCII keyboard has stroke groups represented by alphabetic buttons for single and double Chinese strokes;

(h) double Chinese stroke group in the ASCII keyboard neglects the stroke writing order, but care only for their combination existence;

(i) Chinese character structure group has also been used in the ASCII keyboard to assist this Chinese input method;

(j) Chinese character yong the forever is in the ASCII keyboard to note the existence of eight unique basic Chinese strokes;

(k) space bar button in the ASCII keyboard acts as a wild card button;

(l) entry of a stroke group button in the ASCII keyboard, followed by entry of a number key, has the stroke amount multiplied with the number key value to get the total stroke amount, and to save the input symbol amount;

(m) character input buffer screen is present to assist the Chinese input method;

(n) Shift key, number keys, and some specially assigned punctuation mark buttons help to select a listed Chinese character in the input buffer screen;

(o) upon entry of unique symbol code, there is automatic entry into computer; and

(p) similar languages using CJKV characters (CJKV = Chinese, Japanese, Korean, and Vietnamese) (aka Han characters, or Chinese characters) may work as friendly character input method into computer as well.

7. As in Claims 1, 2, 3, and 6, there are applications of Chinese language seeds from Claim 1, 2D key the big memorizable secret using sensitive input sequence from Claim 2, Chinese input method using most frequently used Chinese characters in Chinese language seeds from Claim 3, and Yongzi Guaxiang input method from Claim 6, for computer cryptography using Chinese polyalphabetic cipher, wherein:

(a) more Chinese language seeds may be added to the cipher table;

(b) 2D key the big memorizable secret creation method and the likes may be used;

(c) there may be variants of Chinese polyalphabetic cipher alike the situation in computer-free cryptography;

(d) the more variety of cipher table amount, the longer the unicity distance, the harder the decipherability problem;

(e) bilingual polyalphabetic cipher using a second Latin language like English language; and

(f) it has enough security strength to be alternative of other symmetric ciphers.

8. As in Claims 1, 2, and 6, there are applications of Chinese language seeds from Claim 1, 2D key the big memorizable secret using sensitive input sequence from Claim 2, and Yongzi Guaxiang input method from Claim 6, for Chinese polyalphabetic cipher working in hybrid communications network, wherein:

(a) strong ambiguity power of Chinese characters to form phrase, clause, sentence, and paragraph is applied;

(b) a Chinese language ciphertext is partitioned into a few pieces of secret;

(c) for each communications network, only one or more, but not all, pieces of secret is delivered;

(d) secret pieces may be separated individual Chinese characters, sequence order of phrases, clauses, sentences, and paragraphs;

(e) every secret piece may be recognized by using special tag mark in a transmitted message; and

(f) negative sentence using negative words, especially negative adverbs like not, no, never, nothing, none, excluding, without, etc., may be used to increase the ambiguity strength of Chinese language plaintext, in order to waste message intruder's effort to launch successful attacks.

9. As in Claims 1, 2, and 6, there are applications of Chinese language seeds from Claim 1, 2D key the big memorizable secret using sensitive input sequence from Claim 2, and Yongzi Guaxiang input method from Claim 6, for Chinese polyalphabetic cipher using key strengthening and multihash key, wherein:

(a) every cycle of multihash key helps to select a cipher table of Chinese polyalphabetic cipher and provides key pieces to encode plaintext character by character into ciphertext;

(b) cycles of multihash key induce pseudo-larger secret keys, contributing pseudo-longer unicity distance for harder decipherability; and

(c) the more cipher tables and/or the harder to locate hidden cipher tables, the stronger the cryptographic and steganographic strength.

10. As in Claims 2, 6, and 9, there are applications of 2D key the big memorizable secret using sensitive input sequence from Claim 2, Yongzi Guaxiang input method from Claim 6, and Chinese polyalphabetic cipher using multihash key from Claim 9, for other ciphers using multihash key for sub-key generation, wherein:

(a) other ciphers may be AES (Advanced Encryption Standard) cipher and the likes;

(b) cycles of multihash key generate sub-keys for inner cryptographic operations of cipher; and

(c) cycles of multihash key induce pseudo-larger secret sub-keys, contributing pseudo-longer unicity distance for harder decipherability.

5