Data Encryption Based on Exponential Numbers

Ho-Hsuan Chang
Communication Engineering Department, I-SHOU University, Kaohsiung, Taiwan
E-mail: hhchang@isu.edu.tw

Abstract. A novel data encryption scheme using a set of exponential numbers is proposed in this paper which aims at achieving secure communications over public networks. The study begins with the definition of pattern of exponential number, where the decimal pattern of an exponential number \( g^e \) is the result of taking \( e - 1 \) times multiplication the base element \( g \) by itself. The multiplication operation of two large integers involves the complicated carry conditions in each step, which is digit-dependence and can only be analyzed case by case. This implies that the pattern of a large exponential number is unpredictable. Next, we prove that nature number can be spanned by a set of exponential numbers, and there exists numerous spanning sets for one nature number, demonstrating that the expression of nature number using a set of exponential numbers is not unique. The unpredictable nature of exponential patterns and the abundant available sets of exponential numbers can contribute the concealment of information when data streams are encrypted by the exponential numbers. In addition, we prove that the base elements of a set of exponential numbers applied for spanning a nature number can be arbitrarily assigned. With this property, we can distribute a secret key sequence in advance as the base elements for data encryption, or this secret key can be shared using the Diffie-Hellman Key-Exchange system over public network. Finally, the encrypted message consists of a sequence of exponents and the index data, which indicates the relationship between the base elements and the secret key for decryption. The ciphertext is transmitted via the public channel to the receiver end; however, it can only be decrypted by those who own the secret key. Some examples are shown for demonstration the feasibility of the proposed encryption scheme.

1. Introduction

Encryption is the process of converting ordinary information (plaintext) into unintelligible text (ciphertext), which can only be read if decrypted. A cipher is a pair of algorithms that encrypt the plaintext and decrypt the ciphertext. The operation of a cipher is controlled by the algorithm and by a key in each instance. Cryptosystems are categorized into two types: symmetric and asymmetric. In symmetric systems, the same private key is used to encrypt and decrypt a message. Asymmetric systems use a public key to encrypt a message and a private key to decrypt it. Symmetric models include the commonly used advanced encryption standard, which replaced the older data encryption standard [1]. Asymmetric systems include the Rivest-Shamir-Adleman (RSA) algorithm [2], the Diffie-Hellman key exchange algorithm [3], the digital signature standard [4], and the elliptic curve cryptography [5, 6].

Public-key cryptography do not require a secure channel for the initial exchange of one (or more) secret keys between the parties; thus, it is often used to secure electronic communication over an open networked environment such as the Internet. In contrast, symmetric cryptosystem should deal with private key distribution and management problem, which the cost and delay imposed by key distribution is a major barrier to the transfer of business communications to
large networks or Internet. With the fast development of Internet and the high demand of secure communications across public networks, public-key cryptography attracts much attention than private-key cryptography because it is less expensive and affordable.

A novel data encryption scheme using a set of exponential numbers is proposed in this paper which aims at achieving secure communications over public channel. At first, we define the pattern of an exponential number in Section 2, where the decimal pattern of $g^e$ is the result of taking multiplication $e - 1$ times the base element $g$ by itself. We prove in Section 3 that nature number can be spanned by a set of exponential numbers, and show also the existence of numerous spanning sets for a nature number. In addition, the base elements of a set of exponential numbers applied for spanning a nature number can be arbitrarily assigned. With this property, we can distribute a secret key in advance as the base elements for data encryption, or this secret key can be shared using the Diffie-Hellman Key-Exchange system. In Section 4, encryption examples are shown for demonstration the feasibility of the proposed encryption scheme. Finally, we draw conclusions in Section 5.

2. Pattern of Exponential Number $g^e$

Exponentiation is a mathematical operation, written as $g^e$, involving two numbers, the base $g$ and the exponent or power $e$. In this paper both $g$ and $n$ are considered nonnegative integers, which $g^e$ is an exponential number. Let $g^e = d_md_{m-1} \cdots d_0$ denote the result of $g^e$ from taking $e - 1$ times multiplication operations upon $g$, and $g^e$ has $m + 1$ digits in this expression. In a base-$k$ numerical expression the value of the leftist digit is $d_m \times k^m$, the second digit is $d_{m-1} \times k^{m-1}$, etc., and $d_n \in \{0,1,\ldots,k-1\}$, $n = 0, 1, \ldots, m - 1$.

Definition 1 The decimal pattern of an exponential number, denoted by $(g^e)$, $(g^e) \equiv (d_md_{m-1} \cdots d_0)$ is defined as the distribution of decimal digits $\{d_n\}^m_0$ in an $(m + 1)$-tuple vector evaluated using a base-$k$ number system, where $d_m \neq 0$.

In base-10 number system, $(7^{108}) \equiv (18646,11341,71613,14493,26161,68039,56698,97144,64158,58248,58341,59291,20740,30215,43182,06422,51722,18248,01)$, which indicates that the decimal pattern $(7^{108})$ consists of 92 decimal digits. Though we can apply addition chain algorithm for fast calculating $g^e$ to obtain its pattern; however, except those of $10^s$, there exist neither the explicit relationship between two patterns of $g^e = d_md_{m-1} \cdots d_0$ and $g^{e+k} = c_mc_{n-1} \cdots c_0$, nor two patterns between $g^e$ and $(gh)^e$. The reason is that the multiplication of two large integers can bring complicated carry conditions which carry is digit-dependence and can only be analyzed case by case e.g., $7^{108}$=$7 \times 7^{107}$=$7\cdot\{26637,30488,16590,20704,65945,25770,80998,53063,77369,40355,11916,56132,96290,03077,59743,77464,53174,03543\}$, where the $(7^{107})$ pattern has 90 digits. It is unlikely one can use a function to express the relationship between $(7^{107})$ and $(7^{108})$ two patterns, which these two patterns are dramatically different.

3. Exponential Spanning Set and It’s Properties

Considering the 92 digits of $(7^{108})$ are the information stream $S$ to be processed, then we prefer using, e.g., a mixed symbol $7 \wedge 108$, to pack and store this set of 92 symbols, which can save significantly the memory space and the channel bandwidth required to transmit them. Even though the information stream may not be able to express using a single exponential pattern, e.g., $S = (7^{108}) + S_d$, we need only to encode the part of $S_d = S - (7^{108})$ using additional exponential numbers.

Definition 2 When a nature number $S_1 > 3$ is not the pattern of any exponential number $g^e$ with $e > 1$, there exists $g_1^{e_1}$ such that $S_1 = g_1^{e_1} + S_2$. In this expression $S_2$ is defined as the residual part of $S_1$. 
With the definition of the residual part of a nature number, any nature number larger than 3 can be the pattern of exponential number after the deduction of its residual part, e.g., \( g_1^3 = S_1 - S_2 \). The residual part of a nature number can be the pattern of another exponential number after the deduction of its residual part, etc. This process can continue to find a set of exponential numbers which can be applied to express a nature number instead of a base-\( k \) number system. We can call this set of exponential numbers the \textit{spanning set} of a nature number.

**Theorem 1** shows that arbitrary set of digits can be spanned by a set of exponential patterns.

**Theorem 1** Let \( S_1 = b_0b_{n-1} \cdots b_0 \) be a nature number with \( n + 1 \) decimal digits which is not an exponential number \( g^e \) with \( e > 1 \). There exist a set of exponential numbers \( \{g_1^e, g_2^e, \ldots, g_k^e\} \) such that \( S_1 = g_1^e + g_2^e + \cdots + g_k^e \), where \( k \geq 2 \).

**Proof:** Let \( h = \lfloor S_1 \rfloor \) and \( g_1 = \lfloor h \rfloor \), where \( \lfloor h \rfloor \) denotes the greatest integer less than or equal to \( h \). Then we have \( S_1 = g_1^e + S_2 \), where \( S_2 \) is the residual part of \( S_1 \). Assume \( (g_1^e) \equiv (b_0b_{n-1} \cdots b_{n-m+i+1}b_{n-m+i} \cdots b_0) \), where the first \( m - i \) digits, \( b_0b_{n-1} \cdots b_{n-m+i+1} \), are the same as those of \( S_1 = b_0b_{n-1} \cdots b_0 \). This implies that \( S_2 = b_{n-m+i} \cdots b_{n-m+i} - b_{n-m+i+1} \cdots b_0 \) and the length of \( S_2 \) is denoted by \( |S_2| = n - m + i + 1 \), where \(| \cdot |\) denotes the length of argument. This process successfully packs the first \( m - i \) digits of \( S_1 \) into \( g_1^e \). Similarly, we can locate the second exponential number \( g_2^e \) such that \( S_2 = g_2^e + S_3 \), where \( S_3 \) is the residual part of \( S_2 \) and \( |S_3| \leq |S_2| \). These iterations continue until \( S_{k-1} = g_{k-1}^e + S_k \) and \( S_k = g_k^e \). We would mention that in case of \( S_k \in \{0, 1, 2, 3, 5, 6, 7\} \) all \( e_k = 1 \).

Let’s present an example for demonstration, e.g., \( 10001 = 10^4 + 1 = 80^2 + 60^2 + 1 = 21^3 + 9^3 + 11 \). This example implies that there exist different spanning set for a nature number.

**Remark 1** Considering \( S_1 = (b_0b_{n-1} \cdots b_0) \) be a \((n + 1)\)-tuple vector, there exist different sets of exponential numbers \( \{g_1^e, g_2^e, \ldots, g_k^e\} \) that can span \((b_0b_{n-1} \cdots b_0)\). In other words, the decomposition of \( S_1 = g_1^e + g_2^e + \cdots + g_k^e \) into a set of exponential numbers is not unique.

When the length \( n + 1 \) of \( S_1 = (b_0b_{n-1} \cdots b_0) \) is large, the computing load from locating proper exponential patterns to span a large integer is heavy. The Corollary 1 can be applied to tackle this problem.

**Corollary 1** The \((n + 1)\)-tuple vector \( S_1 = b_0b_{n-1} \cdots b_0 \) can be partitioned into a disjoint union of \( m \) blocks which are cascading one after another in order, denoted by \( S_1 = S_{11} \cup S_{12} \cdots \cup S_{1m} \), and \( |S_{11}| + |S_{12}| + \cdots + |S_{1m}| = n + 1 \). Here each block can be packed independently such that \( S_{1r} = g_{1r}^e + g_{2r}^e + \cdots + g_{kr}^e \) can be arbitrarily assigned. In other words, we can define in advance a secret decimal sequence of length \( N > n \), \( C = \{c[i]\}_{i=0}^{N-1} \) to locate sequentially \( g_1 = c[0]c[1] \cdots c[a], g_2 = c[a+1]c[a+2] \cdots c[a+k] \), etc., where the values of \( a \) and \( k \) are evaluated according to \( e_1 \) and \( e_2 \) which are chosen to pack \( S_1 \).

**Proof:** 1) Let’s consider a single digit case \( S_1 = b_0 \). If \( b_0 < c[0], g_1^e = c[0]0 \), \( S_2 = S_1 - 1 \). If \( b_0 > c[0], \) there exist \( e_1 > 0 \) such that \( S_2 = S_1 - g_1^e \), where \( g_1 = c[0] \neq 0 \). Similarly, we can pack \( S_2 \) using \( c[1] \) as the base element \( g_2 \) and derive \( S_3 = S_2 - c[1]e_2 \), etc. The value of residual part decreases at least one in each processing, and finally there is no residual part left. This demonstrates that \( b_0 = c[0]e_1 + c[1]e_2 + \cdots \). This expression is still valid even if the value of some elements is zero, \( c[i] = 0 \).

2) Next, for two digits \( S_1 = b_1b_0 \) case, there exist \( e_1 > 0 \) such that \( g_1^e = c[0]e_1 \) and \( S_2 = S_1 - g_1^e \), where the residual part \( S_2 \) is either one digit or is still two digits. If \( S_2 \) is one digit, the processing of packing \( S_2 \) is similar to the steps described in 1). When \( S_2 \) is
still two digits, there exist $c_2 > 0$ such that $S_3 = S_2 - c[1]c^2$. If $S_2$ is still two digits, it has $S_4 = S_3 + [2]c^3$, etc. Finally the residual part will be one digit and can be processed using steps described in 1). This proves $C = \{c[i]\}_{i=0}^{N-1}$ can serve as the base elements for arbitrary two digits $S_1 = b_1b_0$ case.

3) The packing of $S_1 = b_n b_{n-1} \cdots b_0$, $n > 2$, can be carried out using the similar procedures described in steps 1) and 2). However, when $n$ is not small, it may exist $c_1$ such that $(c[0]c[1] \cdots c[a])c^1 < b_n b_{n-1} \cdots b_0$ is true, then $g_1 = c[0]c[1] \cdots c[a]$ is assigned. The residual part is given by $S_2 = S_1 - g_1c^1$, which is then processed for further decomposition. It follows that $S_3 = S_2 - (c[a+1]c[a+2] \cdots c[a+k])c^2$, etc. This proves the secret sequence $C = \{c[i]\}_{i=0}^{N-1}$ can serve as the base elements of a spanning set.

4. Novel Data Encryption Examples

According to Corollary 1, a large nature number can be partitioned into some blocks which each block can be packed independently, thus we demonstrate in this section how to pack a seven digits nature number 1230186 chosen from the first seven digits of RSA-768. RSA-768 has 232 decimal digits (768 bits), and was factored on December 12, 2009 over the span of two years [7]. The RSA-768 can be factored into two factors with 116 digits, shown as follows

\[
\text{RSA} - 768 = 1230186684, 5301177551, 3049495838, 4962720772, 8535695953, 3479219732, 2452151726, 400502636, 5751874520, 2199786469, 3899564749, 4277406384, 5925192557, 3263034537, 3154826850, 7917026122, 1429134616, 7042921431, 1602201240, 4703747377, 9408066535, 1419597459, 8569021434, 13
\]

\[
= 3347807169, 8956898786, 0441698482, 1269081770, 4794983713, 768569124, 3138898288, 3793878002, 2876147116, 5253174308, 7737814467, 999489 \\
\times 3674604366, 6799590428, 2446337996, 2795263227, 9158164343, 8076426760, 3228381573, 9666511279, 2333734171, 4339681027, 0092798736, 308917.
\]

4.1. Example 1

We choose arbitrarily 62 digits from the 89th to the 150th of RSA-768 as the secret key sequence $C$, which is given by $C = 202199786469389595647494277406384592519255732630345373154826850$. The 40 base elements $\{g_i\}$ of the spanning set $\{g_i^{40}\}_{i=1}^{40}$ for representing 1230186 are chosen sequentially from $C$, which are given by

\[
\{g_i\}_{i=1}^{40} = \{20, 21, 9, 7, 86, 46, 9, 3, 8, 99, 56, 47, 49, 42, 77, 40, 63, 84, 59, 25, 19, 2, 5, 7, 3, 2, 6, 3, 0, 3, 4, 5, 3, 7, 3, 1, 5, 4\}. \tag{1}
\]

These data are the first 56 elements of $C$ which are underlined for recognition, and the associated 40 exponents of the spanning set are given by

\[
\{e_i\}_{i=1}^{40} = \{4, 4, 4, 3, 3, 3, 5, 5, 5, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2\}. \tag{2}
\]

Using equations (1) and (2) to express 1230186, it has

\[
1230186 = 20^4 + 21^4 + 9^4 + 9^4 + 7^3 + 86^3 + 46^3 + 9^5 + 3^5 + 8^5 + 99^2 + 56^2 + 47^2 + 49^2 + 42^2 + 77^2 + 40^2 + 63^2 + 84^2 + 59^2 + 25^2 + 19^2 + 2^2 + 5^2 + 5^2 + 7^2 + 3^2 + 2^2 + 3^2 + 0^2 + 3^2 + 4^2 + 5^2 + 3^2 + 7^2 + 3^2 + 1^1 + 1^1 + 1^1. \tag{3}
\]

The encrypted message (ciphertext) consists of $C_1$ and $C_2$ two sequences

\[
C_1 = (22111221112222222222111111111111111111111111111111111)
\]
and
\[ C_2 = (4443335552222222222222222222222222222111) \].

The first element 2 of \( C_1 \) indicates the first base element is two digits which is the first two elements \( c[0]c[1] = 20 \) of \( C \), and the second element 2 of \( C_1 \) is the consecutive two elements \( c[2]c[3] = 21 \) of \( C \), etc. \( C_2 \) is the values of the associated exponents, respectively. Given that the secret key \( C \) is available, the plaintext of nature number 1230186 of (3) can be derived using \( C_1 \) and \( C_2 \).

In equation (3) the spanning set consists of 40 exponential numbers, which is not an efficient encryption example. A more efficient encryption is present in following subsection.

4.2. Example 2

We can apply the following exponent set, which has 25 elements, to span the same nature number 1230186.
\[ \{e_i\}_{i=1}^{25} = \{2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 2, 2, 4, 2, 2, 3, 2, 2\}, \]

and the associated base set is as follows
\[ \{g_i\}_{i=1}^{25} = \{20, 21, 99, 786, 469, 3, 89, 95, 64, 74, 94, 27, 7, 40, 63, 8, 45, 92, 5, 19, 25, 5, 7, 3, 2\}. \]

We have
\[
1230186 = 20^2 + 21^2 + 99^2 + 786^2 + 469^2 + 3^2 + 89^2 + 95^2 + 64^2 + 74^2 + 94^2 + 27^3 + 7^3 + 40^3 + 63^3 + 8^2 + 45^2 + 92^2 + 5^3 + 19^2 + 25^2 + 5^3 + 7^2 + 3^2 + 2^2.
\] (4)

The encrypted message (ciphertext) consists of \( C_1 \) and \( C_2 \) two sequences
\[ C_1 = (22231222221221221221111), \]

and
\[ C_2 = (2222222222233332224223222). \]

In (4) only the first 44 elements of the secret key \( C \) are used to express 1230186; however, it requires 56 elements in (3). The problem of how to achieve pack efficiency should be the topic of our future study, and no further discussion of this topic is addressed here for brevity.

The security of the proposed scheme depends on the secret key sequence, which can be assigned in advance via the secure channel or can be shared using the Diffie-Hellman Key-Exchange system over public networks. In the later case it is based on the difficult of solving discrete logarithm problem.

4.3. Example 3

When there is no secret key sequence to govern how to choose the base elements of a spanning set, the expression of 1230186 can be more efficient, shown as follows
\[
1230186 = 1109^2 + 17^2 + 4^2 = 33^4 + 210^2 + 5^3 + 6^2 + 2^2.
\]

Practically, we can apply a spanning set with a fixed exponent, e.g., \( e_i = 2 \), for data encryption. Taking 1230186 as an example, which is expressed using three exponential numbers \( \{1109, 17, 4\} \), and it has six permutations. The ciphertext can be one of the six permutations of a base set of three elements \( \{1109, 17, 4\} \), which is
\[ C_1 = (1109174)_{(4,2,1)} = (1741109)_{(2,4,1)} = (1711094)_{(2,1,4)} = ... \]
The subscript $(4, 2, 1)$ of the first term indicates that seven ciphertext digits 1109174 should be partitioned into 4, 2, and 1 digits, which are associated with three base elements of the spanning set, respectively. For secure communications purpose only the base elements information, either $1109174$, $(1741109)$ or $(1711094)$, is transmitted, while both the exponent $e_i = 2$ and the partition code, $(4, 2, 1)$, $(2, 4, 1)$ or $(2, 1, 4)$, should be kept secret, which these data may be transmitted via a secure channel. We know that using $(1, 2, 4)$ to partitioning ciphertext $(1109174)$ obtains absolutely a different number

$$(1109174)_{(1,2,4)} = 1^2 + 10^2 + 9174^2 = 84162377 \neq 1230186.$$ 

Because the brute-force attack can grow exponentially with the increasing number of permutations, this implies that higher security level can be achieved if the number of base elements of a spanning set is large enough. Though a smaller number of elements of a spanning set can achieve higher level of packing efficiency, the level of security should be sacrificed.

5. Conclusions

This is the first study of using a set of exponential number for data encryption. The pattern of exponential number is defined, which the pattern of a large exponential number is unpredictable and the distribution of digits in one pattern can approach uniform distribution. We show that any positive integer can be expressed using a set of exponential numbers which is called the spanning set. In addition, there exist numerous spanning sets that can span a nature number. We can apply a spanning set for data encryption, where the secret key sequence is assigned to govern the selection of base elements for information concealment. The data encryption can be processed more efficiently when the same exponent is applied throughout the entire exponential numbers. The security level of this scheme depends on the cardinality of spanning set, thus there is tradeoff between packing efficiency and the level of security.

References

[1] M.S. Sharbaf, Quantum cryptography: An emerging technology in network security, *2011 IEEE International Conference on Technologies for Homeland Security (HST)*: pp. 13–19, 2011.

[2] R. L. Rivest, A. Shamir, and L. Adleman, A Method for Obtaining Digital Signatures and Public-Key Cryptosystems, *Communications of the ACM*, 21 (2): pp. 120–126, 1978.

[3] W. Diffie and M. E. Hellman, New directions in cryptography, *IEEE Trans. Inform. Theory*, vol. 22, pp. 644–654, Nov. 1976.

[4] National Institute for Standards and Technology, Digital signature standard (DSS), *Fed. Reg.*, vol. 56, p. 169, Aug. 1991.

[5] N. Koblitz, Elliptic curve cryptosystems, *Math. Comput.*, vol. 48, no. 177, pp. 203–209, Jan. 1987.

[6] A. J. Menezes, *Elliptic Curve Public Key Cryptosystems*. Norwell, MA: Kluwer, 1993.

[7] Thorsten Kleinjung, Kazumaro Aoki, Jens Franke, Arjen K. Lenstra, Emmanuel Thom, Pierrick Gaudry, Alexander Kruppa, Peter Montgomery, Joppe W. Bos, Dag Arne Osvik, Herman te Riele, Andrey Timofeev, and Paul Zimmermann, *Cryptology ePrint Archive: Report 2010/006*. 