Novel Encryption Schemes Based on Catalan Numbers

[1]D. Sravana Kumar [2]CH. Suneetha [3]A. Chandrasekhar
[1]Reader in Physics, SVLNS Government College, Bheemunipatnam, Visakhapatnam Dt., India
[2]Assistant Professor in Engineering Mathematics, GITAM University , Visakhapatnam, India,
[3]Professor in Engineering Mathematics, GITAM University, Visakhapatnam, India

Abstract:- Applied Number theory has so many applications in cryptography. Particularly integer sequences play very important in cryptography. Earlier cryptographic algorithms were designed using the integer sequences of Fibonacci numbers and Lucas numbers. In the present paper novel cryptographic algorithms were proposed basing on integer sequences of Catalan numbers.

Key Words:- Catalan number, encryption, decryption

Introduction:-

Binomial Coefficients:- Binomials which are sums of two terms always occur in mathematics. There is symmetric way of expanding the positive integral powers of binomials. Let n and r be nonnegative integers. The binomial coefficient $\begin{pmatrix} n \\ r \end{pmatrix}$ is defined as

$$\begin{pmatrix} n \\ r \end{pmatrix} = \frac{n!}{r!(n-r)!} \text{ where } 0 \le r \le n, \text{ if r} > n \text{ then } \begin{pmatrix} n \\ r \end{pmatrix} \text{ is defined as zero.}$$

Integer Sequence: - An integer sequence is a sequence of integers which may be specified explicitly by giving a formula for its nth term, or implicitly by giving a relationship between its terms. For example, the sequence 0, 1, 1, 2, 3, 5, 8, 13, … (the Fibonacci sequence) is formed starting with 0 and 1 and then adding any two consecutive terms to obtain the next one: an implicit description. The sequence 0, 3, 8, 15, … is formed according to the formula $n^2 - 1$ for the *n*th term: an explicit definition.

Definition:- Catalan numbers [1] were discovered by Belgian mathematician Eugene C. Catalan in 1838. Catalan numbers are sequence of natural numbers. The Catalan number C_n is defined as

$$C_n = \frac{2n!}{(n+1)!n!} = \frac{1}{n+1}\begin{pmatrix} 2n \\ n \end{pmatrix}, n \ge 0$$

Every Catalan number is an integer. The various Catalan numbers are 1,1,2,5,14,42,132,429,1430,4862,…………………..

Definition: - The second way of defining the Catalan number is

$$C_n = \begin{pmatrix} 2n \\ n \end{pmatrix} - \begin{pmatrix} 2n \\ n-1 \end{pmatrix} = \frac{2n!}{(n!)^2} - \frac{2n!}{(n-1)!(n+1)!} = \frac{2n!}{n!(n+1)!}$$

This definition confirms that every Catalan number is an integer.

Segner's Recursive Formula:- In 1761, Segner using the addition and multiplication principles , published a second order recurrence formula T_n, where n if greater than or equal to 3. The Segner's recurrence formula can be developed using an n-gon with vertices $A_1, A_2 \ldots\ldots A_n$. The Segner's recurrence formula T_n is

$$T_n = \sum_{k=2}^{n-1} T_k T_{n-k+1} = T_2 T_{n-1} + T_3 T_{n-2} + \ldots\ldots + T_{n-1} T_2, n \ge 3$$

Definition:- From Euler's triangulation problem the Catalan number can also be defined [3] recursively as

$$C_0 = 1,\ C_1 = 1,\ C_n = \frac{4n-2}{n+1} C_{n-1}, n \ge 0 \ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(1)$$

In 1941 H. Urban conjectured the recurrence relation. He first computed C_3, C_4 , C_5 and then he made the following observations

$$\frac{C_1}{C_0} = \frac{1}{1} = \frac{2}{2} = \frac{4.1-2}{1+1} \qquad \frac{C_2}{C_1} = \frac{2}{1} = \frac{6}{3} = \frac{4.2-2}{2+1}$$

$$\frac{C_3}{C_2}=\frac{5}{2}=\frac{10}{4}=\frac{4.3-2}{3+1} \qquad \frac{C_4}{C_3}=\frac{14}{5}=\frac{14}{5}=\frac{4.4-2}{4+1}$$

$$\frac{C_5}{C_4}=\frac{42}{14}=\frac{18}{6}=\frac{4.5-2}{5+1}$$

With this clear pattern Urban inferred that $C_n = \dfrac{4n-2}{n+1}C_{n-1}, n \geq 1$..........(1)

Explicit formula of C_n:- From the recursive formula the explicit formula of Catalan number C_n can be derived as

$$C_n = \frac{4n-2}{n+1}C_{n-1}$$

$$= \frac{(4n-2)(4n-6)}{(n+1)n}C_{n-2}$$

$$= \frac{(4n-2)(4n-6)(4n-10)}{(n+1)n(n-1)}C_{n-3}$$

$$\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots$$

$$= \frac{1}{n+1}\binom{2n}{n}$$

An approximate value of Catalan number C_n can be found using Stirling's approximation for factorials as

$$C_n \approx \frac{2^{2n}}{(n+1)\sqrt{n\pi}}$$

Ubiquitous Nature of Catalan Numbers :-

Like Fibonacci numbers and Lucas numbers, Catalan numbers are also ubiquitous [7]. Catalan sequence is the most frequently encountered sequence presents everywhere in combinatorics. There are many counting problems in combinatorics whose solution is given by Catalan numbers. Some examples are

1.Finding the number of lattice paths of mountain ranges with steps (1,1) and (1,-1) that never fall below the x-axis which are called Dyck paths.

2. Successive applications of binary operations can be represented in terms of full binary tree. There the Catalan number C_n is the number of full binary trees with n+1 leaves.

3. In parenthesization problem using postfix expressions and cyclic shifts of such exprssessions.

4. In the theory of partitioning in Combinatoraics.

5. In Abstract algebra and sports

6. In Pascal's triangles and inComputer Science.

Properties of Catalan numbers:-

Parity of Catalan numbers:-Odd and Even Catalan numbers [6]

Theorem:- For n > 0, C_n is odd if and only if n is a Mersenner number.

Proof:- It follows from the Segner's recurrence relation that

$$C_n = \begin{cases} 2(C_0C_{n-1} + C_1C_{n-2} + \dots\dots\dots\dots + C_{\frac{n}{2}-1}C_{\frac{n}{2}}, & \text{if n is even} \\ 2(C_nC_n + C_nC_n + \dots\dots\dots\dots + C_{\frac{n-3}{2}}C_{\frac{n+1}{2}}, & \text{otherwise} \end{cases}$$

Consequently for n > 0, C_n is odd if and only if both n and $C_{\frac{n-1}{2}}$ are odd. From this it implies that the Catalan number C_n is odd if and only if $\dfrac{n-1}{2}$ and $C_{\frac{n-3}{4}}$ are both odd or $\dfrac{n-1}{2}$ is zero.

Primality of Catalan numbers:-

Theorem:- The only prime Catalan numbers[2] are C_2 and C_3.

Proof :- From the formula (1) $(n+2)C_{n+1} = (4n+2)C_n$. Let C_n be prime for some n. It follows that if n > 3, $\frac{n+2}{C_n}$ < 1. So, $C_n > n+2$. Consequently $\frac{C_n}{C_{n+1}}$, so $C_{n+1} = kC_n$ for some positive integer k. Then (4n+2)=k(n+2) where $1 \le k \le 3$ and thus $n \le 4$.

Therefore, it follows that C_2 and C_3 are the only Catalan numbers that are prime.

L.W. Shapiro of Howard University introduced triangular array of numbers B(n,r) Called Catalan triangle.

$$B(n,r) = \begin{cases} 1, \text{ if } r = 1 = n \\ B(n-1,r-1) + 2B(n-1,r) + B(n-1,r+1), \text{ if } 1 \le r \le n \\ 0, \text{ otherwise} \end{cases}$$

Generalization of Catalan numbers:- Catalan numbers are special class of numbers

$$C(n,k) \text{ defined by } C(n,k) = \frac{1}{kn+1}\binom{(k+1)n}{n} \text{ where } k \ge 0$$

Clearly C(n,1) = C_n.

Some Catalan Numbers for n=1 to 20are given below in the table for ready reference. As the n value increases the number of digits in C_n increase rapidly. For example number of decimal digits in C_{100}=57.

n	C_n	n	C_n
1	1	11	58786
2	2	12	208012
3	5	13	742900
4	14	14	2674440
5	42	15	9694845
6	132	16	35357670
7	429	17	129644790
8	1430	18	477638700
9	4862	19	1767263190
10	16796	20	6564120420

Application of Number Theory in Cryptography:- Applied Number theory has so many applications in cryptography [8,9]. Cryptographic algorithms were designed using Fibonacci numbers and Lucas numbers. The general idea of the Fibonacci cryptography is similar to the Fibonacci coding and based on the application of the generalized Fibonacci matrices. Lucas sequences can also be used for public key cryptosystems and signature systems similar to RSA, but using Lucas sequences modulo a composite number instead of exponentiation.

Proposed Method 1:- Both the communicating parties before communicating the messages agree upon to use a large random number n. Then they write the corresponding Catalan number C_n which is a very large natural number.

Step 1. The sender divides the text message into data blocks of m characters each, where m is a natural number equal to the number of decimal digits in the Catalan number C_n . The size of message always may not be of m characters or multiples of m characters. If the message contains less than m characters then the sender adds three hash (###) characters at the end of the message representing that the message space is over and the remaining characters may be filled at random.

Step 2:- Consider the first data block. All the m characters of the first data block are coded to equivalent binary numbers using ASCII code table. Let the ASCII binary code of the j^{th} character M_j be

b_7	b_6	b_5	b_4	b_3	b_2	b_1	b_0

Step 3:- Let the jth digit of the key i.e., the Catalan number C_n be K_j. Then compute $K_j^l = K_j(\mathrm{mod}\,8)$ and $K_j^{ll} = K_j + 4(\mathrm{mod}\,8)$. Logical NOT Operations are applied on $b_{k_j^l}$ and $b_{k_j^{ll}}$ bits of the number.

Example: Let the M_j has the binary code 1001 0110 and let K_j be 9 then compute $K_j^l = K_j(\mathrm{mod}\,8) = 9(\mathrm{mod}\,8) = 1$ and $K_j^{ll} = K_j + 4(\mathrm{mod}\,8) = 13(\mathrm{mod}\,8) = 5$. Then the logical NOT operation is applied on b_1 and b_5. Then the number M_j is encrypted as 10100100.This encrypted number is coded back to text characters using ASCII code table.

Step 4:- The procedure explained in steps 2 and 3 is applied to all the 8 bit binary codes of the characters in the nth data block. This procedure is applied to all data blocks.

Step 5 :- Then this cipher text is communicated to the receiver in public channel.

Decryption:- The receiver after receiving the cipher text decrypts the cipher text as follows:

Step 1:- The receiver divides the cipher text into data blocks of m characters each.

Step 2:- He decrypts the message by encrypting the cipher text as explained in the encryption algorithm. If the receiver finds three consecutive hash (###) characters in the cipher text then he recognizes that the message is over and then discards the remaining cipher characters.

Example:-

Encryption:-Suppose that the communicating parties agree upon to use the random natural number n=7 in the process of communication. The corresponding Catalan number C_7 = 429. The message is divided into blocks of 3 characters each. Let the message be 'rrr'. This message constitutes only one text block i.e. 'rrr', as m=3 in the selected example.

$$K_1^l = 4(\mathrm{mod}\,8) = 4, K_1^{11} = 4 + 4(\mathrm{mod}8) = 0$$

$$K_2^l = 2(\mathrm{mod}\,8) = 2, K_2^{11} = 2 + 4(\mathrm{mod}8) = 6$$

$$K_3^l = 9(\mathrm{mod}\,8) = 1, K_1^{11} = 9 + 4(\mathrm{mod}8) = 5$$

The binary equivalent of 'r' from the ASCII code table is 01110010. The first character 'r' is coded as 01100011 which corresponds to the character 'c' in the ASCII code table. The second character' r' is coded to 00110110 which corresponds to the character '6' in the ASCII code table. The third character ' r' is coded to 01010000 which corresponds to 'P' in the ASCII code table. So, the same character 'r' in the message is coded to different characters in this proposed method. 'rrr' is encrypted as 'c6P'. Here a simple case of a small random natural number n = 7 is considered as the example. But, in practice very large random natural number n should be selected so that C_n contains large number of decimal digits. Such selection ensures m is large. For example if n = 100 , m =57

Decryption:- The receiver receives the cipher text 'c6P'. He computes C_7 and calculates

$$K_1^l = 4(\mathrm{mod}\,8) = 4, K_1^{11} = 4 + 4(\mathrm{mod}8) = 0$$

$$K_2^l = 2(\mathrm{mod}\,8) = 2, K_2^{11} = 2 + 4(\mathrm{mod}8) = 6$$

$K_3^l = 9(\mathrm{mod}\,8) = 1, K_1^{11} = 9 + 4(\mathrm{mod}8) = 5$ He writes the corresponding binary code for 'c6P'. The corresponding binary code of the character 'c' is 01100011. To decrypt this binary number he encrypts it using the encryption algorithm. This results in decrypted code 01110010 corresponding to the message character' r'. Similarly the other characters in cipher '6' and 'P' are decrypted to 'r' and 'r'. So, the receiver recovers the original message 'rrr'.

Proposed Method II: - Both the communicating parties before communicating the messages agree upon to use a large random number n. Then they write the corresponding Catalan number C_n which is a very large natural number.

Step 1. The sender divides the text message into data blocks of m characters each, where m is a natural number equal to the number of decimal digits in the Catalan number C_n . The size of message space may not be of m characters or multiples of m characters. If the message contains less than m characters then the sender adds three hash (###) characters at the end of the message representing that the message space is over and the remaining characters may be filled at random.

Step 2:- Consider the first data block. All the m characters of the first data block are coded to equivalent binary numbers using ASCII code table. Let the ASCII binary code of the jth character M_j be

b_7	b_6	b_5	b_4	b_3	b_2	b_1	b_0

Step 3:- Let the j^{th} digit of the key i.e., the Catalan number C_n be K_j. Then K_j which is a decimal digit is converted to 4bit binary number. Then the jth digit Mj which is in the 8 bits binary format is XORed with 8421 code of K_jK_j. For example, let the binary code of 4th digit of the message be 10100100 and the 4th digit of the key be 5. Then 10100100 is XORed bit wise with the 8421 code of 55.

(10100100) XOR (01010101) = 00001110.

Step 4:- The procedure explained in steps 2 and 3 is applied to all the characters of each data block which are in binary format.

Step 5:- Then this cipher text is communicated to the receiver in public channel.

Decryption: - The receiver after receiving the cipher text decrypts the cipher text as follows:

Step 1:- The receiver divides the cipher text into data blocks of m characters each.

Step 2:- He decrypts the message by encrypting the cipher text as explained in the encryption algorithm. If the receiver finds three consecutive hash (###) characters in the decrypted text then he recognizes that the message is over and then discards the remaining cipher characters.

Example:-

Encryption:-Suppose that the communicating parties agree upon to use the random natural number n=7 in the process of communication. The corresponding Catalan number $C_7 = 429$. Here $K_1=4$. The corresponding 8421 code is 0100. The 8421 code of 44 = 01000100. $K_2=2$. The corresponding 8421 code is 0010. The 8421 code of 22 = 00100010. $K_3=9$. The corresponding 8421 code is 1001. The 8421 code of 99 =10011001. Let the message be 'rrr'. This message constitutes only one text block i.e. 'rrr', as m=3 in the selected example. The binary equivalent of 'r' from the ASCII code table is 01110010. The first character 'r' is encrypted as 01110010 XOR 01000100 = 00110110. This binary code corresponds to the cipher character '6' in the ASCII code table. The second character 'r' is encrypted as 01110010 XOR 00100010 = 01010000. This binary code corresponds to the cipher character 'P' in the ASCII code table. The third character 'r' is encrypted as 01110010 XOR 10011001 = 11101011. This binary code corresponds to the cipher character 'ë' in the ASCII code table. So, the same character is coded to different characters in this proposed method. 'rrr' is encrypted as '6P ë'. Here a simple case of a small random natural number n = 7 is considered as the example. But, in practice very large random natural number n should be selected so that C_n contains large number of decimal digits. Such selection ensures m is large. For example if n = 94 , m =54

Decryption:- The receiver receives the cipher text '6P ë' . The ASCII binary codes of the characters of cipher message are 00110110, 01010000 and 11101011.He computes C_7 =429 and recognizes K_1=4,K_2=2 and K_3=9. The corresponding 8421 codes are 0100, 0010 and 1001 respectively. He calculates the strings required to decrypt the first, second and third characters in the cipher message as 01000100, 00100010 and 10011001. He decrypts the cipher as follows.

1)00110110 XOR 01000100 = 01110010. This binary code corresponds to the cipher character 'r' in the ASCII code table.

2) 01010000 XOR 00100010 = 01110010. This binary code corresponds to the cipher character 'r' in the ASCII code table.

3)11101011 XOR 10011001 = 01110010. This binary code corresponds to the cipher character 'r' in the ASCII code table.

So, the receiver retrieves the original message 'rrr' from the cipher text.

Conclusions:- In the proposed method of encryption using Catalan numbers a large random natural number 'n' is agreed upon by the communicating parties as the secret key. The decimal digits in the Catalan number C_n corresponding to the agreed upon secret key 'n' are used to encrypt/decrypt the message. The proposed two methods of encryption use logical NOT and logical XOR operations on the bits of the ASCII binary codes of the characters in the message. The message is divided into data blocks of size 'm' characters each, where 'm' is the number of decimal digits in the Catalan number C_n. In the algorithms proposed decryption is effected by encrypting the cipher text, i.e. encryption of plain text gives cipher text and encryption of cipher text gives the plain text. The encryption algorithms proposed using Catalan numbers resistant to most of the types of known attacks such as plain text attacks and chosen cipher text attacks [4]. The identical characters of the plain text are coded to different cipher characters. Hence, the brute force attack [5] and exhaustive key search are difficult to execute. The time for enciphering or deciphering is independent of characters in the data block. The time required to encipher or decipher a data block is same for all data blocks. Even though the original message contains the characters whose number is less than the number of digits of the Catalan number the remaining characters are filled at random, so that each data block contains exactly the same number of characters. Therefore the cipher proposed here is less prone to timing attacks. The main advantage of the methods proposed in this paper are higher level of security at relatively low computational overhead.

References:-

[1] Alter.R, "Some Remarks and Results on Catalan Numbers", proceedings of Louisiana Conference on Combinatorics, Graph Theory and computing (1971).

[2] Alter.R, and K.K. Kubota "Prime power Divisibility of Catalan Numbers", Journal of Combinatorial Theory, series A, 15 (1973), 243-256.

[3]. Brualdi, R. A. Introductory Combinatorics, 4th ed. New York: Elsevier, 1997

[4]. Canetti R. Halevi, S. and Katz, J. "Chosen cipher text security from identity-based encryption", Advances in Cryptography-EUROCRYPT 2004, Vol. 3027 of LNCS, Springer-Verlag.

[5]. Eli Biham, "Cryptanalysis of multiples modes of operation", Journal of Cryptology, 1998, Vol.11, No.1, pgs 45-58

[6] Jarvis.F. "Catalan Numbers", Mathematical spectrum 36 (2003-04), 9-12.

[7] Koshy.T, "The Ubiquitous Catalan Numbers", Mathematics Teacher 100 (October 2006).

[8] D.R. Stinson: Cryptography. Theory and Practice, CRC Press, Boca Raton, 2002.

[9] W. Trappe, L.C. Washington: "Introduction to Cryptography with Coding Theory" , Prentice Hall, Upper Sadle River, 2002