Research and Reviews: Discreate Mathematical Structures

Vol: 11, Issue: 3, Year: 2024, ISSN(e): 2394 –1979

Algorithmic Encryption of Natural Numbers Based on the Collatz Conjecture

1, Bahar Kuloğlu

1 Assistant Professor, Department of Engineering Basic Sciences, Sivas University of Science and Technology, Sivas, Türkiy

Email: bkuloglu@sivas.edu.tr

Research Paper

Received Date: 19 December 2024

Accepted Date:6 January 2025

Published Date:10 January 2025

Abstract

This paper creates an encryption technique based on the well-known unresolved mathematical problem known as the Collatz conjecture. This conjecture states that every natural number larger than one eventually lowers to one by a set of precise procedures called the Collatz sequence. The study's methodology associates these sequences' terms with the Turkish alphabet's letters. The Collatz transformation is applied to the integer that corresponds to each letter. The words created from these sequences are subjected to addition and subtraction operations during the encryption process. It is challenging to reverse-engineer the original text because of the intricate pattern created by these procedures. This approach uses the Collatz conjecture's intrinsic complexity and unpredictability to produce a strong encryption scheme. It is intended that this encryption technique will provide a degree of security based on the conjecture's unproven status, potentially rendering it immune to contemporary cryptographic attacks. In order to advance the science of cryptography, the study aims to investigate the viability of using this unsolved issue as a basis for developing extremely safe encryption algorithms.

Keywords: Collatz Conjecture, Encoding, Decoding, Information Security, Cryptology.

* *Author for Correspondence* E-mail: **bkuloglu@sivas.edu.tr**

1. INTRODUCTION

Cryptography is a method of securing information using mathematical techniques. Cryptology, on the other hand, is the science that deals with the concealment and decryption of information. Protecting and securing information has always been of great importance, both in the past and today. For this reason, the history of cryptography dates back to ancient times. Around 1900 BCE, hieroglyphic symbols in Egypt were considered the first examples of secret writing. Between 60–50 BCE, Julius Caesar used a cryptographic method, now known as Caesar cipher, to secure state communications. This method involved replacing each letter in the text with the letter three positions later in the alphabet. The Vigenère cipher (1586 CE) improved upon Caesar's method by creating a type of polyalphabetic substitution table. The Hill cipher, on the other hand, systematically increased the reliability of encryption through the use of block methods.

In cryptography, encryption algorithms are used to encrypt or decrypt pieces of information. Keys, consisting of a series of bits, are utilized to perform this process.

With advances in technology, it has become evident that previously used methods can now be easily broken, even by amateurs. Therefore, stronger encryption and decryption techniques are required to secure sensitive information, such as military records, medical records, financial data, and protection against cyberattacks [1].

Our literature review revealed that Fibonacci and Lucas number sequences are widely studied, with many encryption and decryption algorithms based on these sequences. Furthermore, in later years, more generalized forms of these sequences have been used to develop more diverse cryptographic algorithms [2,3].

In the following years, coding studies continued with different number sequences

Inspired by these developments and considering previously studied encryption systems, we propose a method where the natural numbers greater than 0 corresponding to the Turkish alphabet are mapped to the terms of sequences derived from the Collatz conjecture. The sequence length is then used to encrypt the letters of the message. In the encryption and decryption steps, the terms generated by the Collatz conjecture for the positions 1,2,3,4, ... (corresponding to the letters in the word or sentence) are first added and then subtracted to complete the encryption and decryption processes [4].

The primary reason for using the Collatz conjecture in our encryption method is its unique property: it posits that all natural numbers greater than 1 eventually reduce to 1. While this conjecture has been proven for numbers up to $2^{68} \approx 2.951 \times 10$, it remains unproven for numbers larger than this. These larger numbers are still the subject of ongoing mathematical research, making the conjecture a promising foundation for a secure encryption system [5,6].

2. MAIN RESULTS

The Collatz conjecture is based on reducing all numbers to one [7]. This conjecture operates in two stages, depending on whether a number is even or odd. If the number is even, it is divided

by two, while if it is odd, the algorithm multiplies it by three and adds e. The resulting value is then re-evaluated under the same rules, continuing the process. If the result is odd, the same operation is repeated, and if it is even, it is divided by two until it is reduced to one.

In our study, all letters in the Turkish alphabet are initially assigned numerical values in sequence. If any symbols or special characters are used, numbering continues sequentially from 29 onward. Lastly, since spaces between words also have a binary representation in computing systems, a numerical value is assigned to the space after the symbols. This ensures that the numerical terms corresponding to the letters are finalized (see Table 1).

Next, encoded numerical terms corresponding to the letters are generated. This step is critical. At this stage, the Collatz conjecture's algorithm is applied to the numerical value of each letter, producing the sequence length, denoted as L (see Table 2) [8,9]. A key complexity in this process arises from letters sharing the same L value. To address this, the smallest unused L value is assigned to the first letter with a duplicate L, and the same approach is applied sequentially to all letters sharing the same L value. This method ensures that the encoding becomes more intricate [10].

Below, the sequence lengths (L values) corresponding to all letters in our alphabet, calculated using the Collatz conjecture algorithm, are provided:

$A = \{1\}, L = 1$

$B = \{2,1\}, L = 2$

$C = \{3,10,5,16,8,4,2,1\}, L = 8$

Ç $= \{4,2,1\}, L = 3$

$D = \{5,16,8,4,2,1\}, L = 6$

$E = \{6,3,10,5,16,8,4,2,1\}, L = 9$

$F = \{7,22,11,34,17,52,26,13,40,20,10,5,16,8,4,2,1\}, L = 17$

$G = \{8,4,2,1\}, L = 4$

Ğ $= \{9,28,14,7,22,11,34,17,52,26,13,40,20,10,5,16,8,4,2,1\}, L = 20$

$H = \{10,5,16,8,4,2,1\}, L = 7$

$I = \{11,34,17,52,26,13,40,20,10,5,16,8,4,2,1\}, L = 15$

İ $= \{12,6,3,10,5,16,8,4,2,1\}, L = 10$

$J = \{13,40,20,10,5,16,8,4,2,1\}, L = 10$

$K = \{14,7,22,11,34,17,52,26,13,40,20,10,5,16,8,4,2,1\}, L = 18$

$L = \{15,46,23,70,35,106,53,160,80,40,20,10,5,16,8,4,2,1\}, L = 18$

$M = \{16,8,4,2,1\}, L = 5$

$N = \{17,52,26,13,40,20,10,5,16,8,4,2,1\}, L = 13$

$O = \{18,9,28,14,7,22,11,34,17,52,26,13,40,20,10,5,16,8,4,2,1\}, L = 21$

Ö $= \{19,58,29,88,44,22,11,34,17,52,26,13,40,20,10,5,16,8,4,2,1\}, L = 21$

$P = \{20,10,5,16,8,4,2,1\}, L = 8$

$R = \{21,64,32,16,8,4,2,1\}, L = 8$

$S = \{22,11,34,17,52,26,13,40,20,10,5,16,8,4,2,1\}, L = 16$

$Ş = \{23,70,35,106,53,160,80,40,20,10,5,16,8,4,2,1\}, L = 16$

$T = \{24,12,6,3,10,5,16,8,4,2,1\}, L = 11$

$U = \{25,76,38,19,58,29,88,44,22,11,34,17,52,26,13,40,20,10,5,16,8,4,2,1\}, L = 24$

$Ü = \{26,13,40,20,10,5,16,8,4,2,1\}, L = 11$

$$V = \{27,82,41,124,62,31,94,47,142,71,214,107,322,161,484,242,121,364,182,91,274,137,412,206,103,310,155,466,233,700,350,175,526,263,790,395,1186,893,1780,890,445,1336,668,334,167,502,251,754,377,1132,566,283,850,425,1276,638,319,958,479,1435,719,2158,1079,3238,1619,4858,2429,7288,3644,1882,911,2734,1367,4102,2051,6154,3077,9222,4616,2308,1154,577,1732,866,433,1300,650,325,976,488,244,121,61,184,92,46,23,70,35,106,53,160,80,40,20,10,5,16,8,4,2,1\}, L = 112$$

$Y = \{28,14,7,22,11,34,17,52,26,13,40,20,10,5,16,8,4,2,1\}, L = 19$

$Z = \{29,88,44,22,11,34,17,52,26,13,40,20,10,5,16,8,4,2,1\}, L = 19$

$Space = \{30,15,46,23,70,35,106,53,160,80,40,20,10,5,16,8,4,2,1\}, L = 19$

The table below is obtained by numbering the letters in alphabetical order:

A	**B**	**C**	**Ç**	**D**	**E**	**F**	**G**	**Ğ**	**H**
1	2	3	4	5	6	7	8	9	10
I	İ	J	K	L	M	N	O	Ö	P
11	12	13	14	15	16	17	18	19	20
R	S	Ş	T	U	Ü	V	Y	Z	Space
21	22	23	24	25	26	27	28	29	30

Table 1. Numerical terms corresponding to letters

The table given below was obtained by writing the string length values created during the reduction of the letters in our alphabet to 1 with the Collatz assumption and giving the first smallest value of the unused L value to the first letter with the same *L* value and applying the same method to all letters with the same *L* value in order.

A	**B**	**C**	**Ç**	**D**	**E**	**F**	**G**	**Ğ**	**H**
01	02	08	03	06	09	17	04	20	07
I	İ	J	K	L	M	N	O	Ö	P
15	10	12	18	14	05	13	21	23	25
R	S	Ş	T	U	Ü	V	Y	Z	Space

26	16	27	11	24	28	22	19	29	30

Table 2. Coded numerical terms corresponding to letters

Let's consider the following example for a more systematic understanding of the steps and processes mentioned above.

Example: Let's consider the sentence: HİÇ HATA YAPMAMIŞ BİR İNSAN ASLA YENİ BİR ŞEY DENEMEMİŞTİR

The coded numbers corresponding to each letter are obtained using Table2 as follows and Table3 is created.

H	**İ**	**Ç**	**Space**	**H**	**A**	**T**	**A**	**Space**	**Y**
07	10	03	30	07	01	11	01	30	19
A	P	M	A	M	I	Ş	Space	B	İ
01	25	05	01	05	15	27	30	02	10
R	Space	İ	N	S	A	N	Space	A	S
26	30	10	13	16	01	13	30	01	16
L	A	Space	Y	E	N	İ	Space	B	İ
14	01	30	19	09	13	10	30	02	10
R	Space	Ş	E	Y	Space	D	E	N	E
26	30	27	09	19	30	06	09	13	09
M	E	M	İ	Ş	T	İ	R		
05	09	05	10	27	11	10	26		

Table 3. Numerical terms corresponding to the coded sentence

If we sum the encoded numerical values corresponding to each letter in the obtained sentence sequentially from left to right, as shown in Table 2, the following table is generated. During the summation process, once the final value in Table 2 is reached, the process loops back to the beginning. This ensures that the same letter corresponds to a different numerical value each time, thereby increasing the complexity of decryption. This feature sets our study apart from similar approaches. In other words, this process is designed to make decrypting the transmitted message as challenging as possible.

H	**İ**	**Ç**	Space	**H**	**A**	**T**	**A**	Space	**Y**
08	12	11	33	13	10	28	05	50	26
A	P	M	A	M	I	Ş	Space	B	İ
16	35	17	19	19	20	40	51	25	35
R	Space	İ	N	S	A	N	Space	A	S

52	46	37	24	40	29	35	49	30	46
L	A	Space	Y	E	N	İ	Space	B	İ
15	03	38	22	15	22	27	34	22	17
R	Space	Ş	E	Y	Space	D	E	N	E
41	40	39	27	33	35	18	30	36	34
M	E	M	İ	Ş	T	İ	R		
31	25	32	21	51	39	32	45		

Table 4. Encryption of the coded sentence

It can be observed that since the same letters will correspond to different numerical values each time, the decryption of the message becomes significantly more difficult compared to similar encryption and decryption methods. This makes it immediately clear that the encryption is much harder to break.

The message transmitted to the recipient will be:

08121133131028055026163517191920405125355246372440293549304615033822152227342217414039273335183036343125322151393245

Thus, the encryption algorithm is concluded, and the decryption algorithm is initiated.

The decryption process, which involves obtaining the original sentence, is carried out by reversing the previous operations. Specifically, the values given in Table 2 are subtracted from each corresponding letter in the sequence, going from left to right. Using these operations, Table 5 is generated.

H	**İ**	**Ç**	Space	**H**	**A**	**T**	**A**	Space	**Y**
08-01	12-02	11-08	33-03	13-06	10-09	28-17	05-04	50-20	26-07
A	P	M	A	M	I	Ş	Space	B	İ
16-15	35-10	17-12	19-18	19-14	20-05	40-13	51-21	25-23	35-25
R	Space	İ	N	S	A	N	Space	A	S
52-26	46-16	37-27	24-11	40-24	29-28	35-22	49-19	30-29	46-30
L	A	Space	Y	E	N	İ	Space	B	İ
15-01	03-02	38-08	22-03	15-06	22-09	27-17	34-04	22-20	17-07
R	Space	Ş	E	Y	Space	D	E	N	E
41-15	40-10	39-12	27-18	33-14	35-05	18-13	30-21	36-23	34-25
M	E	M	İ	Ş	T	İ	R		
31-26	25-16	32-27	21-11	51-24	39-28	32-22	45-19		

Table 5. Deciphering the coded sentence

As seen from Table 5, by initially obtaining the encoded values for each letter, we successfully reach the sentence of the message we sent. Therefore, we conclude the process by proving that the encryption and decryption algorithms work correctly.

From the operations performed, it is evident that due to the performance and flexibility of our algorithm, it can appeal to a wide user base and offer a solution that can be used across different platforms. This demonstrates that our algorithm has the potential to be applied in various fields in future studies.

3. CONCLUSIONS

The primary aim of this study is to verify whether the transmitted message is correctly delivered through encoding and decoding steps, using addition and subtraction operations based on the Collatz conjecture. By doing so, we have developed an encryption-decryption algorithm that, whether implemented without computers or through the specified steps in a program, is more reliable and applicable than previously established coding algorithms.

The performance of our algorithm allows for its use in encrypting database files and in dynamic password generation programs. This enables the secure storage of data and the creation of strong passwords in future applications.

Additionally, due to the flexibility of our algorithm within a defined text pool, we can appeal to a broad user base. This means that users can view our algorithm as a solution for encrypting applications, photos, audio, and textual content on various platforms, including computers, tablets, and smartphones.

In this way, users can securely protect their data and encrypt the content they wish to secure across different platforms. Thanks to the algorithm's flexibility, it can be used to encrypt various types of data, offering solutions tailored to users' needs.

Suggestions:

This study demonstrates that the encryption method based on the defined L value can be extended to different encryption approaches. For example, by considering the largest and smallest numbers in the cycle of each letter under the Collatz conjecture, or by summing these extreme values and applying the Collatz operations again, new encryption and decryption algorithms could be developed. Instead of relying on a single number sequence, using more complex sequences enhances the system's resistance to various types of attacks. At this stage, exploring different conjectures and even developing algorithms that use multiple conjectures simultaneously would significantly increase the reliability and robustness of the encryption system.

4. ACKNOWLEDGEMENT

We would like to thank the entire jury team of the TÜBİTAK Science Human Support Program, who awarded our project the second-place prize in the regional final of the research projects competition for middle school students held in 2024, for their constructive feedback and support in developing our work.

5. REFERENCES

1. Basu, M., Prasad, B. (2009). The generalized relations among the code elements for Fibonacci coding theory. Chaos Solitons Fractals, 5 2517-2525.
2. Dişkaya, O., Avaroğlu, E., Menken, H., and Emsal, A. (2022). A new Encryption Algorithm Based on Fibonacci Polynomials and Matrices. International Information and Engineering Technology Association, 39 (5), 1453-1462
3. Uçar, S., Taş, N., and Özgür, N. H. (2019). A New Application to Coding Theory via Fibonacci and Lucas Numbers. Mathematical Sciences and Application E-Notes, 7 (1), 62-70.
4. Eser. E., Kuloğlu, B., Özkan, E. (2024). An Encoding-Decoding Algorithm Based on Fermat and Mersenne Numbers, Applied Mathematics E-Notes, 24, 274-282
5. Kuloğlu B., Özkan E. (2023). Creating Encryption and Decryption Algorithm Using Pell Numbers. EYFOR-14, Antalya, Türkiye, pp. 505.
6. Kuloglu, B., Özkan, E. (2023). Applications of Jacobsthal and Jacobsthal-Lucas Numbers in Coding Theory, ¨ Mathematica Montisnigri, Lvii, 54–64.
7. Van Bendegem JP. The Collatz conjecture. A case study in mathematical problem solving. Logic and Logical Philosophy. 2005 Jun 23;14(1):7-23.
8. Kuloğlu, B. (2024). Creating a New Coding and Decoding Algorithm Based on Violin Notes, Research & Review: Discrete Mathematical Structure, 10 (3), 25-30.
9. Wang F, Ding J, Dai Z, Peng Y. An application of mobile phone encryption based on Fibonacci structure of chaos. In2010 Second World Congress on Software Engineering 2010 Dec 19 (Vol. 2, pp. 97-100). IEEE.
10. Tuncer T, Kurum HY. A novel Collatz conjecture-based digital image watermarking method. Cryptologia. 2022 Mar 4;46(2):128-47.