A Hybrid Encryption and Decryption Algorithm using Caesar and Vigenere Cipher

Camille Merlin S. Tan, Gerald P. Arada, Alexander C. Abad and Elmer R. Magsino
Electronics and Communications Engineering Department, Gokongwei College of Engineering, De La Salle University, Manila, Philippines

camille_merlin_s_tan@dlsu.edu.ph, gerald.arada@dlsu.edu.ph

Abstract. This paper presents a hybrid encryption and decryption algorithm, which combines the concept of Caesar Cipher and Vigenere Cipher algorithms. This was done by comparing the proposed design with some of the well-known ciphers such as Hill Cipher, Caesar Cipher and Vigenere Cipher through MATLAB simulations. Parameters such as letter frequency and graph behavior were evaluated. Results show that the proposed algorithm performs excellently. It has the most evenly distributed graph among other ciphers being compared in terms of their frequency of the ciphertext. Also, the decryption algorithm was able to achieve 99.99% accuracy.

1. Introduction
Cryptography deals with the science of making information secure and involves both encryption and decryption of data. It involves and provides several security measures, namely, authentication, confidentiality, and integrity. Modern cryptography now combines different cryptographic algorithms in generating cryptographic keys similar to this paper [1]. However, with the improvements in encryption, different methods of attacks were also developed. There are two main types of attacks such as passive attack and active attack. Passive attack refers to an attacker that only listens on a network segment and tries to read sensitive information, which can be done either online or offline. Whereas active attack refers to an attacker that intercepts, reading and altering the content before transmitting it to the target receiver, they may also delete it completely without reaching the receiver [2]. These cryptographic algorithms are most used in the field of Internet of Things (IoT) [3],[4], cybercrime prevention [5], WLAN, video broadcasting, secure payment and card transactions, and mobile communications [6].

There are two types of classical cryptography namely symmetric and asymmetric cryptography. For symmetric cryptography in Figure 1, it uses only one secret key for both the encryption and decryption process. Also, this type of cryptography does not have a good scalability. Symmetric cryptography can process fast and functions even without requiring lots of overheads on the central processing unit (CPU) and network resources, it is suitable for encrypting huge amounts of data such as the whole database and disk partition. In addition, symmetric encryption is prone to key exhaustion problems and key hierarchy must be maintained properly, else it is possible that the information would be leaked, and the attacker can remodel the secret key [2],[7].
On the other hand, the asymmetric cryptography in Figure 2 also known as the public key cryptography uses two different keys for the encryption and decryption process, all users of the asymmetric cryptosystem have both public and private keys [2]. The public key is accessible to everyone and it is used to encrypt the plain text message before sending it to the receiver. However, for the message to be decrypted, the private key is required. Therefore, the private key should be kept secret all the time [7].

For this paper, the proposed algorithm is compared to some well-known ciphers such as Hill Cipher, Caesar Cipher and Vigenere Cipher in terms of letter frequency of the ciphertext. A ciphertext with evenly distributed letter frequency is desired so that other people except the target recipient cannot guess the cipher pattern and unlock the original message. To have a better understanding of the Hill Cipher, Caesar Cipher, Vigenere Cipher as well as the proposed algorithm, each of them are discussed below.

The Hill Cipher was invented in 1929 by Lester Hill, it uses the concept of linear algebra, a background in matrices and modulo arithmetic are necessary to use this cipher. To encrypt, we need to have a secret key, this secret key should be converted into a 2x2 matrix when working with digraphs. The secret key is converted into numbers using A=0, B=1, C=2 down to Z=25 as reference. Whereas, the original text is grouped into column vectors, the letters are also converted in the same way as the secret key, matrix multiplication is done for each column vector and the secret key, the modulo 26 is applied to each of the resulting vector, then each vector is converted back into letters and combined in order to obtain the ciphertext. To decrypt, we should find the inverse matrix of the secret key first, then we multiply the inverse matrix with each of the column vectors of the ciphertext, apply modulo 26 to the resulting vector and convert the numbers back into letters [8].

As mentioned earlier, Caesar Cipher is one of the simplest and widely used algorithms, it is a type of symmetric cryptography. The original message is encrypted by shifting a specific amount of position then shifts back the same amount of position for decryption [9]. It is commonly applied in parts of other complex encryption algorithms due to its simplicity; its main disadvantage is that the ciphertext is easy to decrypt by anyone reading the text due to the different distribution of letters in the ciphertext. The equation for encryption can be expressed as Encryption(x)=(x+n) mod26, where x is the letter number (e.g., A=0, B=1 etc.) and n is the number of shift positions. While for decryption, a reverse process is performed and its equation can be expressed as decryption(x)=(x-n) mod26 [9],[10]. Figure 3 shows a Caesar Cipher with n=+3 shift positions.
The Vigenere Cipher algorithm in Figure 4 obtains the ciphertext by adding the secret key to the original message, when the secret key is shorter than the original message, the secret key is repeated until it reaches the last letter of the original text, the sum of modulo 26 gives the letter of the ciphertext. For decryption, it is basically the reverse process of encryption, it is done by taking the letters of the ciphertext and subtracting it with the secret key letters. When the sum is a negative number, 26 shall be added and converted to a letter, to obtain the original text [11].

Different types of techniques are used to protect data from being attacked by a third party such as multiple level image captcha [12], Neural Cryptography [13],[14], Elliptic Curve Cryptography (ECC) [15], Lettuce-based Cryptography (LBC) [16], Deoxyribo Nucleic Acid (DNA) key-based cryptography [17], Quantum Cryptography [13],[14],[18] and steganography [19].

2. Problem Formulation

Cryptographers apply number theory and some mathematical computations to develop new algorithms that are complex, hard to brute force with high security performance. However, many of these codes were still broken by hackers. There are trade-offs between simple algorithms and complex algorithms, simple algorithms can easily be processed but it can easily be decrypted by just observing the pattern of an understandable text or message. On the other hand, complex algorithms take more time to process, it can be hard to decrypt but sometimes to the point that it cannot be decrypted back to the original message after encryption, especially for asymmetric cryptosystems.

Table 1 shows the frequency distribution analysis of letters occurring in the Concise Oxford Dictionary, the letters are arranged from the most common to the least common letter appearing in the English text. One of the most common letters are E, A, R, I, O occupying 11.16%, 8.50%, 7.58%, 7.54%, 7.16% of an average English text, respectively. On the other hand, the least common letters are V, X, Z, J, Q occupying 1%, 0.29%, 0.27%, 0.20%, 0.20%, respectively. In addition, the third and sixth column show the proportion based on the least common letter which is Q. Setting Q as 1, the letter E is the most common letter, and it is around 57 times more common than the letter Q in forming English words [21].

Figure 5 shows the bar plot version of Table 1, it can be observed that the letters in the alphabet do not appear evenly distributed in an English text, some appear often compared to other letters, while some are very rare. Therefore, a ciphertext from a Caesar Cipher can easily be predicted, even though a shift is applied, the most common letter appearing in the text can be used to guess the number of shifted positions done in the alphabet. Thus, one way of ensuring that a ciphertext is hard to decrypt without knowing the algorithm used is by making all the letter frequencies as balanced as possible at the same time simple enough to process encryption and decryption when the algorithm is known [21].
Table 1. Average Letter Distribution in English Text [19]

| Letter | Frequency | Proportion | Letter | Frequency | Proportion |
|--------|-----------|------------|--------|-----------|------------|
| E      | 11.1607%  | 56.88      | M      | 3.0129%   | 15.36      |
| A      | 8.4966%   | 43.31      | H      | 3.0034%   | 15.31      |
| R      | 7.5809%   | 38.64      | G      | 2.4705%   | 12.59      |
| I      | 7.5448%   | 38.45      | B      | 2.0720%   | 10.56      |
| O      | 7.1635%   | 36.51      | F      | 1.8121%   | 9.24       |
| T      | 6.9509%   | 35.43      | Y      | 1.7779%   | 9.06       |
| N      | 6.6544%   | 33.92      | W      | 1.2899%   | 6.57       |
| S      | 5.7351%   | 29.23      | K      | 1.1016%   | 5.61       |
| L      | 5.4933%   | 27.98      | V      | 1.0074%   | 5.13       |
| C      | 4.5388%   | 23.13      | X      | 0.2902%   | 1.48       |
| U      | 3.6308%   | 18.51      | Z      | 0.2722%   | 1.39       |
| D      | 3.3844%   | 17.25      | J      | 0.1965%   | 1          |
| P      | 3.1671%   | 16.14      | Q      | 0.1962%   | 1          |

Figure 5. The Average Letter Distribution of English Text [19]

3. Methodology
The proposed algorithm in Figure 6 is based on the concept of Caesar Cipher and Vigenere Cipher with an improved performance by adding different number of shift positions depending on the line number.

Figure 6. The proposed algorithm’s flow chart
The number of shift position are designed differently to balance the frequency of the letters occurring in the ciphertext as much as possible, since some letters appear more frequently than others, we must change the number of shift positions so that common English letters can fill in other uncommon letters, while the uncommon letters get to be moved to the common letters to avoid increasing rapidly. In addition, the proposed algorithm is more suitable for long messages, as the number of lines increase, the frequency of each letter is filled in, which leads to a more balanced distribution. Some letters may not be able to be balanced out yet when the message is very short. Whereas for decryption, a reverse process is applied to obtain the original message, this can be done by shifting number of positions backwards depending on the number of shifted positions applied on the line. All the simulation were done in MATLAB.

4. Simulation Result, Analysis and Discussion

4.1. Encryption

Figure 7(a) shows the bar graph for sample message, this bar plot shows the total number of times a letter appears in the original text. The message used in this example is “Cryptography is the science of using mathematics to encrypt and decrypt data. Cryptography enables you to store sensitive information or transmit it across insecure networks such as the internet so that it cannot be read by anyone except the intended recipient. Cryptography is the study of secure communication techniques which allow only the sender and the target recipient of the message to read its contents. The term was derived from the Greek word kryptos, which means hidden. It is closely related to encryption, the act of scrambling ordinary text into ciphertext” [22]. The total number of letters for this text is 475. It can be observed that the frequency distribution of the letters in the alphabet is not evenly distributed. Indeed, there are a lot of letters E and T. On the other hand, there are no letters J and Z, while letters K, Q, V and X are very rare for the sample message.

![Figure 7](image)

**Figure 7.** The (a) original message and the resulting ciphertext obtained from (b) the proposed encryption algorithm (c) Hill Cipher (d) Caesar Cipher, and (e) Vigenere Cipher

Figures 7(b)-7(e) show the resulting ciphertext obtained from the proposed encryption algorithm, Hill Cipher, Caesar Cipher, Vigenere Cipher, respectively. Comparing their results with Figure 7(a), it can be observed that Figures 7(a) and 7(d) are the most similar, the only difference is the shifting of letter positions. Therefore, in terms of letter frequency comparison, the Caesar Cipher does not show a good
encryption, since the original message can easily be predicted by computing the number of shift positions from the original message. Figures 7(b), 7(c) and 7(e) show that letters behave differently compared to the original message, a pattern cannot easily be detected. However, Figure 7(b) shows the most desired graph among other ciphers, the overall bar graph shows an evenly distributed plot with few outliers. In addition, letters V, K and Z are the most common letters in the resulting ciphertext for the sample message, the result is ideal because these letters are one of the most uncommon in Figure 7(a), which allows attackers to hardly guess the original message without knowing the secret key of the message.

Table 2 shows the letter distribution of the original message and the ciphertext of each Cipher as well as the proposed algorithm. Since there are 26 letters in the alphabet, the mean percentage is 3.85%, the frequency should be around 18 for each letter to have an ideal frequency distribution. The proposed algorithm was able to have majority of the letters fall between the frequency range of 13% to 23%, which belongs to the ±5% of the ideal value. Also, it can be observed that the most common letter E having 13.30% of the original message, is now the most uncommon letter in the ciphertext with only 1.26% of the entire ciphertext after applying the encryption. This shows that the encryption was able to make the ciphertext very different from its original text. Therefore, it can be concluded that the proposed algorithm is most ideal among all types of ciphers presented since it displays the most evenly distributed and it behaves differently as compared to the original message.

Table 2. Letter Distribution in Sample Message and Ciphertext

| Letter | Original Message | Proposed Algorithm | Hill Cipher | Caesar Cipher | Vigenere Cipher |
|--------|------------------|--------------------|-------------|---------------|-----------------|
| a      | 13.30%           | 2.26%              | 3.16%       | 5.67%         | 3.21%           |
| b      | 12.45%           | 5.95%              | 5.01%       | 6.91%         | 3.99%           |
| c      | 7.73%            | 3.15%              | 3.16%       | 2.11%         | 1.95%           |
| d      | 7.81%            | 3.15%              | 6.32%       | 0.21%         | 3.95%           |
| e      | 7.00%            | 2.73%              | 2.20%       | 7.56%         | 4.20%           |
| f      | 6.65%            | 2.94%              | 6.84%       | 7.35%         | 5.25%           |
| g      | 6.22%            | 1.68%              | 3.95%       | 2.15%         | 3.75%           |
| h      | 7.79%            | 2.73%              | 4.39%       | 0.14%         | 2.73%           |
| i      | 6.15%            | 5.04%              | 3.15%       | 4.83%         | 4.62%           |
| j      | 4.94%            | 3.99%              | 2.45%       | 1.60%         | 3.19%           |
| k      | 3.65%            | 3.35%              | 4.56%       | 5.67%         | 2.10%           |
| l      | 3.65%            | 2.73%              | 3.96%       | 0.63%         | 1.05%           |
| m      | 3%               | 3.78%              | 2.06%       | 0.03%         | 1.95%           |
| n      | 2.36%            | 3.36%              | 1.93%       | 1.02%         | 6.72%           |
| o      | 1.72%            | 3.36%              | 1.56%       | 1.28%         | 5.87%           |
| p      | 1.72%            | 2.99%              | 4.91%       | 0.63%         | 1.95%           |
| q      | 1.72%            | 4.20%              | 5.09%       | 12.19%        | 1.05%           |
| r      | 1.29%            | 3.76%              | 5.61%       | 13.03%        | 7.77%           |
| s      | 1.29%            | 4.83%              | 2.91%       | 0.42%         | 3.95%           |
| t      | 0.86%            | 3.15%              | 2.63%       | 0.98%         | 5.25%           |
| u      | 0.86%            | 6.30%              | 3.33%       | 0%            | 3.36%           |
| v      | 0.54%            | 3.36%              | 5.09%       | 1.26%         | 5.25%           |
| w      | 0.54%            | 4.81%              | 4.74%       | 1.00%         | 4.82%           |
| x      | 0.54%            | 4.83%              | 1.75%       | 2.94%         | 4.62%           |
| y      | 0.54%            | 3.36%              | 4.21%       | 7.14%         | 0.63%           |
| z      | 0%               | 3.36%              | 4.35%       | 3.57%         | 3.15%           |

4.2. Decryption

Figure 8 shows the decrypted message using the proposed algorithm, it was able to convert the encrypted message back into its original message without any error. Therefore, it can be concluded that the decryption algorithm has an accuracy of 99.99%.
5. Conclusion

There are various types of cryptography algorithms, important concepts such as encryption, decryption, secret key, ciphertext and the overall process in cryptography. Cryptography is applied to achieve data integrity, confidentiality, authentication, security, and many more by transmitting information through insecure networks such as the internet. Cryptography is necessary especially in today’s digital world, almost everything is accessible through the internet such as education, shopping, banking, social media, and web browsing. With all these services being offered online, the problem of security in communication becomes more challenging since there is so much data, which exceeds the capacity of what the computer software can process known as the big data, this leads to a serious problem when it comes to data security and privacy. In this paper, the authors successfully implemented hybrid encryption and decryption processes by combining the Caesar Cipher and Vigenere Cipher algorithms.

For future research, the proposed algorithm can further be improved by considering the frequent pairs in the English text as well such as NG, TT, TH, IN, AN, ST, HE, TO, AS, IT, IS, OF, AT, ON and many more, a certain rule can be deployed for the common pairs in order to make the ciphertext more unpredictable. Also, the author recommends focusing on making cryptosystems that are more flexible, so that it can process technical changes quickly. With the advanced technology nowadays, quantum computers are widely used by various attackers, which allow them to attack cryptographic algorithms faster compared to classical computers. Therefore, systems that are using cryptography algorithms should continuously be updated and improved to fight against new types of attacks in the future. While there are more threats and challenges happening due to the available technology resources, researchers should also make use of these technologies to make a more powerful cryptography, in which it can alter some parameters or key automatically when an attack is detected. Applying several layers of different cryptography algorithms can also be a solution.

References

[1] Liu D et.al 2009 Next Generation SSH2 Implementation, Securing Data in Motion Syngress pp. 41-64.
[2] Cryptography 2020 [Internet] Retrieved from: https://www.synopsys.com/glossary/what-is-cryptography.html
[3] Rawat D B and Ghafoor K Z 2019 Smart Cities Cybersecurity and Privacy Elsevier pp. 75-88.
[4] Kumar A Jain V and Yadav A 2020 A New Approach for Security in Cloud Data Storage for IOT Applications Using Hybrid Cryptography Technique International Conference on Power Electronics & IoT Applications in Renewable Energy and its Control (PARC) 514-517.
[5] Shinder L and Cross M 2008 Scene of the Cybercrime 2nd edition Syngress pp. 505-554.
[6] Martin K M 2017 Everyday Cryptography: Fundamental Principles and Applications 2nd edition Oxford University Press UK.
[7] What is Encryption & How Does it Work? 2017 [Internet] Retrieved from: https://medium.com/searchencrypt/what-is-encryption-how-does-it-work-e8f20e340537
Accessed on: Jan. 29, 2021

[8] Barr T H 2002 Invitation to Cryptology *Prentice Hall, Inc.*, New Jersey.

[9] Caesar Cipher in Cryptography 2019. [Internet] Retrieved from: https://www.geeksforgeeks.org/caesar-cipher-in-cryptography/ Accessed on: Jan. 30, 2021.

[10] Caesar Cipher [Internet] Retrieved from:http://practicalcryptography.com/ciphers/caesar-cipher/ Accessed on: Jan. 30, 2021.

[11] Vigenere Cipher [Internet] Retrieved from:https://www.dcode.fr/vigenere-cipher Accessed on: Jan. 30, 2021.

[12] Ogiela M R and Ogiela L 2018 Cognitive cryptography in advanced data security *IEEE 32nd International Conference on Advanced Information Networking and Applications (AINA)* pp. 740-743.

[13] Sharma A and Sharma D 2016 Big data protection via neural and quantum cryptography 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom).

[14] ÖzÇakmak B Öz bílen A Yavanoğlu U and Cın K 2019 Neural and quantum cryptography in big data: A review 2019 *IEEE International Conference on Big Data (Big Data)* pp. 2413-2417.

[15] Prathama D R Putro P W and Naviangga D I 2018 Secure mobile payment based on elliptic curve cryptography 2018 *International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)* pp. 140-145.

[16] Nejatollahi H Dutt N and Cammarota R 2017 Special session: trends, challenges and needs for lattice-based cryptography implementations 2017 *International Conference on Hardware/Software Codesign and System Synthesis (CODES+ISSS)* pp. 1-3.

[17] Akiwate B and Parthiban L 2018 A dynamic DNA for key-based cryptography 2018 *International Conference on Computational Techniques, Electronics and Mechanical Systems (CITEMS)* pp. 223-227.

[18] Pawar H R and Harkut D G 2018 Classical and quantum cryptography for image encryption & decryption 2018 *International Conference on Research in Intelligent and Computing in Engineering (RICE)* pp. 1-4.

[19] Gladwin S J and P Lakshmi Gowthami 2020 Combined cryptography and steganography for enhanced security in suboptimal images 2020 *International Conference on Artificial Intelligence and Signal Processing (AISP)* pp. 1-5.

[20] The Frequency of the Letters of the Alphabet in English 2021 [Internet] Retrieved from: https://www3.nd.edu/~busiforc/handouts/cryptography/letterfrequencies.html Accessed on: Feb. 1, 2021.

[21] Specktor B 2018 Can You Guess the Most Common Letters in the English Languages? [Internet] Retrieved from: Available: https://www.rd.com/article/common-letters-english-language/ Accessed on: Feb. 1, 2021.

[22] Reddy P 2021 Real Life Applications of Cryptography [Internet] Retrieved from:Available: https://medium.com/@prashanthreddyt1234/real-life-applications-of-cryptography-162ddf2e917d Accessed on: Feb. 2, 2021.