Algorithmic Decryption of Substitution Cipher's

A Technical Report submitted to the Department of Computer Science

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Kevin Bruzon

Spring, 2023

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Briana Morrison, Department of Computer Science

Rosanne Vrugtman, Department of Computer Science

Algorithmic Decryption of Substitution Ciphers

CS4991 Capstone Report, 2023

Kevin Bruzon
Computer Science
The University of Virginia
School of Engineering and Applied Science
Charlottesville, Virginia USA
kb5ux@virginia.edu

ABSTRACT
Although the recent wave of technological advancements offers many benefits, society is not fully aware of the increased threats to data privacy and security yielded by this wave. This meta-study examines, analyzes and synopsizes a variety of research papers and articles regarding advancements in decryption algorithms for substitution ciphers and discusses the implications posed to data privacy. The goal of this review is to shed light and raise awareness on the growing data privacy threat yielded by technological and algorithmic advancements, which are facilitating not only the access of secure and private data, but also the development of methods for access. Future work could involve analyzing the other side of cryptography, and evaluating advancements in more secure encryption algorithms that can combat the increased privacy threat yielded by technological advancements.

1. INTRODUCTION

The 21st century has seen a rapid increase in the use of technology embedded into everyday life. This growth has led to both increased concerns and threats regarding data privacy, with Feistel, a German-American cryptographer, warning in the late 20th century that "there is growing concern that computers now constitute, or will soon constitute, a dangerous threat to individual privacy" (1973). To achieve data privacy, many concepts of cryptography must be employed. "Cryptography is an algorithmic process of converting a plain text or clear text message to a cipher text or cipher message based on an algorithm that both the sender and receiver know" (Babu et al., 2014).

Cryptography consists of two major components: encryption and decryption. According to Babu, et al. (2014), encryption is the process of coding information, either a file or message, into a cipher text, yielding unreadable data which can only be interpreted by someone who possesses the decoding key. Decryption is the process of decoding the encrypted data using the provided key, yielding the original plain text form of the data. A variety of algorithms exist for encrypting data, otherwise referred to as ciphers. A substitution cipher consists of codes in which every single letter in the alphabet has one fixed substitute (Peleg & Rosenfeld, 1979).

As the use of advanced computing technologies increases and becomes more embedded in society, it is critical to assess and document any new algorithms that might compromise data privacy and security. These algorithms, which are capable of deciphering data at faster and more efficient rates than traditional methods, pose a risk to data privacy.

2. RELATED WORKS

A variety of research papers and literature regarding advances in secure encryption algorithms exist. However, very few discuss

advances in decryption algorithms besides the previously developed methods, and even fewer discuss the threat posed to data privacy by these advances.

On the topic of decryption, a number of experts in the field of cybersecurity conducted an empirical study on the effectiveness of newly developed decryption tools for ransomware data. In this study, Filiz, et al. (2020) tested a total of 78 different decryption tools against 61 different ransomware samples. Their results indicated that 55% of the decryption tools tested were successful at decryption and recovering the data from the ransomware samples. Although the effectiveness of the decryption tools was somewhat high, the researchers discuss a need for advancements to improve the tools so that they are more effective at decrypting and recovering ransomware data. This research revealed how new developments and advancements in decryption are facilitating the access of encrypted data. However, I take a different lens, analyzing the drawback of advances in decryption, rather than the positive impact that can be yielded by advances.

On the other hand, the CEO of Theon Technology recently wrote a Forbes article on the topic of threats posed by decryption algorithms. In his article, Bledsoe (2022) discusses methodologies on ways to protect data against the threat posed by algorithmic decryption. Although the majority of the content consists of the methodologies to combat advances in decryption algorithms, Bledsoe posits: "Algorithmic decryption advancements by cybercriminals pose a growing threat to decades-old standard cryptographic algorithms" (para. 2). Furthermore, he indicates that various government organizations are currently developing and upgrading cryptographic methods due to the growing threat yielded by advances in computing, more specifically quantum computing. This serves as reinforcement to the overall significance of shedding light on how advances in algorithmic decryption poses a grave threat to the privacy and security of data.

3. METHODOLOGY

For this meta-study, I analyzed and synthesized a variety of research papers, articles, and literature. The papers examined focused on the topic of advancements in decryption algorithms and their implications for data privacy.

3.1 Search Strategy

I found these sources by conducting an extensive literature search using various search engines such as Google Scholar and the UVA Library Database. The search terms used were "decryption algorithms", "substitution ciphers", "data privacy", and "algorithmic decryption"

3.2 Selection Criteria

I then individually analyzed the sources for their content. The inclusion criteria were papers written after 2010, published in English, written by computer scientists or cryptographers, and relevant to the research question. If any of the inclusion criteria was not met, I excluded the source.

3.3 Meta-Analysis

I analyzed findings and conclusions made by the authors regarding their algorithms and decryption success. I then synthesized the data and information collected from the analyzed sources then drew conclusions and implications on the topic of data privacy and security in relation to decryption algorithms.

3.4 Meta-Synthesis

One of the only algorithms that has existed for decryption of substitution ciphered data was a human implementing a brute-force algorithm to decode through letter frequency analysis and pattern matching (Jain et al.,

2015). In the case of a shift substitution cipher, the brute-force approach consists of testing all possible shifts, both left and right. In order to better visualize the process of this brute-force decryption, Figure 1 depicts the process of first encrypting the original data.

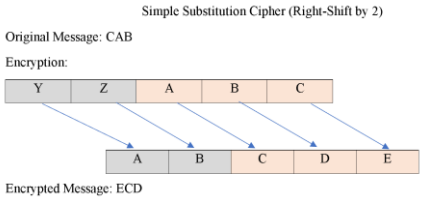

Figure 1 : Encryption of a simple substitution cipher using a right-shift of 2

Depicted in this figure is the encryption process for a message, CAB, using a substitution cipher shifting letters to the right by two. Letter Y maps to A, Z to B, A to C, B to D, and C to E. The resulting encrypted message is ECD. Figure 2 depicts the brute-froce decryption of th encrypted message, ECD.

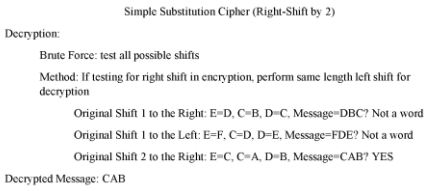

Figure 2: Decryption of a simple substitution cipher using a brute-force approach

Depicted in this figure is the decryption process for a message, ECD, using a brute-force solution. The brute-force solution depicted in the figure consists of testing all possible shifts in both directions. After testing a shift by one in both directions, testing a shift by two to the right yields the original message, CAB. With the rapid growth of advanced computing technologies, technological embeddedness in society, and growing privacy concerns, it is essential to analyze and report on new algorithms, which pose potential threats to data privacy and security, as they can decrypt ciphered data faster and more efficiently than previously established approaches.

Corlett and Penn (2010), developed an adaptation of the Viterbi algorithm to solve letter-substitution ciphers. The Viterbi algorithm is a dynamic programming algorithm used to find the most likely sequence of hidden states—called the Viterbi path—that results in a sequence of observed events, especially in the context of Markov information sources and hidden Markov models (HMMs). However, in this case, Corlett and Penn utilize a generalized version, wrapping it in an A* search, which is a heuristic search algorithm used in pathfinding and graph traversal to find the shortest path between two points, using a combination of the cost of each step along the path and an estimate of the remaining cost to the goal.

Eight years later, Alkazaz et al. (2018) developed various Prediction by Partial Matching (PPM) algorithms to decrypt substitution ciphered data. These algorithms use the PPM text compression scheme, which is “an adaptive statistical coding approach, which dynamically constructs and updates fixed order Markov-based models that help predict the upcoming character relying on the previous symbols or characters being processed” (p. 58).

4. RESULTS & DISCUSSION

Alkazaz et al.’s PPM text compression algorithms, yield 92% of simple substitution ciphers being deciphered without errors (2018). Corlett and Penn’s adaption of the Viterbi algorithm yielded 100% accuracy on their test set. (2010). Technological advances are made every day, as the research and development of technologies tends to progress with time. Advanced decryption algorithms, such as the ones proposed by Alkazaz et al. or Corlett and Penn, create room for

vulnerabilities regarding data privacy, as private ciphered data could potentially be accessed and decrypted with these advanced decryption algorithms.

The advancement of decryption algorithms poses a great threat to data privacy and security. This is due to the fact that these algorithms are able to decipher encrypted data much more quickly and efficiently than traditional methods, thus making it easier for malicious actors to gain access to sensitive data. In addition, these algorithms are capable of deciphering data that has been encrypted using strong encryption algorithms, such as the substitution cipher. This means that data that was considered secure and private, can now be accessed with relative ease.

The threat posed by decryption algorithms is not only limited to data privacy and security, but also to the concept of trust in technology. In the past, the security of data was based on the assumption that the encryption used was strong enough to prevent any malicious actors from gaining access to the data. However, with the advent of these new algorithms, this assumption no longer holds true, and thus people are left with a feeling of distrust and unease when it comes to trusting technology with their data.

5. CONCLUSION

In conclusion, this meta-study has examined, analyzed and synthesized a variety of research papers and articles regarding advancements in decryption algorithms for substitution ciphers and discussed the implications posed to data privacy. The results of this meta-study indicate that the advancement of decryption algorithms poses a great threat to data privacy and security, as these algorithms are able to decipher encrypted data much more quickly and efficiently than traditional methods, thus making it easier for malicious actors to gain access to sensitive data. Furthermore, this meta-study has highlighted the need for advancements in secure encryption algorithms that can combat the increased privacy threat yielded by technological advancements.

6. FUTURE WORK

The research conducted in this paper is only the first step in analyzing the implications of advances in decryption algorithms for data privacy. Future research should focus on evaluating and analyzing the advancements in secure encryption algorithms that can counter the increases threats posed by decryption algorithms. Furthermore, research should also be conducted on the development of decryption algorithms for other ciphers such as stream ciphers, block ciphers, and elliptic curve ciphers, as these ciphers are also used to encrypt data. Additionally, research should be done to analyze how advances in computing, such as quantum computing, are further impacting the development of decryption algorithms and how this poses a greater threat to data privacy. With the continued advances in technology, it is essential to analyze and document the implications posed to data privacy and security.

REFERENCES

Alkazaz, N. R., Irvine, S. A., & Teahan, W. J. (2018). An automatic cryptanalysis of simple substitution ciphers using compression. *Information Security Journal: A Global Perspective*, 27(1), 57-75. https://doi.org/10.1080/19393555.2018.1426799

Bledsoe, S. (2022, December 28). Council post: How to protect your data from algorithmic decryption. *Forbes*.

Corlett, E., & Penn, G. (2010). An exact A* method for deciphering letter-substitution ciphers. *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, 1040–1047. https://aclanthology.org/P10-1106

Feistel, H. (1973). Cryptography and Computer Privacy. *Scientific American*, 228(5), 15–23. http://www.jstor.org/stable/24923044

Filiz, B., Arief, B., Cetin, O., & Hernandez-Castro, J. (2021). On the Effectiveness of Ransomware Decryption Tools. *Computers & Security*, 111, 102469. doi:10.1016/j.cose.2021.102469

Jain, A., Dedhia, R., & Patil, A. (2015). Enhancing the security of Caesar cipher substitution method using a randomized approach for more secure communication. *International Journal of Computer Applications*, 129, 6–11. doi:10.5120/ijca2015907062

Peleg, S., & Rosenfeld, A. (1979). Breaking substitution ciphers using a relaxation algorithm. *Communications of the ACM*, 22(11), 598–605. doi:10.1145/359168.359174

Magesh Babu, V., Shankar Ganesh, T., & Ramraj, K. (2018); A comparative analysis on encryption and decryption algorithms; *International Journal of Scientific and Research Publications*, 4(12). http://www.ijsrp.org/research-paper-1214.php?rp=P363462