ALGORITHMIC DECRYPTION OF SUBSTITUTION CIPHERS

SOCIO-POLITICAL INFLUENCES ON DATA PRIVACY AND SECURITY

A Thesis Prospectus
In STS 4500
Presented to
The Faculty of the
School of Engineering and Applied Science
University of Virginia
In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science in Computer Science

By
Kevin Bruzon

December 2, 2022

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.

ADVISORS

Catherine D. Baritaud, Department of Engineering and Society

Briana Morrison, Department of Computer Science

The 21st century served as a platform for rapid technological growth, with technology becoming more embedded into society and the everyday lives of humans. The integration of technology endured an even greater growth in recent years, during the COVID-19 pandemic, where many human interactions began to be exercised through technology, rather than in-person. With this technological growth, public concern regarding data privacy also experienced rapid growth, however, the concern was not a new one. In the late 20th century, Horst Feistel, a German-American Cryptographer, revealed "there is growing concern that computers now constitute, or will soon constitute, a dangerous threat to individual privacy" (1973, p. 15). Technological growth and embeddedness into society has fueled the growth of both privacy concerns and privacy threats. It is crucial to raise awareness on these concerns, and ensure the implementation of measures to mitigate data privacy concerns and threats. In order for data privacy to be achieved, many concepts of cryptography must be employed. Cryptography is a well-known method, dating back to 400 BC, used to secure the transmission and storage of data. In simpler terms, "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" (Magesh Babu et al., 2014, p. 1). This process consists of two major components: encryption and decryption. According to Magesh Babu et al., 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 whoever possesses the decoding key, while decryption is the process of decoding the encrypted data using the provided key, yielding the original plain text form of the data (2014).

The technical component of this thesis will examine established decryption methods and algorithms for substitution ciphers, analyze and synthesize research regarding more efficient decryption methods and algorithms for substitution ciphers, and discuss the potential implications for data privacy and security. The STS component of this thesis will examine data privacy legislation from various nations around the globe, analyze the role social and political contexts play in shaping the protection of data privacy, and discuss recommendations and potential courses of action to ensure effective data privacy legislation is implemented. The technical and STS components are tightly coupled, both addressing the topic of data privacy and security. The objective is that, combined, the research will shed light on the threat technological advances and developments pose for data privacy and security, as well as the importance of strong data privacy legislation as a device for ensuring the protection of data. Research pertaining to both the technical and STS components is ongoing, and will continue through the Spring 2023 semester. The technical component will be completed in CS 4991 during the Spring 2023 semester, and the STS component will be completed in STS 4600 during the Spring 2023 semester. The timetable for both components is outlined by a Gantt chart in Figure 1. The personnel involved in the completion of both components include Kevin Bruzon, an undergraduate student in the Department of Computer Science, as the primary author for both reports, Brianna Morrison, an associate professor in the Department of Computer Science, as the advisor for the technical component, and Catherine Baritaud, a senior lecturer in the Department of Science, Technology and Society, as the advisor for the STS research component.

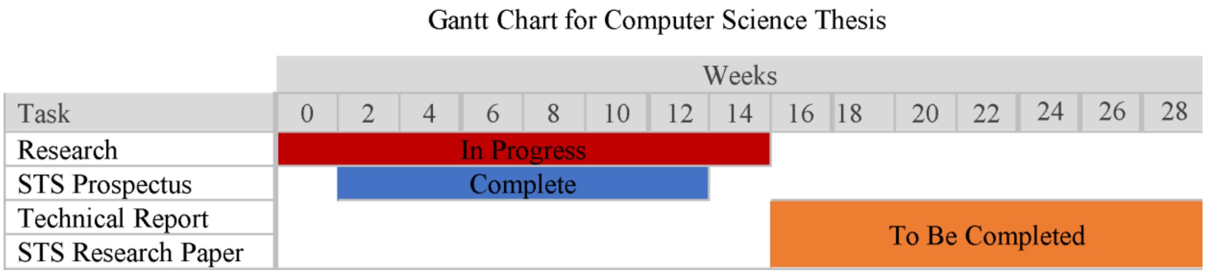


Figure 1: Gantt Chart for Computer Science Thesis. This figure depicts a timetable, outlining the major components of the thesis, and the expected timeline for completion of each component. The timeline consists of the Fall 2022 semester, and Spring 2023 semester. (Bruzon, 2022).

ALGORITHMIC DECRYPTION OF SUBSTITUTION CIPHERS

A substitution cipher consists of codes in which every single letter in the alphabet has one fixed substitute (Peleg & Rosenfeld, 1979). David Oranchak, an engineer from Walden University, similarly indicates, “simple substitution ciphers encrypt plaintext messages using symbols which map to individual plaintext letters” (2008, p. 1717). Figure 2 depicts a simple substitution cipher with a shift, outlining the mapping of letters to a substitute and the rendered ciphered text. The shift substitution cipher in Figure 2 uses a shift of two to the right, therefore letter A maps to C, B maps to D, C maps to E, and so on. The message being encrypted in the figure, is a simple message, CAB, using the first three letters of the alphabet for simplicity.

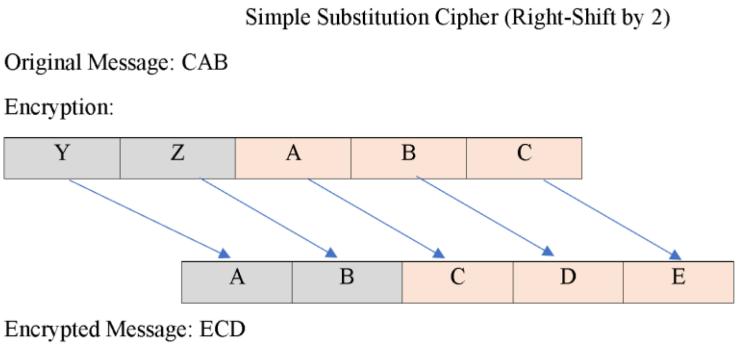
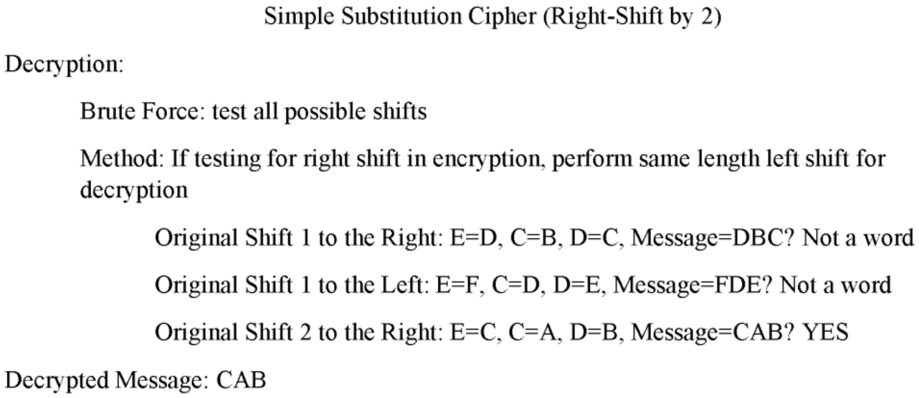


Figure 2: Encryption of a simple substitution cipher using a right-shift of 2. Depicted in this figure is the encryption process for a message, CAB, using a substitution cipher shifting letters to the right by two. Letter Y maps to A, Z to B, A to C, B to D, and C to E. The resulting encrypted message is ECD. (Bruzon, 2022).

Figure 3 outlines a brute-force solution, as was previously done before the development of advanced computing technologies. The methodology behind the solution is testing all possible shifts, left and right, and determining if the decryption resembles a potential message.



Simple Substitution Cipher (Right-Shift by 2)

Decryption:

Brute Force: test all possible shifts

Method: If testing for right shift in encryption, perform same length left shift for decryption

Original Shift 1 to the Right: E=D, C=B, D=C, Message=DBC? Not a word

Original Shift 1 to the Left: E=F, C=D, D=E, Message=FDE? Not a word

Original Shift 2 to the Right: E=C, C=A, D=B, Message=CAB? YES

Decrypted Message: CAB

Figure 3: 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. (Bruzon, 2022).

One of the only algorithms that existed for decryption of substitution ciphered data was a human implementing a brute-force algorithm to decode through letter frequency analysis as well as pattern matching (Jain et al., 2015). 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. According to Alkazaz, a senior lecturer for the Department of Computer Science at the University of Baghdad, Irvine, a computer scientist and R&D engineer, and Teahan, a lecturer for the Department of Computer Science at Bangor University, using a Prediction by Partial Matching (PPM) text compression algorithm, yields 92% of simple substitution ciphers being deciphered without errors (2018). Corlett, a master's of computer science from the University of Toronto, and Penn, a professor in the Department of Computer Science at the University of Toronto, discuss using an adaptation of the Viterbi algorithm to solve letter-substitution ciphers, which yielded 100% accuracy. (Corlett and Penn, 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 one 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 (2018; 2010). Data privacy and security is at risk, and the public must be aware of the implications technological growth is posing for the privacy and security of personal data.

The technical project will describe the current state of the art in research on algorithmic decryption of substitution ciphers. The objective of the technical work is to present an overview of established algorithms for the decryption of substitution ciphers, analyze the state of the art of research in finding advanced and more efficient decryption algorithms for substitution ciphers, and discuss the implications for data privacy and security. The analysis will offer insights on the threat technological growth, and more specifically, advances in algorithmic decryption, poses for data privacy and security. The findings of the technical project will be presented in a state-of-the-art report. The primary sources for research will consist of academic articles, research papers, and any relevant open-source code. These documents will provide insight into advances in decryption algorithms, and the risks posed to data privacy as a result of advanced decryption techniques. The anticipated outcome of this state-of-the-art report is for not only engineers, but all members of society, to be aware of the growing privacy threat yielded by technological advances. Advances in algorithmic decryption are facilitating not only the access of secure and private data, but also the development of methods for access.

SOCIO-POLITICAL INFLUENCES ON DATA PRIVACY AND SECURITY

Data privacy legislation provides a legal framework on the collection, use, and storage of personal data. In order to mitigate data privacy concerns, many governments around the world have implemented data privacy legislation However, a grave issue exists in the fact that data privacy legislation is not always effective, with many nations implementing weak frameworks that don’t address public concerns regarding the privacy and security of personal data. In China, the emergence of technology embedded in urban development has driven the public to raise privacy concerns as a result of weak legal frameworks for data privacy protection (Yang & Xu, 2018). Meanwhile in Australia, due to criminal acts and terrorism, the government issued legislation which grants law enforcement agencies access to “an unlimited range of technical assistance, extending beyond decryption to include modifying consumer products and services” (Hardy, 2020, para. 3). Similarly, many European nations have joined the debate of granting access to private encrypted data dueto crime and terrorism. (Severson, 2017). According to Severson, a Harvard Law School graduate, in France there is legislation in place that grants law enforcement agencies technical assistance in gathering information during criminal investigations (2017). Many data privacy legislations have shifted to loosening access restrictions, with many beginning to provide data access to government groups and law enforcement agencies. Rather than mitigating concerns of the public, different socio-political contexts are driving data privacy legislation into a state of weakness and ineffectiveness, yielding more room for threats and concern. This marks a grave issue, as throughout the years, the wide-spread application of technology and internet in everyday lives has drastically risen, thus presenting a further danger to personal data with more and more data being stored on the internet.

According to Vernon Andrews, a bachelor of cybersecurity from Columbus State University, today's methods of data storage on the internet have resulted in an emergence of concerns and issues pertaining to data privacy (2019). Furthermore, according to Yang, a researcher at the University of Melbourne and Xu, and Xu, a senior lecturer in communication at Deakins University, privacy issues have been an ongoing debate in the United States since the Internet was first made public (2018). Andrews also reveals, from the results of a questionnaire, that society has a lack of knowledge regarding data privacy issues (2019). This presents a large complication and a need for research, as technological growth has fueled the growth of data privacy issues, which much of the public lack's awareness of, yet various nations have begun opening the doors for access of private data with loosened restrictions in their legal frameworks. In order to mitigate data privacy threats and concerns, it is crucial for governments and nations to implement strong legal frameworks, driven by aware socio-political contexts, which effectively address public concern, and protect the access of personal data.

The STS research will examine the data privacy legislation of various nations around the globe and analyze the role social and political contexts play in shaping said data privacy legislation. The objective of the STS project is to gain insights on both the socio-political contexts surrounding weak data privacy legislation and the socio-political contexts surrounding strong data privacy legislation, and explore methodologies which social and political contexts can employ in the development of data privacy legislation. Applying these methodologies in data privacy legislation would promote positive reinforcement from socio-political contexts, while addressing the concerns of the public, thus yielding effective data privacy protections. Bart

Jacobs and Jean Popma, Dutch computer scientists and cybersecurity experts at Radboud University, discuss potential methodologies that contribute to protecting data privacy. The authors discuss four processes that must be implemented in medical research to effectively protect data privacy, and they apply this framework to a Parkinson's research project they both were involved with at the time. They provide a detailed analysis of these four processes, concluding that data can be securely and privately managed if the processes are implemented by a multidisciplinary network of stakeholders. Yang and Xu arrive at similar conclusions in their research, arguing that interactions and cohesiveness between political, commercial, and social contexts can help ensure data privacy is protected and preserved.

Through the examination of differing data privacy legislations, the processes, contexts, and methodologies shaping said legislation will be analyzed, with the aim of revealing pitfalls of legislation and potential improvements to enhance data privacy protection. The analysis will apply the Social Construction of Technology (SCOT) framework, originally introduced in 1984 by Trevor Pinch and Wiebe Bijker. Although this framework is geared towards a technical subject, it will be adapted to a non-technical subject, data privacy legislation, still outlining the relationships between the subject and social and political groups. Figure 4 depicts this framework in the context of strong data privacy legislation, outlining the positive relationships between data privacy legislation and the different socio-political contexts and groups that shape data privacy legislation, as well as the protection of data privacy.

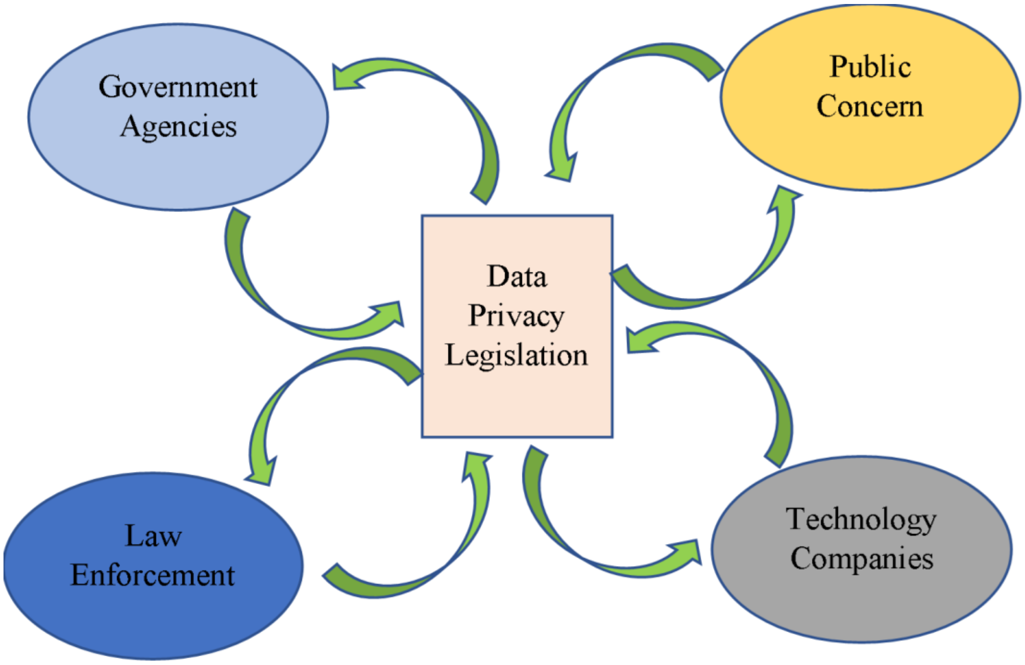


Figure 4: Strong Data Privacy legislation SCOT model. This figure depicts the application of the SCOT framework, outlining the different social and political groups that shape strong data privacy legislation. A positive relationship between the groups and data privacy legislation is outlined by the green color in the arrows, and the size of the shapes depicts a balanced influence across all the groups. (Adapted by Bruzon (2022) from Carlson, 2009)

Groups such as the government, law enforcement agencies, technology companies, and the public all play a role in shaping the strength of data privacy legislation, and the influences of these groups are driven by social and political contexts. However, data privacy legislation also holds a role and influences these groups and contexts as well. This can be noted in Figure 4, where public concern positively reinforces the implementation of protections in legislation, while modifications and improvements to the legislation serve as a tool for mitigation, positively affecting the public by addressing their concerns. Figure 5 serves as a juxtaposition, and depicts the SCOT framework in the context of weak data privacy legislation, outlining the relationships between data privacy legislation, and the different socio-political contexts and groups that shape data privacy. It outlines the failures in weak data privacy legislation as it reveals incomplete relationships, where not all groups and contexts have an impact on legislation, or the impacts serve as negative reinforcement which rather than protecting privacy, they open the door for access to private data. In Figure 5, public concern is depicted with only one arrow, as most weak legal frameworks tend to ignore the concerns of the public, and fail to address their needs. In the case of law-enforcement and government agencies, the impact they have on legislation in situations of weak frameworks is negative, as it loosens protections on data privacy. However, they receive a positive impact as the loosening of protections facilitates the carrying out of their agendas.

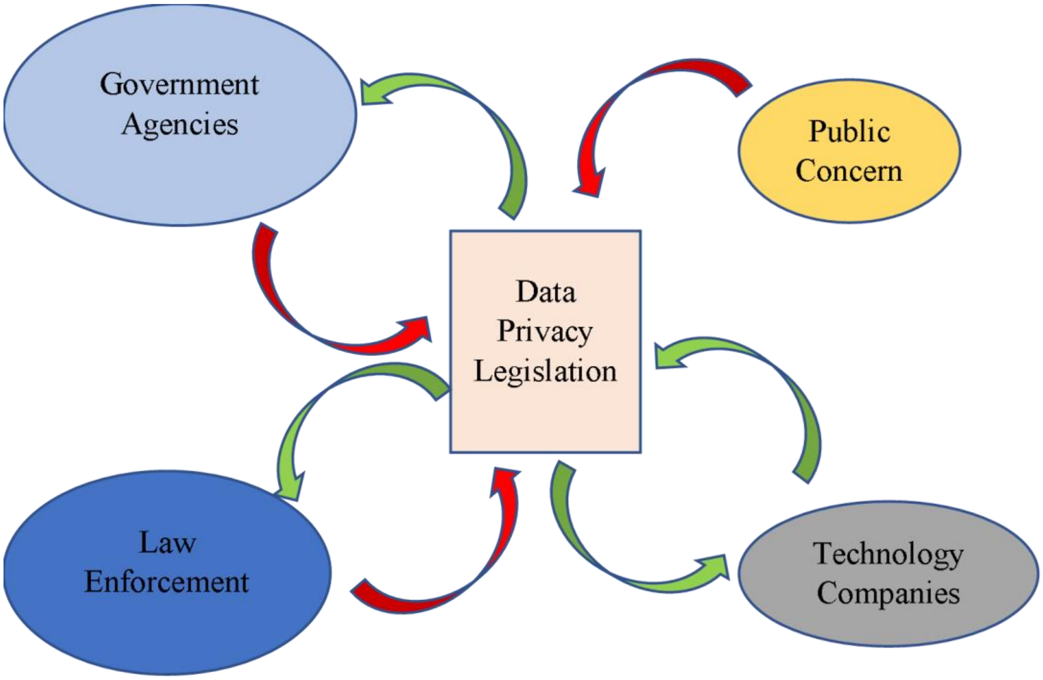


Figure 5: Weak Data Privacy legislation SCOT model. This figure depicts the application of the SCOT framework, outlining the different social and political groups that shape weak data privacy legislation. A negative relationship between the groups and data privacy legislation is outlined by the reed color in the arrows, and the size of the shapes depicts an imbalanced influence across all the groups. (Adapted by Bruzon (2022) from Carlson, 2009)

Research for the STS topic will come from a variety of sources such as journal articles, research papers, and legal documents. This research will be presented in the form of a scholarly article, with the anticipated outcome that governments, technology companies, and the public are not only aware of privacy concerns and threats, but that they are responsive to the concerns and growing threats and employ methodologies, with the consideration and influence of socio-political contexts, to ensure effective data privacy legislation is implemented to promote the protection of personal data.

DATA PRIVACY LEGISLATION COMBATS PRIVACY THREATS

Modern society is constantly striving for progress, aiming to move forward and make everyday life easier. The growth of technology has fueled societal progress, with technology being a contributing factor to the completion of nearly every task possible. The impact of technological growth has been tremendous, with advances in algorithmic decryption, and increased embeddedness in everyday life. The advances, growth, and increased embeddedness of technology has posed grave risks for data privacy, and it is essential for society to be aware of the implications technological progress poses. A solution, which attempts to address growing privacy concerns as a result of technological growth, exists in data privacy legislation. Many nations across the world developed legislation, aiming to protect data, and mitigate privacy concerns, however, many have been weak. Technological advances and developments pose a grave threat to data privacy and security. Data privacy legislation must be used as a device to combat the threats posed by technological growth, employed by a multidisciplinary network of socio-political contexts, ultimately yielding the protection of data privacy.

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

Andrews, V. (2019) Analyzing awareness on Data Privacy. *Proceedings of the 2019 ACM Southeast Conference* (pp. 198-201). Association for Computing Machinery. https://doi-org.proxy01.its.virginia.edu/10.1145/3299815.3314458

Bijker, W. E., & Pinch, T. J. (1984). The social construction of facts and artefacts: or how the sociology of science and the sociology of technology might benefit each other. *Social Studies of Science,* 14(3), 399–441. https://doi.org/10.1177/030631284014003004

Bruzon, K. (2022). Gantt Chart for Computer Science Thesis. [Figure 1]. *Prospectus* (Unpublished undergraduate thesis). School of Engineering and Applied Science, University of Virginia. Charlottesville, VA.

Bruzon, K. (2022). Encryption of a simple substitution cipher using a right-shift of 2. [Figure 2]. *Prospectus* (Unpublished undergraduate thesis). School of Engineering and Applied Science, University of Virginia. Charlottesville, VA.

Bruzon, K. (2022). Decryption of a simple substitution cipher using a brute-force approach. [Figure 3]. *Prospectus* (Unpublished undergraduate thesis). School of Engineering and Applied Science, University of Virginia. Charlottesville, VA.

Bruzon, K. (2022). Strong Data Privacy legislation SCOT model. [Figure 4]. *Prospectus* (Unpublished undergraduate thesis). School of Engineering and Applied Science, University of Virginia. Charlottesville, VA.

Bruzon, K. (2022). Weak Data Privacy legislation SCOT model. [Figure 5]. *Prospectus* (Unpublished undergraduate thesis). School of Engineering and Applied Science, University of Virginia. Charlottesville, VA.

Carlson, B. (2009). [SCOT figure description and example]. *Class handout* (Unpublished). School of Engineering and Applied Science, University of Virginia. Charlottesville, VA.

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

Hardy, K. (2020). Australia's encryption laws: practical need or political strategy? *INTERNET POLICY REVIEW*, 9(3, SI). doi:10.14763/2020.3.1493

Jacobs, B., & Popma, J. (2019, January). Medical research, Big Data and the need for privacy by design. *Big Data & Society*, 6(1), 205395171882435. https://doi.org/10.1177/2053951718824352

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

Oranchak, D. (2008). Evolutionary algorithm for decryption of monoalphabetic homophonic substitution ciphers encoded as constraint satisfaction problems. *Proceedings of the 10th Annual Conference on Genetic and Evolutionary Computation*, 1717–1718. doi:10.1145/1389095.1389425

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

Severson, D. (2017). The encryption debate in Europe. *Hoover Institution Aegis Paper Series*, (1702).

V. Magesh Babu, T. Shankar Ganesh, K. Ramraj (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

Yang, F., & Xu, J. (2018). Privacy concerns in China's smart city campaign: The deficit of China's Cybersecurity Law. *ASIA & THE PACIFIC POLICY STUDIES*, 5(3, SI), 533–543. doi:10.1002/app5.246