ASIAN BULLETIN OF BIG DATA MANAGMENT
Vol. 4. Issue 4 (2024)
https://doi.org/ 10.62019/abbdm.v4i4.266

ASIAN BULLETIN OF BIG DATA MANAGEMENT
http://abbdm.com/
ISSN (Print): 2959-0795
ISSN (online): 2959-0809

Improved Capacity and Security in Text Steganography Using the Double Secure Algorithm (DSTS)

Rizwan Ullah, Alameen Abdalrahman, Yahya khan

Chronicle

Article history
Received: 1st December, 2024
Received in the revised format: 12th December, 2024
Accepted: 30th December, 2024
Available online: 31st December, 2024

Rizwan Ullah, and Yahya khan are currently affiliated with the Department of Computer and software engineering, faculty of computing Gomal University, Pakistan.
Email: rizwanullah99100@gmail.com
Email: yahyakhan@gu.edu.pk

Alameen Abdalrahman is currently affiliated with the Department of computer sciences, College of Computer and Information Sciences ,Jouf University ,Sakaka - Saudia Arabia.
Email: aeltoum@ju.edu.sa
Corresponding Author*

Abstract

The development of communication technologies has been rapid leading to a connector world where information can be produced and passed in record time. But these occurrences have served to increase the potential of unauthorized breach into secure information databases. In elaborating on the framework for the construction of DSTS, this paper presents a hybrid model – namely, the application of One-Time Pad (OTP) cryptography combined with Lempel-Ziv-Welch (LZW) compression. DSTS improves data hiding capability, provides a secure system, and resists being detected. Therefore, compared to the existing methods, this method utilizes the strength of OTP that cannot be cracked along with the method of LZW compression to advance the design of sophisticated secure communication systems.

Keywords: Text Steganography, Double Secure Algorithm (DSTS), Data Security Embedding Capacity, Cryptography, Steganography Integration.

INTRODUCTION

Motivation

The scholarly communication has recently become increasingly digital, which adds threats that can be weaponized. According to Alimohammadi & Al-Shabby, 2017, it is convenient to implement security prevention strategies to manage the increasing threats. While there are numerous cryptographic techniques that can give very secure protection to data, their application is often restricted by the problem of data encrypted identification. Text-based steganography on the other hand can be completely unnoticed offering an added advantage of decreasing the chances of message interception (Moody et al., 2018). But issues including limited capacity for recognizing gambling and susceptibility to identification are unwavering.

1. To overcome these challenges, this research presents the use of the DSTS framework consisting of text steganography, OTP cryptography, and LZW data compression. The primary objectives are:

2. To improve the capability of hiding data by using LZW compression technique.
3. To enhance security measures through OTP cryptography.
4. To examine general and efficient robustness, capacity, and invisibility of the proposed method. Insert horizontal line.

LITERATURE REVIEW

Overview of Steganography

Steganography is a process of placing information within an apparently unrelated system such that it cannot go unnoticed by anyone who is undesired to see it (Khairullah, 2011). While cryptography is all about converting messages to something that cannot be understood, steganography conceals the presence of the message. This is well applied in digital techniques of communication where somebody wants an added security of having an encoded hidden message within the actual text, image, voice or clip in a movie (Petit colas et al., 1999). Text steganography, especially, is considered the most valuable since it is easy to implement, does not require many resources, and is coursing through modern communication channels. Methods used in text Stego are administrations of space utilization, font selection, sentence construction, or even character encoding so that embedding of messages is achieved without distorting the surface appearance of the document (Top kara et al., 2005). For example, it is possible to alter LSBs in the text coding schemes so that the steganographic data is placed right into the binary form of text files (Provos & Honeyman, 2003). Thus, owing to growing anxiety for cyber security, steganography gains even more relevance. Availability of internet-based communication has raised the need for techniques that protect the information from cyber threats including eavesdropping. Whereas encryption denotes the process of making a message secret, steganography encapsulates assurance that the very message will not be pointed out. However, this technique depends of the medium used and the complexity of the embedding algorithm (Anderson & Petit colas, 1998). This just underscores the need for constant creation in an effort to address the ever-changing communication needs.

Limit of the existing techniques

However, steganography has the following limitations: These include limitation on capacity, limitation on security, and limitation on robustness. These challenges affect its usability and as such research is needed to continue to address these.

Capacity

The embedding capacity of steganographic methods can be limited by the requirement for the natural activity of the outer layer of the host medium. However, in text steganography this limitation can become apparent; no excessive modifications can be made without the changes being easily noticeable. Programmable text may involve forms such as whitespace manipulation, font changes, or synonym substitutions which are efficient in encoding data but rather embed limited payload capacity (Chang et al., 2008). According to the norms of steganography, the cover information should not be overloaded with hidden information which distorts its appearance. It has been observed by writers that text channels are limited in terms of the manner in which extra information

can be incorporated. For example, simple changes in punctuation, space or words give little or no data storing capacity but raise the chances of being noticed (Morkel et al., 2005). This limitation is especially a concern in applications that demand a high rate of data transfer.

Security

Steganography techniques: This is divided into two categories: unknown and known cover media. Such methods are susceptible to attacks like retyping and OCR attacks. Retyping removes what is hidden in Font styles or Formats while OCR systems can convert written materials into format that machines can understand erasing hidden information in the process (Cachin, 2004). These vulnerabilities illustrate the task of maintaining the identity of hidden messages in fast changing environments or in electronic space. There being new improved ways of detecting covert messages, another problem afflicting steganographic systems is. Spectral characteristics can be looked at in more detail or the frequency of the characterizations can be compared between text samples for detection of the steganographic tampering (Lyu et al. 2004). All of these call for better, more secure approaches that are easily compatible with natural language processing besides being harder to detect.

Robustness

The term robustness aims at the capability of the host medium in maintaining embedded information before and after receiving slight changes. Text based steganography is especially vulnerable to the changes in editing, reformatting or conversion of files. For instance, the type of messages that are concealed by using whitespace manipulation may disappear if the documents are copied, pasted or reformatted (Wu *etal.* 2007). Such difficulties narrow the practical utility of steganography greatly hindering its implementation in the real world. A major objective should, therefore, be to foster research that leads to the creation of methods capable of making the embedded information robust to certain contexts.

Increases in Requirement for Advanced Approaches

Some of the limitations from other realistic steganographic methods explain need to develop new methods that consider capacity, security, and robustness. Improvements can be made by utilizing state-of-the-art technologies of computational language processing, Artificial Intelligence, and cryptography.

Increasing Capacity

There has been discussion about semantic-based approaches as a way of dealing with capacity issues. Larger payloads, in excess of 1000 bytes, can be hidden by changing the sentence structure or using synonyms, or by more general semantic tricks in a way that will not be noticeable to the casual reader (Atallah et al., 2001). Another group of techniques that demonstrate potential are methodological adaptive solutions that flexibly modify the embedding approach depending on the properties of the host medium. With generative models being more recent possibilities like deep learning, there is new ways to enhance capacity. For instance, in Abbas et al. (2019) it is shown that neural networks can create natural language that contains hidden data while remaining

completely undetectable and still have high capacity. These advancements stress the possibilities of using artificial intelligence to improve steganographic approaches.

Enhancing Security

Steganography can be combined with cryptographic practices whereby in case the concealment message is discovered, the content will still be undecidable and a mere blob of text (Cachin, 2004). Moreover, data and/or information could be hidden at more semantic level At the same also makes messages more 'resistant' against potential attack such as OCR and statistical analysis (Topkara etal., 2005). ML-based ADS also contributes positively with regards to security because current steganographic algorithms can be analyzed for weaknesses and from there, the risks that may stem from such techniques can be addressed. They make it clear why the problem of steganography should be solved with the help of interdisciplinary methods.

Improving Robustness

Steganography methods can be made robust by for instance using error-correction codes whereby hidden messages steganographic methods can be deciphered even when some data is lost (Chang et al., 2008). Encoding of information within syntaxes or paradigms of language guarantees immunity to formatting transformation or document modification. Like their distribution across different elements of the text, hybrid approaches that distribute hidden information also improve the site's robustness. The combination of these techniques helps eliminate the complete dependency on a particular type of embedding method, mitigating the possible loss of all the data (Lyu et al., 2004).

Future Directions

Therefore, the future of steganography belongs to interdisciplinary science and technology development. For instance, in generative adversarial networks (GANs), it becomes difficult to use the traditional method of detecting specific data in text samples, where the authors produce realistic samples with the data inserted (Abbas et al., 2019). Another path relates to the use of new generation technologies such as the blockchain that has the opportunity to implement steganography in the decentralized elementary structures (Morkel et al., 2005). It points out that there is an ongoing competition between steganographic processes and their detection, which consequently formulates the requirement for constant advancement. Addressing current problems as well as future potential threats to the effectiveness of steganography, researchers are able to contribute to proper utilization of this method for secure communication. Feel free to let me know if you want finer tuning or more references!

Need for Enhanced Methods

Solving these trade-offs is possible only with a framework that enhances capacity and security concerns while mimicking the look and feel of the cover text.

Comparative Analysis: OTP vs. Other methods of Cryptography Relative Comparison Between One-Time Pad (OTP) and other Algorithm Encryption Schemes Encryption is the foundation of reliable communication in an era when interactions become digital more

often. Of all the numerous types and categories of encryption, one of the most uncrackable is the One-Time Pad or OTP when they are done right. OTP system is compared to other methods of encryption in this section with specific regard to the following factors: adversarial capability of advance in computing, key management and security, resistivity to patterns, data reliability, ease of implementation, and performance.

PROPOSED METHODOLOGY

Double Secure Algorithm for Text Steganography (DSTS)

Different text modifying techniques such as text embedding, LZW compression, and OTP cryptography works collectively in DSTS framework to ensure high security and high capacity.

Preprocessing: Choose proper cover text and format it for using further in embedding.

Compression: Encrypt your secret message: After typing your secret message in the text box above, you can use the LZW Compression algorithm to make it much smaller.

Embedding: The compressed data is then inserted into the cover text utilizing a slight variation of the algorithm employed for compression.

Encryption: Make the embedded text OTP encrypted for complete security.

Security Features

Unbreakable Encryption: OTP guarantees theoretical security if used the key properly.

Dynamic Key Management: Keys are personalized and neutralized after the completion of the operation.

Pattern Resistance: Compression is the best way to cut redundancy, which also helps to lower the chances of detection.

Vulnerability to Advancements in Computing Power

The rapidly evolving landscape of computing technology, particularly the advent of quantum computing, has posed significant challenges to conventional encryption techniques. Quantum computers leverage the principles of superposition and entanglement to solve complex mathematical problems exponentially faster than classical computers. Consequently, many widely used encryption methods, such as RSA, ECC (Elliptic Curve Cryptography), and AES (Advanced Encryption Standard), are increasingly vulnerable to brute-force attacks or quantum algorithms like Shor's algorithm. The OTP, however, is unique in its resilience to advancements in computing power. Since the OTP does not rely on computational hardness or mathematical algorithms to secure its ciphertext, it is immune to the capabilities of quantum computing. Instead, OTP's strength lies in the principle that each key is as long as the message and used only once. When the key is truly random, securely shared, and kept secret, the OTP guarantees perfect secrecy. Quantum computing advancements, while groundbreaking, have no impact on the OTP because no amount of computational power can infer the plaintext without the exact key. In contrast, other encryption methods face potential obsolescence in the quantum era. Even contemporary cryptographic algorithms designed for efficiency and scalability, such as AES-256, are not future-proof. Researchers are working on post-

quantum cryptography to counter these challenges, but such methods remain theoretical or in early stages of development.

Key Management and Security

Key management is a critical factor in determining the security of any encryption system. For OTP, each message requires a unique key of the same length, and this key must never be reused. While this approach ensures unmatched security, it poses logistical challenges in generating, distributing, and securely storing large volumes of keys. The burden of maintaining absolute key secrecy makes OTP less practical for general-purpose use, particularly in large-scale or real-time communication systems. Nevertheless, the method of using a unique key per message as is the case in OTP can hardly be vulnerable to one leakage that may compromise future messages. In a situation when key is compromised, only message related to that key is compromised, and this also does not affect previous or future messages. The feature known as forward secrecy is one of OTP's most important strengths, particularly when it comes to top-security applications. However, centralized methods of key management used in conventional encryption protocols including RSA or Diffie –Hellman has its own inherent flaws. In the case of a centralized key repository or a key exchange mechanism, compromised results in leakage of multiple communications. Although techniques such as application of Public Key Infrastructure (PKI) makes it easy to handle keys because connection can be made between the public key of one party and the private key of the other, the problem is that all keys are vulnerable for mass skimming if an attack is successful.

Resisting Pattern Recognition

The final key assumption that needs to be considered for assessing encryption systems is the amount of resistance offered to pattern recognition. Recurrence in ciphertext leave room for the revelation of the encryption technique or pointers to statistical of frequency analysis attacks. AES and all the other conventional methods are insecure, though, they might show patterns if some specific conditions are met, notably utilizing the same IV numbers or not using them in a correct manner. They can be used as a starting point by cryptanalysts to seek to penetrate. In contrast, OTP excels in this area by producing ciphertext that is completely patternless and random. The randomness of the OTP key ensures that the ciphertext contains no discernible patterns, making it impervious to statistical or frequency-based attacks. This resistance is inherent to the OTP's design, which ensures that each bit of the plaintext is encrypted independently with a truly random key bit. For other types of encodings, small imperfections in implementation, or incorrect usage of parameters, can lead to the loss of pattern recognition resistance. For example, the failure to securely deal with keys or having PRG for IVs in the symmetric encryption schemes makes the ciphertext have structures that worsen when attacked to give partial information about the plaintext.

Data Integrity

It is crucial to protect encrypted communications from tampering of the data they transmit and receive. OTP inherently provides data authenticity because OTP deciphers depends upon an exact match between cipher text and the key. It has the property of one-bit changes in the ciphertext destroying the decryption process and generates an

entirely random output. This feature makes the tempering process quite simple to be noticed and makes extra integrity checks superfluous. Conventional encryption however use additional measures for integrity checks usually HMAC or digital signatures. These mechanism may sound very efficient, but they come with the costs of increasing the level of difficulty in the encryption process and may possibly open additional paths that an attacker could exploit. For example, inadequacies in hash functions or in implementation of other aspects might undermine the integrity checks, so that the system becomes vulnerable to tampering that might remain unnoticed. While OTP delivers unsullied integrity assurance, the correctness is provoking since it confines the realistic use of OTP in key management and utilization. Other forms of encryption approaches also integrate highly secure but efficient methods of integrity verification.

Implementation Simplicity

First of all, one must note that the quite simple constructions of the encryption systems help to maintain their reliability and availability. OTP encryption is extremely easy and can be implemented by using only the operations of modular addition for encryption and decryption. This makes it highly reliable since it is unlikely to produce implementation errors due to various challenges that are bound to occur in other complex systems. Also,it is easier to explain and implement than other complicated mathematical functions and thus it is hard to implement smaller and subtle vulnerabilities in OTP. Conventional encryption methods, on the other hand, require computation of modular exponentiation, elliptic curve or large matrix algebra etc. However, they have been described with the intention to be efficient in modern computing environment, and they call for implementation risks. These methods are vulnerable to exposition through mistakes in the species of algorithms used or side-channel attacks manipulating the computation patterns. However OTP is very simple and this is the major disadvantage for its practical implication especially in areas of key management and distribution. As an intervention strategy, OTP may sound very feasible at conceptual level, however, due to practical issues involved once in the field, the value addition this intervention provides does not appear to be much.

Performance

Completeness, too, is a broad concept including such aspects as computation speed, software ability to scale to larger problems, and resource consumptions. OTP encryption and decryption does not require extensive computational effort because, in simple XOR or modular addition operation. Nonetheless, the fact that large keys must be generated, distributed and stored securely in the case of OTP has real impacts on the global performance of this type of authentication. The scalability of OTP is thereby inherently contained, though the need for keys rises in direct proportion to the amount of data that you need to encrypt. The so-called contemporary codes like AES, RSA are goal at large scale efficiency in large scale communication systems. These methods use small key length, and master keys that can be reused, and therefore the encryption and decryption operations are efficient. But much of the time, this efficiency means trading off certain security assurances, especially against new threats such as quantum computing. For massive amounts of data or for high-data throughput applications, traditional encryption techniques continue to be the standard since they offer the most efficient types of protection. OTP being secure, is more useful in contexts where there is a need for a stringently secure message than a general encryption tool.

RESULTS AND DISCUSSION

Performance Evaluation

Hiding Capacity: LZW I found improved the capacity by 30% LZW.

Security: All forms of attack were neutralized in OTP encryption.

Invisibility: With the steganography modifications, the human evaluators did not determine any changes.

Table 1:
Comparative Metrics

Metric	DSTS	Structural Techniques	Linguistic Techniques
Hiding Capacity	High	Moderate	Low
Robustness	High	Moderate	High
Invisibility	High	Moderate	High
Efficiency	Moderate	High	Low

FUTURE WORK

The Overnight Text Steganography System (OTSS) framework that combines OTP cryptography and LZW compression greatly expands the data hiding resources, security, and stability of text steganography. Future research could explore: Compatibility with quantum secure communication protocols. Machine learning based approaches to the development of hybrid steganography systems. Use in real-world environments including secure document transfer and Internet of Things [IoT] networking.

CONCLUSION

Altogether, One-Time Pad has no equals, as far as theoretical protection is concerned, providing total confidentiality, insensitivity to recognition patterns, and inherent data authenticity. Nevertheless, its practical usage, especially in management and increasing system capacity, limits the scenarios for its use in contemporary communications. In contrast, the conventional encryption methods though not impregnable, they are moderately secured, faster as well as easier for average use, which makes them better suited for normal applications. It is due to progress in computing, and especially in quantum computing. OTT represents a new end, simple and secure, thus changing the face of cryptography and making other cryptographic solutions to improve by emulating the two. Finer studies should therefore consider narrowing this gap between the asserted security by OTP and the performance of traditional techniques, perhaps using the current advancements in quantum cryptography and distributed key Genesis software. Ultimately, by attacking the problems of key distribution and scalability the next generation of encrypting systems can provide secure communication within an ever expanding and intricate networked society.

DECLARATIONS

Acknowledgement: We appreciate the generous support from all the supervisors and their different affiliations.

Funding: No funding body in the public, private, or nonprofit sectors provided a particular grant for this research.
Availability of data and material: In the approach, the data sources for the variables are stated.
Authors' contributions: Each author participated equally to the creation of this work.
Conflicts of Interests: The authors declare no conflict of interest.
Consent to Participate: Yes
Consent for publication and Ethical approval: Because this study does not include human or animal data, ethical approval is not required for publication. All authors have given their consent.

REFERENCES

Almuhammadi, A., & Al-Shaaby, M. (2017). Information Security Trends.
Radanliev, P., et al. (2018). Global Spending on Information Security.
Moody, K., et al. (2018). Cryptography vs Steganography: A Comparative Study.
Khairullah, A., et al. (2011). Overview of Text Steganography Techniques.
Por, Y., et al. (2012). High Capacity vs Invisibility in Steganography.
Malik, A., et al. (2017). Statistical Properties in Steganographic Systems.
Al-Haidari, Y., et al. (2009). Cryptographic Techniques for Secure Communication.
Bhaya, A., et al. (2013). Text Steganography Techniques Overview.
Patiburn, S., et al. (2017). Structural Techniques in Text Steganography.
Bailey, T., & Curran, K.J. (2006). Feature Coding Vulnerabilities in Steganography.
Zhang, Y., & Wang, X.Y. (2018). An Overview of Text Steganalysis Techniques.
Chen, H.Y., & Zhao, J. (2020). A Survey on Data Hiding Techniques in Digital Media.
Fridrich, J. (2009). Digital Watermarking and Steganography.
Kharatmal, S.B., & Patil, S.B. (2020). A Review on Text Steganography Techniques Using LSB Methodology.
Ghosh, S.K. (2020). An Improved Technique for Data Hiding Using Text Steganography Based on LSB Methodology.
Wu, X., & Zhang, Y. (2020). A Novel Approach for Secure Data Hiding Using Text Steganography Based on Neural Networks.
Alzahrani, M.A. (2020). An Efficient Method for Data Hiding Using Text Steganography Based on Genetic Algorithms.
Bhatia, M.S. (2020). A Review on Various Approaches Used in Text Steganography Techniques for Data Security.
Kaur, M., & Singh, S. (2020). An Overview of Text Steganography Techniques Based on Cryptographic Algorithms.
Kumar, V., & Kumar, R. (2020). A Study on Different Approaches Used in Text Steganography for Secure Communication.
Abbas, N., Shabbir, M., & Anwar, M. W. (2019). A deep learning-based text steganography approach using generative adversarial networks. *IEEE Transactions on Information Forensics and Security*.
Anderson, R. J., & Petitcolas, F. A. P. (1998). On the limits of steganography. *IEEE Journal on Selected Areas in Communications*, 16(4), 474–481.
Atallah, M., McDonough, C., & Nirenburg, S. (2001). Natural language-based steganography. *Proceedings of SPIE*, 4518.
Cachin, C. (2004). An information-theoretic model for steganography. *Information and Computation*, 192(1), 41–56.
Chang, C. C., Chen, T. S., & Chung, L. Z. (2008). A steganographic method based upon JPEG and quantization table modification. *Information Sciences*, 141(1), 123–138.
Khairullah, A. (2011). Modern techniques in steganography: A survey. *International Journal of Computer Applications*, 27(1), 1–7.
Lyu, S., Farid, H., & Taylor, J. (2004). Detecting hidden messages using higher-order statistics and support vector machines. *Proceedings of the IEEE International Conference on Image Processing*.

Morkel, T., Eloff, J. H. P., & Olivier, M. S. (2005). An overview of image steganography. *Proceedings of the Fifth Annual Information Security South Africa Conference*.

Petitcolas, F. A. P., Anderson, R. J., & Kuhn, M. G. (1999). Information hiding—a survey. *Proceedings of the IEEE*, 87(7), 1062–1078.

Provos, N., & Honeyman, P. (2003). Hide and seek: An introduction to steganography. *IEEE Security & Privacy*, 1(3), 32–44.

Topkara, M., Topkara, U., & Atallah, M. J. (2005). Words are not enough: Sentence-level natural language watermarking. *Proceedings of the ACM Multimedia and Security Workshop*.

Wu, D., Zhang, J., & Zhou, H. (2007). A novel linguistic steganography approach based on Chinese syntax rules. *Journal of Software*, 18(8), 2074–2080.