Grenze International Journal of Engineering and Technology, January Issue

Covert Data Intelligence: Advance Data Hiding Techniques

Ayush Ther[1], Soham Sawalakhe[2], Ayush Shirbhate[3], Bhushan Onkar[4] and Dhiraj Shirbhate[5]
[1-5]Government College of Engineering, Computer Engineering, Yavatmal, India
Email: ayushthercse@gmail.com, sohamsawalakhecse@gmail.com, ayushshirbhtecse@gmail.com, bhushanonkarcse@gmail.com, dhirajshirbhatecse@gmail.com

***Abstract*—— In the recently developed malware, new methods have been adopted to conceal malicious code in the documents due to which it has become difficult to detect malware with traditional security systems. The proposed research work presents a new data-hiding technique based on PVD.**
It will take advantage of the document structure to embed secret messages. A document-structure-based embedding and extraction algorithm presented in this paper can effectively hide and retrieve data in an undetectable and secure manner. Experimental results illustrate that our approach can embed data seamlessly into documents without visible distortion, meaning the integrity of the cover document remains intact. In particular, the method would be robust against typical detection mechanisms operating in cybersecurity. Data hiding techniques play a pivotal role in covert communication, where information needs to be concealed to ensure privacy and security.
This paper explores different data hiding techniques, their methodologies, and comparative effectiveness in terms of security, capacity, and imperceptibility. Key techniques discussed include steganography, cryptography, watermarking, and innovative methods like Pixel-Value Differencing (PVD). The paper also highlights recent advances in document-based data concealment and their applications in cybersecurity.

***Index Terms*— Data Hiding, Steganography, Pixel-Value Differencing (PVD), Document Security, Malware, etc.**

I. INTRODUCTION

Data hiding is one of the ways to secure sensitive information by hiding it inside other data, which is apparently innocuous. This makes the data be embedded or concealed in such a way that the unsuitable users do not easily find or access it. In this way, such important information is kept safe from unwanted eyes, yet it is available for authorized users through techniques like encryption, steganography, and obfuscation. This process is widely used in cybersecurity, digital forensics, and the protection of priva-cy, especially because the integrity and confidentiality of data are completely protected [1].
The rapid evolution of digital communication and information exchange has led to a parallel increase in the need for secure communication channels. Traditional cryptography focuses on securing the content by en-crypting it, but data hiding aims to conceal the very existence of the information. This hidden infor-mation can be safely transmitted across seemingly innocuous carriers, such as images, audio files, or documents, without drawing attention to itself.

Grenze ID: 01.GIJET.11.1.243

A. Problem Statement

While digital information continues to rise, techniques of data-hiding have become increasingly com-mon-not only for legitimate uses, such as privacy protection, but also for malicious applications in-volving concealing illicit content or evading forensic detection. File systems such as FAT and NTFS afford different opportunities in terms of data hiding through the use of techniques like hidden files, deleted files, alternate data streams, utilization of slack space, and steganography. The challenge re-mains with digital forensic investigators to effectively detect and analyze these hidden data within complex file systems. The current study tends to probe, compare, and assess different methods of data hiding in both the FAT and NTFS file systems, as well as the abilities of forensic tools in finding out these hidden data, in order to meet this pressing demand in digital forensics.

B. Overview

Data-hiding techniques take advantage of structural features of file systems to hide information in a manner that is not normally visible under regular, day-to-day operations. The techniques include the rather simplistic hiding of files and folders, advanced manipulation of slack space, and embedding of data in the so-called alternate data streams in file systems like FAT and NTFS. In the technique that is becoming increasingly sophisticated, the difficulty in detection similarly develops, hence presenting enormous challenges for a digital forensic investigator. The study has summarized a few usual techniques of data hiding and discussed the implementation in FAT and NTFS, concluding with the effectiveness of different forensic tools in detection of the hidden data [2]. This will definitely help forensic analysts to further develop a better approach toward detection and improve protection from data hiding in cybercrime investigations effectively.

II. Literature Review

Since the beginning of digital forensic research, various researchers have discussed data-hiding tech-niques of file systems. Carrier (2005) presented an overview of the basic knowledge with regard to file system forensic analysis, including file structure and the possibility of data hiding in FAT and NTFS file systems. Davis et al. (2005) discussed several hacking techniques, such as the concealment of files and partitions, and their forensic consequences. Shu-fen et al. (2009) restrict their discussion to the FAT file system-based file hiding and further discuss the methods and difficulties concerning hidden data detection.

Kessler (2004) provides a very enlightening discussion concerning steganography, which is a high-level method for concealing data within the files of multimedia. Since the changes are highly subtle, there are difficulties regarding their detection. These works together point toward the challenge of hidden data detection in file systems and the further need for new forensic tools and methodologies in order to stay current with evolving techniques for hiding data.

Recent studies have further built on these themes to discuss modern techniques for uncovering hidden data. In this regard, Liu et al. (2018) studied ext4 and NTFS file systems with vulnerabilities in modern times, proposing an enhanced tool for the detection of forensic investigators. Lin and Chang (2021) introduced machine learning models for finding steganographic patterns within multimedia files with a dramatic increase in the detection rates. This collection of works focuses on the increasingly sophisticated development of data-hiding techniques and the urgency for novel forensic tools and methodologies

A. Information Hiding: Definition And Importance

Information Hiding is the art and science of hiding information in other, innocuous-appearing data or structures. The very concept plays a vital role in protecting sensitive information to the extent of dis-guising an illicit activity from being detected. In the context of digital forensic investigation, under-standing and finding hidden data is crucial in cybercrime investigations involving sophisticated techniques to cover up evidence. Accompanying the increased importance of information hiding is its dual use to satisfy both benign purposes, such as the protection of privacy, and malicious, such as for avoiding forensic investigators.

B. Legitimate and Malicious uses of Information Hiding

Hiding information can be a perfectly legal example if it protects proprietary information, guarantees privacy in communication, or serves security goals. Indeed, quite a big part of information protection schemes relies on the use of cryptographic algorithms or steganographic methods to preserve information against unauthorized access. However, information hiding is also used maliciously-for example, by cybercriminals to embed malicious payloads in seemingly innocent files, hide stolen data, or actually escape detection during forensic investigations.

The knowledge of these dual uses has important implications for the elaboration of efficient forensic strategies, since they have to make a dis-tinction between legitimate uses of hidden data and malicious ones [3].

C. Significance in Digital Forensics

Detection and analysis of hidden data are very important aspects to unearth the evidences in digital forensics, which could be hidden and remain undisclosed. A forensic analyst must be versed and pre-pared in different ways of hiding data to conduct a reconstruction of an event, identification of the re-sponsible party, and to not miss any evidence. Advanced file systems, such as NTFS with features like ADS, add many pitfalls in the way of forensic tools and methodologies, making the latter ever-changing.

D. Purpose of the Study

The main objective of this study is to present and compare different data hiding methods in the FAT and NTFS file systems. By analyzing the implementation of these methods and how they can be detected using forensic tools, the study will give insight into current forensic technologies' capability and point out where further enhancement is needed. These help in understanding various techniques, thereby helping to improve forensic practices and increasing the capability of detection in the real world.

III. File System Overview

A. Overview of FAT and NTFS

FAT is one of the older file systems that has been widely employed on many platforms such as MS-DOS and early versions of Windows. Due to its simplicity and compatibility, FAT is a common choice for removable storage devices such as USB drives and memory cards. The FAT file system uses a relatively simple data structure: the File Allocation Table keeps track of the location of files on disk. The directo-ry entries provide basic file information like name and size, but also starting cluster [4]. NTFS is the more advanced file system developed by Microsoft, and most of the modern versions of Windows OS are using it. NTFS provides additional security, as through encryption of files and ACLs, along with supporting large volumes and sizes of files. As in all file systems, the most important data structure in NTFS is the Master File Table, which contains one record for every file and directory on the volume, including metadata. NTFS also hosts attributes, including Alternate Data Streams (ADS), where com-plex data can be hidden inside a single file.

B. FAT and NTFS Comparison

The richness of NTFS over FAT bears essential implication w.r.t. data hiding. The fact that FAT is simple makes it easy to understand and work on; however, that equally means the data structures are also sim-ple, providing very few opportunities for sophisticated data hiding techniques. On the other hand, NTFS enhanced capabilities, such as ADS and custom-defined attributes, increase possibilities for con-cealment but simultaneously make it harder to detect. Forensic tools should consider such difference while doing forensic analysis on a file system in order to detect hidden information.

IV. Data Hiding Techniques Overview

A. Steganography

Steganography is the practice of hiding information within other, seemingly innocent, digital objects like images, audio, video, or text.Table I It ensures the communication of a secret message without re-vealing that a message is being sent. The most common forms of steganography include:

TABLE I. DATA HIDING TECHNIQUE OVERVIEW

Technique	Description	Advantages	Disadvantages
Steganography	Conceals data in carrier files such as images, audio, video, or text by altering minor features.	High imperceptibility and flexibility	Vulnerable to detection by advanced forensic techniques.
Encryption	Encrypts data in unreadable form using cryptographic algorithms	Strong security and confidentiality	Key management and decryption process
Obfuscation	Conceals data through renaming files, altering extensions, or modifying metadata	Simple, effective, and quick	Easily detectable by heuristic or forensic tools.
Alternate Data Streams (ADS)	Hides data within unused streams of NTFS file systems, other than standard file data	High capacity and difficult to detect manually	Detection requires specialized tools and knowledge.

File Splitting	Divides a file into smaller parts to obscure its original purpose and structure.	Makes identifying the original data challenging	Requires reconstruction and may leave detectable patterns.
Compression Hiding	Embeds data into unused portions of compressed files, especially in lossless formats.	Data remains intact and retrievable	Limited embedding capacity; lossy compression may lose data.

B. Cryptography and Data Hiding

Though cryptography does not directly hide data, it plays a critical role in concealing the meaning of information. When combined with steganography, cryptography ensures the security of the content, even if its existence is detected. For example:

- RSA and AES Encryption: Cryptographic algorithms such as RSA and AES are frequently used to en-crypt messages before they are embedded into a carrier medium. This enhances security by providing multiple layers of protection.

C. Watermarking

Digital watermarking is a form of data hiding that embeds information into a digital object, primarily for authentication and ownership purposes. Watermarking can be classified into two types:

- Visible Watermarking: Typically used for copyright protection, this technique embeds a visible wa-termark into an image or video.
- Invisible Watermarking: This is a more secure method that inserts a hidden watermark, which can later be extracted or detected to prove ownership or authenticity.

Watermarking is widely used in multimedia applications and is resilient against common forms of im-age and audio manipulation, such as resizing or compression.

D. Document-Based Data Hiding

An emerging area of interest is the concealment of data within documents. Fig 1 This technique utilizes the structure and content of text documents, such as font styles, paragraph formatting, and spaces, to embed secret information.

Methods include

- Text Steganography: Hides information within text by altering words or lines, using whitespace en-coding, or modifying the font's attributes (such as size, style, or boldness).
- PDF and Word Document Embedding: Leveraging the metadata and structural elements of file for-mats like PDFs or Word documents to conceal data without affecting the visible content of the docu-ment.

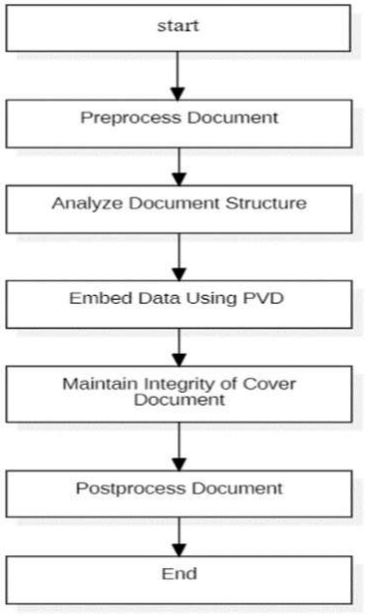

Fig.1 Database Hiding using PVD flowchart

V. Novel Technology and Detailed Analysis of Hiding Techniques

A. Hidden Files and Folders

- Description: Fig 2 Hidden files and directories result from setting file attributes that make them in-visible to normal directory viewing. This is the simplest way of hiding data. In fact, users and mal-ware apply this method to keep files away from casual view.
- Forensic Detection: Hidden files can be detected very easily with forensic tools simply by scanning for the file properties to be set as hidden. This technique usually has limited effectiveness in any se-rious attempt at concealment, as most forensic tools and even operating system settings can expose hidden files with little or no effort.

B. Deleted Files

- Definition: What the file system of the operating system generally does to a file in deleting it is to keep its directory entry and mark it as deleted, thereby releasing that space of the said file for new data to be placed. The data previously in that file remains on the disk until it is overwritten by an-other file. Basically, this defines why "deleted" files can typically be recovered.
- Forensic Detection: The forensic tools are able to determine unallocated file entries in the FAT or MFT. Based on these, the tools can recreate and recover the deleted files unless the data clusters have been written over. The recovery process is quite essential in forensic investigation because critical evidence is just what can be recovered from the deleted files [5].

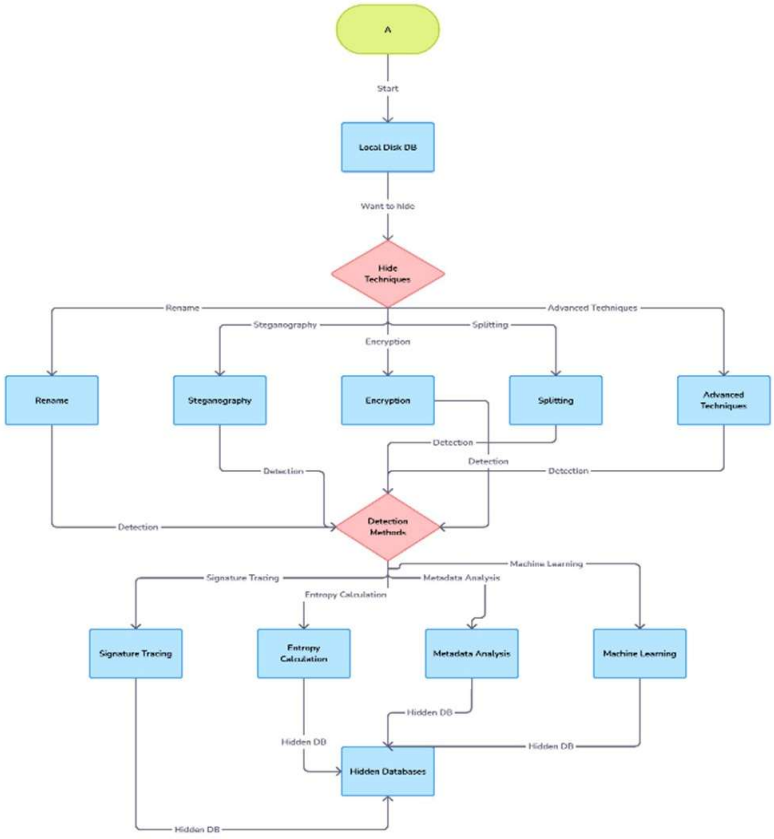

Fig 2. Different local disk techniques for hiding

C. Hidden/Deleted Partitions

- Hiding Partitions: Hidden partitions belong to the kind of partition that may be hidden from the op-erating system by changing the ID bits in the partition table. This is quite easily detectable by foren-sic tools, as the standardized methods for detection can show such hidden partitions.

- Deleting Partitions: Deleting a partition removes its entry from the partition table, but it does not remove the data stored inside the partition. Forensic tools can recover deleted partitions because the data stored inside them is still accessible when the disk is scanned for partition signatures [6].

D. Alternate Data Streams (ADS)

- Explanation: Alternate Data Streams are part of NTFS's design that allow numerous data streams to be associated with a single file. These streams do not show up in normal directory listings and thus, provide a powerful means of hiding data in a file without increasing it's apparent size or changing the nature of the content.
- Forensic Detection: Present forensic tools have the capability to detect and display ADS, by explic-itly searching for them in the NTFS images. During NTFS file system investigation, the investigator is required to consider explicitly hidden streams and carry out ADS analysis as a standard inclusive way.

E. Slack Space and File Slack Space Hiding

- Slack Space: Slack space is that part of the disk cluster that is left unused when a file does not fully occupy an assigned cluster. These slack spaces would then contain residual data unknowingly of deleted, older files. This too can be recovered and scrutinized in any forensic analysis.
- File Slack Space Hiding: In this type of hiding, slack space-the area at the end of allocated clusters for a file-is the target area. This data will be placed there on purpose. Because different files must be used to store the hidden data, this approach makes detection more complex [7].
- Forensic Detection: The forensic tools can analyze disk images to find out hidden data within the slack space by observing difference in cluster usage and file-size varying characteristics. To detect the data hidden within the slack space, the analyst needs to be very careful because many of the times it is left-over data and content.

F. Bad Clusters

- Description: Bad clusters refer to the disk areas marked as unusable because errors are encoun-tered. Bad clusters can be tampered with, however, to make data more covert by intentionally mark-ing them as bad in the file allocation table, ensuring that the specific clusters are not read by the file system.
- Forensic Detection: Several forensic tools exist for detecting clusters marked as bad and allowing an investigator to verify their contents [8]. Though the problem here is how to differentiate between clusters that are actually bad and those that are deliberately marked just for concealing data. Re-covery is possible because of hidden data in these clusters an investigator needs to recover and de-crypt.

G. Steganography

- Definition: Steganography is the art of embedding data in multimedia files such as images, audio, or video by slightly modifying the file data in such a way that the human eye and ear do not sense the modified data [9]. For example, in an image, the least significant bits (LSB) of its pixel values may be slightly altered to carry information not visible in the cover.
- Forensic Detection: Detection of steganography indeed is difficult, and the same requires tools of reasonably high level constructed for statistical study and pattern recognition. Identification of file structural anomaly or pattern may even be done by forensic tools, but to recover the hidden infor-mation, knowledge on the specific steganographic used methodology will most probably be re-quired.

VI. Comparative Analysis of Techniques

Steganography, watermarking, and cryptography provide unique security solutions for data hiding and protection. Table II Where steganography attains high concealment, it is easily accessible to detection methods; watermarking focuses on ownership verification rather than data secrecy, and cryptography ensures data security even if detected.

TABLE II. COMPARATIVE ANALYSIS OF DATA HIDING TECHNIQUES

Technique	***Security***	***Capacity***	***Imperceptibility***	***Disadvantages***
Steganography	-Offers a high degree of concealment. - Susceptible to sophisticated steganalysis methods.	- Capacity varies with the method. - Methods such as PVD can hide large quantities in	- Generally high. - This depends on the method (for example, LSB is less secure).	- Statistical analysis can detect. - It is sometimes difficult to

		images		balance capacity and imperceptibility.
Watermarking	- Primarily for authentication and ownership determination. - Not designed to hide information from detection.	- Not storage-intensive. - Primarily for small pieces of information (e.g. ownership information).	- Normally more resistant to tampering. - Less concerned with hiding.	- Not ideal for large data hiding. - Designed for other uses like copyright protection.
Cryptography	- Provides that the content will not be compromised even if the data is discovered. - Discovery of encrypted data may bring some attention.	- Method dependent. - Not for hiding but securing data.	- N/A (Data security-oriented rather than hiding)	- Encrypted data attracts attention. - Does not hide the presence of data.
Pixel-Value Differencing (PVD)	- More secure than simple LSB. - More resilient to detection.	- Can embed more amount of data with less visible distortion.	- Better preserves the quality of the carrier image. - Hides the secret data better.	- More complex in implementation. - Capacity and security are hard to balance.
Least Significant Bit (LSB) Substitution	- Lower security. - Detectable through compression or statistical analysis.	- It has high capacity in terms of data bits that can be embedded.	- It has lower imperceptibility compared to PVD.	- The image can easily be detected if it is analysed or compressed
Audio/Video Steganography	- Variable security. - Methods like echo hiding and motion vector modification offer varied capacities.	- Much higher capacity due to the size and intricacy of audio/video files.	- Generally high. - Dependent on the method; more intricate media can better conceal data.	- Requires more complex algorithms. Higher processing requirements.

VII. Challenges and Future Directions

While data hiding techniques have proven effective, they are not without challenges. The increasing sophistication of steganalysis poses a significant threat to steganography, necessitating constant inno-vation. Additionally, trade-offs between capacity, security, and imperceptibility must be considered when selecting the appropriate technique for an application [10].

Future research may focus on devel-oping hybrid techniques combining steganography with other security measures such as cryptographic algorithms and blockchain technology. Advances in artificial intelligence could also lead to more ro-bust and dynamic data hiding methods.

VIII. Conclusion

Conclusively, the investigation into various methods of data hiding in FAT and NTFS file systems demonstrates how such techniques are developing due to modern forensic utilities. Despite the fact that some traditional methods are hidden and deleted files, up until now they easily appear to be disclosed by current forensic practices, while the more modern ones-Alternate Data Streams (ADS) and Ste-ganography-present their myriad problems of detection.

These are acceptable proofs as to why forensic methods and techniques have to be constantly developed so as to keep in pace with developing, more sophisticated, and complex strategies of hiding information. As the field of information security changes, the very balance of hidden data and the possibilities for detecting the hidden data will remain a future area of focus for both cybersecurity and digital forensics.

Acknowledgement

We would like to express our sincere appreciation to those who contributed to the completion of this review paper. Their support and assistance have been invaluable throughout the research and writing process. We extend our heartfelt thanks to Prof. Dhiraj D. Shirbhate for his guidance and made this re-search possible. His expertise and insights greatly enriched the content of this paper.

We are also grate-ful to Prof. Chetan V. Andhare for his Guidance & references. Their support played a crucial role in shaping the direction of our work.

Lastly, we would like to express our gratitude to our colleagues and peers who provided valuable feedback during the review process. Thank you all for your contributions and support.

REFERENCES

[1] R. Kumar, S. Chand, "A Reversible High Capacity. Data Hiding Scheme Using Pixel Value Adjusting Feature", Multimedia Tools and Applications, Springer, pp. 241-259,2016.

[2] A.M. Alattar, "Reversible Watermark Using the Difference Expansion of a Generalized Integer Transform", IEEE Transactions on Image Processing, IEEE, pp. 1147- 1156, 2004.

[3] N. F. Johnson, S. Jajodia, exploring steganography: seeing the unseen, Computer Practices (1998) 26-34.

[4] W.L. Tai, C.M. Yeh, and C.C. Chang, "Reversible Data Hiding Based on Histogram Modification of Pixel Differences", IEEE Transactions on Circuits and Systems for Video Technology, IEEE, pp. 906-910, 2009.

[5] D.M. Thodi, J.J. Rodriguez, "Expansion Embedding Techniques for Reversible Watermarking", IEEE Transac-tion on Image Processing, IEEE, pp. 721-730, 2007.

[6] X. Li, J. Li, B. Li, and B. Yang, "High-Fidelity Reversible Data Hiding Scheme Based on Pixel-Value-Ordering and Prediction-Error Expansion", Signal Process, Elsevier, pp. 198-205, 2013.

[7] F. Di, M. Zhang, X. Liao, and J. Liu, "High-Fidelity Reversible Data Hiding by Quadtree-based Pixel Value Ordering", Multimedia Tools and Applications, Springer, pp. 1–17, 2018.

[8] F. Peng, X. Li, and B. Yang. "Improved PVO-based Reversible Data Hiding", Digital Signal Process, Springer, pp. 255–265, 2014.

[9] X. Qul, H.J. Kim, "Pixel-based Pixel Value Ordering Predictor for High-fidelity Reversible Data Hiding, Signal Process, Elsevier, pp. 249–260, 2015.

[10] B. Ou, X. Li, Y. Zhao, and R. Ni, "Reversible Data Hiding Using Invariant Pixel-Value Ordering and Predic-tion-Error Expansion", Signal Process: Image Communication, Elsevier, pp. 760–772, 2014.