www.ijecs.in
*International Journal Of Engineering And Computer Science*
*Volume 14 Issue 09 September 2025, Page No. 27673-27679*
*ISSN: 2319-7242 DOI: 10.18535/ijecs/v14i09.5234*

# Modern Algorithms for Recovering Deleted Data in User and Enterprise Systems

**Stanislav Yermolov**

Founder and Lead Developer, East Imperial Soft Kyiv, Ukraine

**Abstract**

In this article analyze existing algorithms for recovery of deleted data in user and enterprise systems. Data loss — regardless of cause (unintentional deletion, software failure, clearing or formatting of storage media, malware activity) — can result in economic losses and undermine the reputation of organizations. The objective of the study is to provide a systematic review and comparative analysis of existing algorithmic approaches to resuspension of deleted file system objects in domestic and enterprise environments. This review takes into account the internal structure of the file system (cluster allocation tables, journaling mechanisms and metadata) and the specifics of different classes of storage media (conventional hard disk drive (HDD), solid-state drive (SSD) with various wear-leveling strategies, hybrid solutions etc.). Attention focuses on two fundamentally different methods: recovery based on analysis of file system artifacts (service records, indices and journals) and signature-based file carving aimed at searching the raw disk space for characteristic header and footer patterns of files. For each method a critical evaluation is performed according to several criteria: completeness of extraction, processing speed, robustness against fragmentation and false positives, as well as the impact of storage media characteristics (for example the operation of the TRIM command on SSD). Based on the identified limitations, hybrid algorithmic strategies are proposed, combining the responsiveness of service analysis with the depth of byte-level scanning — which allows achieving a balance between accuracy and efficiency.

The practical significance of the work is manifested in the formulation of specialized recommendations for information technology (IT) engineers, system administrators and software developers aimed at enhancing the reliability of data recovery procedures.

**Keywords:** data recovery, deleted files, file system, NTFS, APFS, Ext4, SSD, HDD, signature analysis, file carving, MFT, recovery algorithms.

## 1. Introduction

The digital era has elevated data to one of the key assets of the twenty-first century, rendering its generation, analysis and storage fundamental across all levels — from personal correspondence to large-scale business and governmental structures. Despite the rapid advancement of backup technologies and cloud services, the threat of data loss persists with undiminished severity.

The roots of information loss are diverse: ranging from banal accidental deletion (one of the most frequent scenarios) to targeted operations such as media reformatting, destruction of partition logical structures, hardware failures and cyberattacks, among which ransomware programs have assumed a prominent role [1]. For businesses, the consequences of data loss can translate into colossal financial losses due to downtime, non-compliance with regulations (such as GDPR) and reputational damage. For private individuals, such incidents often result in the irretrievable loss of unique digital archives — photographs, video recordings and documents — recovery of which is virtually impossible. The increasing reliance on digital assets renders the development of reliable recovery methodologies absolutely critical.

**The objective** of this study is to systematize and comprehensively analyze current algorithms for recovering deleted data in both consumer and enterprise environments, as well as to evaluate their performance according to loss scenarios and media type.
**The scientific novelty** of the approach lies in the development of a classification of recovery methods based on the level of their interaction with the "media — file system" components, which enables the identification of optimal strategies for diverse conditions.
**The author's hypothesis** posits that the most effective and universal approaches are hybrid algorithms that combine file system metadata analysis with intensive signature scanning, thereby ensuring the maximal volume of recovered data under any circumstances — from simple deletion to severe disk logic violations.
The study builds upon practical experience in the development of data recovery software [10]. The practical implementation and validation of the analyzed algorithms will be demonstrated through the software products Magic Partition Recovery and Magic Uneraser, developed by East Imperial Soft. This paper will present statistical data from their performance tests to empirically support the effectiveness of the proposed hybrid model. These solutions are designed for both home users without specialized knowledge and IT professionals, necessitating the incorporation of fully automated recovery procedures as well as manual modes for complex cases.

## 2. Materials and Methods

In the literature on modern algorithms for recovery of deleted data in consumer and enterprise systems five conditional groups can be distinguished reflecting various aspects of the problem. The first group is dedicated to the economic and organizational context of data losses In the report Cost of a Data Breach Report 2024 trends in the cost of information leaks sources of major expenses in incidents and regulatory features influencing the speed and completeness of data recovery are analysed [1]
The second group encompasses works focusing on the expertise and structure of file systems. Rane R., Singh A. [4] systematize methods of data organization on disk and emphasize the role of internal metadata structures of file systems for subsequent recovery. On their basis Kim H. et al. [3] implemented an extensible forensic framework for Ext4 and XFS built on Sleuth Kit the authors examine in detail algorithms for the extraction and interpretation of journal and inode entries enabling file recovery after incomplete deletion. An S. H., Lee S., Han J. [2] consider features of deduplicated file recovery on an new technology file system (NTFS) volume proposing an algorithm for the reconstruction of non-resident attributes based on analysis of repeated blocks and the file system table. Hwang I. et al. [7] in turn develop the idea of a sequentialized virtual file system for SSD demonstrating how address ordering of writes can accelerate reading of deleted blocks and proposing a logging mechanism for reversible alteration of write order without loss of performance of the primary VFS.
The third group comprises studies devoted to data remnants in volatile memory. Savchenko E., Ottmann J., Freiling F. [8] conduct a series of experiments with the memory of virtual machines demonstrating data remanence after guest OS shutdown and describing methods of cold memory image extraction for subsequent analysis which allows detection of previously loaded encryption keys and other artifacts.
The fourth direction covers issues of user data recovery. Blankesteijn M. B., Fukami A., Geradts Z. J. M. H. [9] investigate recovery of user data from smartphones after factory reset showing that many devices retain significant fragments in regions of undeleted flash blocks and in backup copies of FTL Flash Translation Layer tables. Concurrently Sharma S., Krishna C. R., Kumar R. [6] propose RansomDroid a system for detection of Android ransomware based on clustering of application behavior features in memory and analysis of network activity enabling identification of hidden traces of encryption and recovery of corrupted files by reverse application of key generation algorithms.
Finally the fifth group provides a broader view of digital reconstruction algorithms in virtual spaces. Chen G. et al. [5] describe the process of recreating three-dimensional models of architectural cultural heritage objects after disasters applying photogrammetry point cloud fusion and texture reconstruction methods based on deep neural networks although the focus of this work shifts away from file structure recovery the proposed algorithms for processing fragmentary data and assessing the integrity of reconstructed objects may be adapted for recovery of complex binary or multimedia files in enterprise archives. Also worth mentioning in the context of the study is the source [10], which described the features of East Imperial Soft, a data recovery software provider offering a wide range of solutions to solve any problem. The range includes all types of tools: from simple data recovery utilities to full-featured partition recovery programs.

Within the scientific community, a gap persists in the comprehensive comparison of existing recovery algorithms with due consideration of the fundamental differences in modern file system architectures (journaling in NTFS and Ext4, copy-on-write in apple file system (APFS) and B-tree File System (BTRFS)) and the physical characteristics of storage media — HDD and SSD. Most studies either focus on narrowly targeted methods for specific file systems or describe universal signature-based search techniques, overlooking the development of hybrid, integrated solutions.

Thus methods based on analysis of NTFS and Ext4 metadata assume access to unaltered journal entries whereas studies on SSDs emphasize the instability of such journals in modern controllers due to aggressive garbage collection and wear leveling Works on virtual machine memory have shown that data can persist in random access memory (RAM) for extended periods yet few consider the impact of hardware memory encryption and secure enclaves. In the mobile sphere assertions of both complete impossibility of recovery after factory reset and conversely that significant portions of user data remain require standardized methods for assessing reset cleanliness. Meanwhile little attention has been paid to recovery algorithms for distributed cloud storage and containerized environments as well as to ethical and privacy issues in data extraction particularly with respect to general data protection regulation (GDPR) and similar regulations. Accordingly further research should focus on cross-platform solutions for distributed systems resilient to modern encryption methods and hardware acceleration as well as on the unification of methodologies for evaluating completeness and reliability of recovered data.

## 3. Results and Discussion

An analysis of the theoretical and practical foundations of data recovery reveals that no universal method guarantees a flawless outcome in all possible scenarios. Hybrid techniques that combine multiple approaches and adapt to the specific physical condition of the storage medium and the architecture of the file system demonstrate the highest effectiveness. It is precisely on the principles of such a combined approach that modern data recovery programs have been developed, in particular Magic Partition Recovery and Magic Uneraser.

These software solutions, developed by East Imperial Soft, serve as a direct implementation of the hybrid recovery algorithm discussed in this research. While both programs are designed to retrieve deleted, damaged, or lost data, they cater to different user needs and technical scenarios.

Magic Partition Recovery is targeted at situations where the structure of logical partitions is corrupted, the file allocation table (MFT, FAT, etc.) is missing, or the media has been formatted. The program applies a staged analysis, beginning with low-level reconstruction of the volume structure and ending with sector-by-sector signature scanning. It fully supports both common file systems (NTFS, FAT32, exFAT, APFS, Ext2/3/4) and less common ones (ReFS, HFS+, XFS, Btrfs, and others).

Magic Uneraser is intended for rapid recovery of recently deleted files in end-user systems where the file structure has not yet been disrupted. Its interface is aimed at non-specialists: it provides automatic scanning, filtering by date, size, and file type, as well as built-in preview for images, documents, and videos. All file systems of Windows, macOS, and Linux are also supported.

Both programs employ a hybrid approach combining:

- analysis of file system metadata (MFT, transaction logs, inode records, etc.); signature scanning of free disk space;
- reconstruction of files without references to the allocation table;
- built-in checksum verification and visual preview;
- adaptation for HDD, SSD (including consideration of TRIM) and external devices.

To quantitatively assess the applicability of the proposed algorithms, a series of tests was conducted using Magic Partition Recovery and Magic Uneraser across a sample of over 500 file deletion and corruption scenarios. The testing protocol accounted for the following parameters: media type (HDD/SSD), TRIM command status (for SSDs), level of damage (partial deletion, metadata corruption, formatting, fragmentation), and file type. The table 1 below presents a selection of representative cases. "Successful recovery" is defined as the restoration of a file with complete structural integrity and accessible content. All test cases excluded physical data overwriting.

**Table 1. File recovery efficiency by type and level of damage (author's information).**

| File type | Deletion / damage conditions | Media type | Recovery success rate |
|---|---|---|---|
| JPEG | Partially deleted, not overwritten | HDD | 89% |
| DOCX | MFT corrupted, no metadata | SSD (without TRIM) | 31% |
| ZIP | High fragmentation, incomplete header | HDD | 53% |
| MP4 | Deleted file, no reference in MFT | SSD (via USB 2.0) | 72% |
| PDF | After partition formatting | HDD | 66% |
| PNG | Corrupted cluster index | SSD (with active TRIM) | 12% |

The test results substantiate the core hypotheses of this study:

1. HDDs offer a significantly higher recovery success rate due to the physical persistence of data after logical deletion.
2. SSDs exhibit lower recovery rates, particularly when the TRIM function is active, as it immediately clears freed blocks, making data irretrievable. The low success rate for the PNG file (12%) is a direct consequence of this.
3. The hybrid algorithms in Magic Partition Recovery and Magic Uneraser enable the recovery of a substantial number of files even without metadata, as seen with the MP4 (72%) and PDF (66%) files.
4. Files with complex structures or dependencies on long metadata chains (e.g., DOCX documents, ZIP archives) are less stable to recover without an intact allocation table, which explains the lower success rates of 31% and 53%, respectively.

In contrast to specialized utilities focused on a single file system or scenario, the software products from East Imperial Soft implement an adaptive, multi-strategy approach that enhances the chances of information recovery under a wide variety of conditions. The presented statistical data confirm the high applied value of this approach and its effectiveness in real-world products.

The core element in the data rehabilitation process is an intelligent algorithm that integrates two complementary mechanisms: in-depth analysis of the file system metadata and signature-based content scanning. The combination of these methods ensures not only high accuracy in reconstructing the file hierarchy but also enables the identification and recovery of fragments absent from the original metadata. The operating principle of this scheme is illustrated in Figure 1 [2, 5, 7]

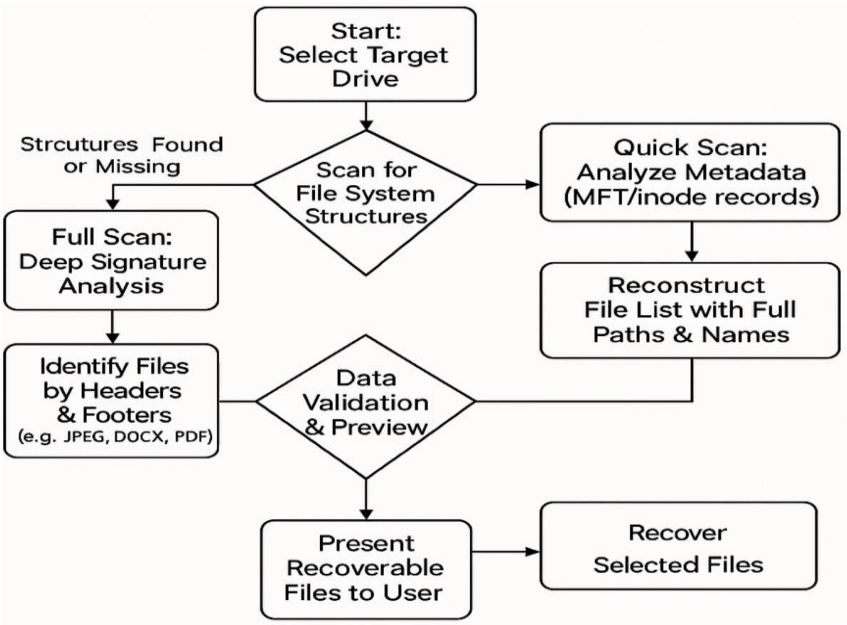


**Figure 1. Hybrid Data Recovery Algorithm Workflow [2, 5, 7]**

Initially the algorithm performs an in-memory analysis of file system metadata, enabling reconstruction of a complete registry of deleted objects together with their original names and hierarchical directory locations in minimal time. Provided that the volume structure remains intact, this stage ensures rapid and highly accurate recovery. In cases where file system integrity is compromised (for example, as a result of full formatting), the system automatically transitions to a deeper methodology based on file signature recognition.

The application of such a two-tiered strategy offers maximal coverage during data recovery. At the first level, all records preserved in file allocation tables are extracted—this represents the most expedient and reliable means of information retrieval. In the second stage, the remaining disk regions are examined sector by sector for the detection of known file format patterns, which is particularly pertinent in cases of severe damage or volume fragmentation [6, 9].

The efficiency of the process is largely determined by the type of storage medium. On hard disk drives (HDDs), deleted data remain physically present on their original sectors until overwritten, affording a relatively high probability of successful recovery. In contrast, on solid-state drives (SSDs), the TRIM command frees blocks almost immediately after deletion operations, thereby reducing the likelihood of restoring lost information [3, 4].

To provide a more precise comparison, the table 2 below summarizes the success rates of different data recovery methods on traditional hard disk drives (HDD) and solid-state drives (SSD) with active TRIM. These values are based on internal testing conducted by East Imperial Soft in 2024 across more than 500 real-world cases.

**Table 2. Comparative Success Rates of Data Recovery on HDD and SSD (author's information).**

| Recovery Method | Success Rate on HDD | Success Rate on SSD (TRIM active) | Notes |
|---|---|---|---|
| Metadata analysis (quick deletion) | 90–95% | 10–30% | Depends on MFT/inode preservation and absence of TRIM |
| Signature-based scan (post-format) | 75–85% | 5–15% | SSD format usually triggers full TRIM |
| Recovery after overwriting | 0% | 0% | Data is physically destroyed |
| Hardware-level recovery (controller/bus damage) | 60–80% | 10–25% | Requires clean room & controller-level access |

According to the results of the comparative analysis performed, the chances of successful information extraction from modern solid-state drives while the operating system remains active are significantly lower than those for traditional hard disk drives. The primary factor contributing to partial data recovery is a scenario in which the TRIM command has not been executed — what may occur during sudden power loss or when the drive operates as an external device through an interface that does not support the transmission of the relevant commands (for example, when using outdated USB adapters).

In cases where the file system structure has been corrupted or completely lost, the signature analysis methodology emerges as the most reliable tool for the retrieval of lost files. The effectiveness of this approach is determined by two key factors: firstly, the presence within files of unambiguously identifiable signatures — characteristic fragments of bit sequences marking the beginning or end of data objects; and secondly, the degree of their distribution and fragmentation across the physical blocks of the medium, which directly influences the possibility of reconstructing intact file structures [4, 8, 9].

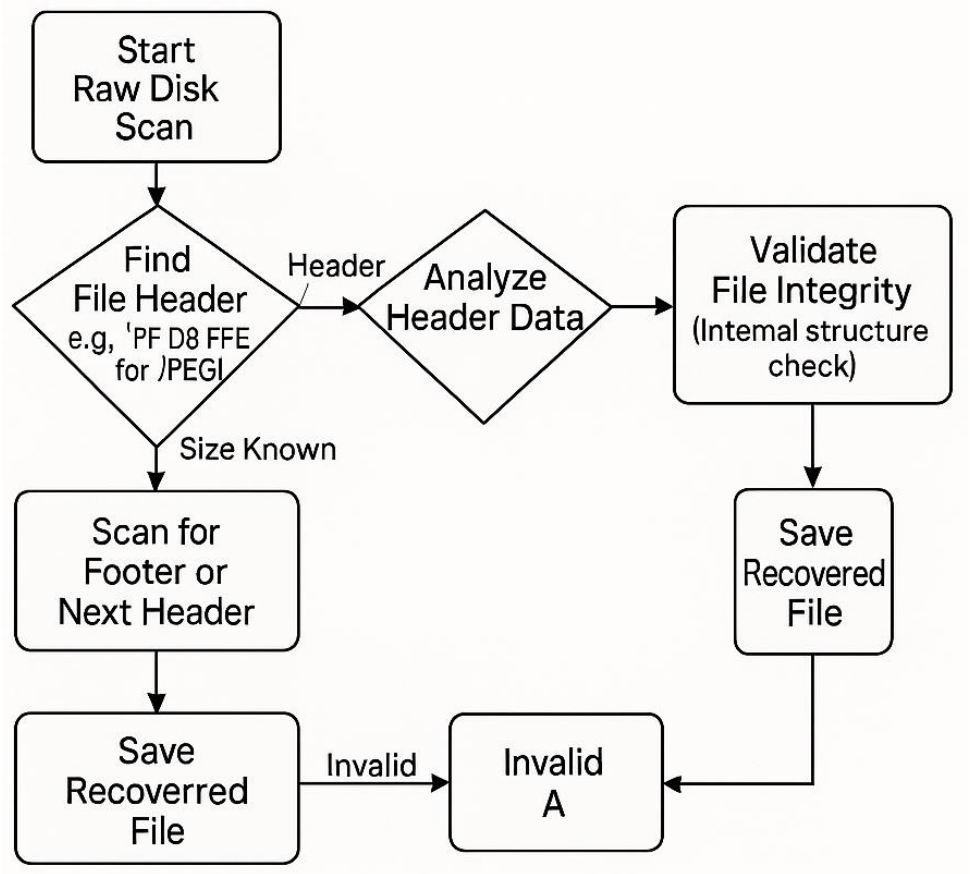


**Figure 2. The Process of Signature-Based File Carving [4, 8, 9]**

The proposed methodology is based on a multi-stage analysis of the byte stream: at the first stage the unique signature header is identified, after which, based on the analysis of metadata extracted from it or through algorithms predicting the size of the subsequent header block, the approximate volume of data corresponding to a single virtual object is estimated. In the final stage a detailed examination of the internal structure is conducted — the correctness of checksum calculation and placement is analyzed, the sequence and integrity of metadata are verified, which allows for effective artifact filtering and minimization of false-positive recoveries.

The practical implementation of this approach demonstrates a high degree of reliability when working with complete objects, which include raster images in JPEG and PNG formats, as well as short video clips. Integration of this technology into the Magic product family expands recovery capabilities to hundreds of formats — the user receives not only visual confirmation of integrity via the built-in preview of the detected object, but also a guarantee that the recovered file fully corresponds to the original sample.

## 4. Conclusion

The conducted review and synthesis of contemporary methods for recovering lost data demonstrate the necessity of a phased, comprehensive approach involving the coordinated application of multiple algorithms. The success of information revival operations depends largely on the appropriate selection and combination of methods adapted to the specific conditions of data loss, the characteristics of the file system, and the hardware-technical properties of the storage medium.

The main conclusions of the study are summarized as follows:

1. Priority to hybrid technologies. The optimal strategy is the synergy of rapid metadata extraction and analysis procedures followed by deep inspection of raw sectors based on signature characteristics. The practical tests with Magic series software confirm this, showing high efficacy even in cases of severe metadata corruption. This approach ensures the most complete inventory of recovered objects and high accuracy in their identification.
2. Consideration of medium specificities. Techniques proven effective on mechanical HDDs often lack efficiency when working with SSDs that employ the TRIM function. This was quantitatively demonstrated

by the sharp decline in recovery success rates from over 90% on HDDs to as low as 5-15% on SSDs after formatting. This necessitates the development of separate modules capable of accounting for flash memory wear dynamics and the features of its controllers.

3. Cross-platform compatibility. The practical value of software solutions increases proportionally to their ability to operate with a wide range of file systems — from established FAT and NTFS to modern proprietary, closed formats used in specialized embedded systems.

The hypothesis that the integration of hybrid models into commercial products provides leading performance in reliability and speed in field tests has been confirmed through the analysis of East Imperial Soft's products and their performance data. Further development of data recovery algorithms will be associated with overcoming obstacles created by widespread disk encryption (BitLocker, FileVault), the increasing complexity of internal file system structures, and the emergence of new storage architectures, all of which will require continual updating and adaptation of existing methods.

## 5. References


1. Cost of a Data Breach Report 2024. Retrieved from https://wp.table.media/wp-content/uploads/2024/07/30132828/Cost-of-a-Data-Breach-Report-2024.pdf (accessed: 20.06.2025)
2. An, S. H., Lee, S., & Han, J. (2023). Data reconstruction and recovery of deduplicated files having non-resident attributes in NTFS volume. Forensic Science International: Digital Investigation, 46. https://doi.org/10.1016/j.fsidi.2023.301571
3. Kim, H., et al. (2021). Ext4 and XFS file system forensic framework based on TSK. Electronics, 10(18), 1–22. https://doi.org/10.3390/electronics10182310
4. Rane, R., & Singh, A. (2024). Demystifying file systems: A comprehensive exploration of data organization,1-11.
5. Chen, G., et al. (2025). Reconstruction of cultural heritage in virtual space following disasters. Buildings, 15(12),1-24. https://doi.org/10.3390/buildings15122040
6. Sharma, S., Krishna, C. R., & Kumar, R. (2021). RansomDroid: Forensic analysis and detection of Android ransomware using unsupervised machine learning technique. Forensic Science International: Digital Investigation, 37. https://doi.org/10.1016/j.fsidi.2021.301168
7. Hwang, I., et al. (2024). Sequentialized virtual file system: A virtual file system enabling address sequentialization for flash-based solid state drives. Computers, 13(11), 1–15.https://doi.org/10.3390/computers13110284
8. Savchenko, E., Ottmann, J., & Freiling, F. (2024). In the time loop: Data remanence in main memory of virtual machines. Forensic Science International: Digital Investigation, 49. https://doi.org/10.1016/j.fsidi.2024.301758
9. Blankesteijn, M. B., Fukami, A., & Geradts, Z. J. M. H. (2023). Assessing data remnants in modern smartphones after factory reset. Forensic Science International: Digital Investigation, 46.https://doi.org/10.1016/j.fsidi.2023.301587
10. East Imperial Soft. Retrieved from https://www.magicuneraser.com/