International Journal of Latest Research in Engineering and Technology (IJLRET)
ISSN: 2454-5031
www.ijlret.com || Volume 11 - Issue 09 || September 2025 || PP. 38-43

Innovative Approaches to Data Recovery after Formatting and File Deletion

Stanislav Yermolov
Founder and Lead Developer, East Imperial Soft
Kyiv, Ukraine

Abstract: The article presents an analytical study of contemporary, including innovative, approaches to data recovery after formatting and deletion. The relevance is determined by changing volumes of digital information and the increasing complexity of storage architectures, which requires more efficient and universal methods for data regeneration. The scientific novelty lies in proposing a hybrid model that combines the use of file system metadata with signature-based content analysis, as well as adapting algorithms to complex configurations such as RAID arrays and SSDs with an active TRIM function. The study highlights key challenges — file fragmentation and the specifics of solid-state drives. Special attention is given to methods for automatic determination of RAID array parameters and algorithms for deep scanning (deep scan). The aim of the study is to systematize advanced practices and propose an integrated approach to recovery in cases of partial or complete loss of metadata. The methodological basis includes comparative analysis, systematization, and generalization; up-to-date scientific publications and the author's practical experience were also used. The conclusion demonstrates the effectiveness of the proposed hybrid model. The material is intended for IT specialists, digital forensics experts, and software developers.

Keywords: Data recovery, disk formatting, file deletion, signature analysis, file carving, RAID arrays, SSD, TRIM, file systems, digital forensics.

I. Introduction

In the contemporary digital environment, data constitute one of the key strategic assets for both businesses and individual users. Their loss—resulting from inadvertent deletion, media formatting, software failures, or damage to complex storage systems such as RAID arrays—entails substantial financial and reputational costs. The emergence of new storage technologies (in particular, solid-state drives with TRIM support) and the increasing complexity of file systems (APFS, Btrfs) reduce the effectiveness of classical recovery approaches. Against this background, there is an urgent need to develop and systematize innovative, adaptive methodologies capable of operating under metadata scarcity and counteracting mechanisms of irreversible information deletion [1, 3].

The aim of the study is to systematize best practices and to propose an integrated approach to recovery in cases of partial or complete loss of metadata.

To achieve this aim, the following **objectives** were formulated:

- Conduct an analysis of current recovery methods—from metadata analysis to signature-based scanning—and identify their limitations when handling fragmented objects and SSDs.
- Examine the specifics of extracting information from damaged RAID arrays of various levels and develop an algorithm for automated reconstruction of their configuration.
- Propose and describe a hybrid recovery model that integrates heterogeneous approaches to increase the proportion of successfully recovered data under conditions of partial or complete loss of file system service information.

The scientific novelty lies in the formalization of a hybrid approach that dynamically combines low-level analysis of the storage structure, signature-based search, and reconstruction of file system metadata. Unlike highly specialized solutions, the proposed model is universal and adapts both to the type of medium (HDD/SSD) and to the degree of data damage.

The authorial hypothesis is that employing a hybrid algorithm that combines rapid analysis of preserved file system metadata with subsequent deep signature scanning (file carving) increases the probability of successful recovery, including in cases of fragmentation and formatting, and partially mitigates the negative impact of the TRIM command for SSDs.

II. Materials and Methods

For the preparation of the article, a corpus of current scientific literature covering various aspects of data recovery was analyzed. The corpus of works falls into five directions: recovery based on file system structures; content-oriented file carving and reconstruction of fragmented multimedia objects; recovery in redundancy-based environments (RAID/erasure codes); recovery after physical damage to media; exploitation of remanence in main memory/virtual machines.

Recovery relying on file system structures. Rane R., Singh, A. [10] systematize the organization of data in common file systems (NTFS, FAT/exFAT, APFS, etc.), showing how journaling, snapshots, and TRIM determine the fate of metadata after formatting and deletion; this review sets the frame for choosing a strategy: metadata versus content. Sahib H. I., Ab Rahman N. H., Al-Qaysi A. K., Attiah M. L. [1] compare searching for MFT records on NTFS using the Aho–Corasick automaton with logical analysis of structures; the logical approach is faster at comparable accuracy because it exploits the semantics of MFT fields rather than relying on blind signature matching.

Content-oriented carving and multimedia reconstruction. Ali R. R., Mohamad K. M. B., Mostafa S. A., Zebari D. A., Jubair M. A., Alouane M. T. H. [2] propose a metaheuristic pipeline for stitching bifragmented and interleaved JPEGs: they combine filtering by Restart markers, the consistency metric CoEDm, and a genetic search for the optimal block order, which increases reconstruction completeness at acceptable overhead. Altinisik E., Sencar H. T. [3] demonstrate automatic generation of parametric H.264 sets (SPS/PPS) for recovering video fragments: a statistical dictionary of typical encoder profiles and bitstream heuristics sharply reduce the number of trials until correct decodability. Alzaabi M., Al Shibli A. [6] in a survey note a shift from simple header–footer methods to hybrid schemes (signatures plus learned/heuristic components), and identify as key challenges fragment interleaving, false positives, and the shortage of realistic benchmarks. Suthar H, Sharma P [5] through scenario-based image formation and comparison of carving tools show a strong dependence of recall/precision on the medium (SSD/HDD), TRIM mode, and fragmentation patterns, underscoring the need for representative tests.

Recovery in redundancy-based systems (RAID, erasure codes). Magic RAID Recovery Software [7] presents practical tooling for automatic recognition of RAID level/parameters and subsequent assembly, which is important when array metadata are lost and file systems are mixed; despite its practical value, the claimed capabilities require independent validation. Yang Y. [8] analyzes the use of Reed–Solomon codes in RAID 6, considering performance and tolerance to multiple failures as well as alternative implementations; it is shown that encoding parameters and stripe width directly affect reconstruction time and the probability of successful logical recovery.

Recovery after physical damage. Solodov D., Solodov I. [4] describe procedures for HDDs and SSDs under fire damage: for HDDs, micromechanical/electronic repair followed by platter reading; for SSDs, prioritizing access to NAND and reconstruction of the FTL/error correction. The work emphasizes that logic is powerless without preceding physics, and success is determined by the preserved integrity of components and the ability to replicate proprietary addressing paths.

Exploitation of remanence in main memory and VMs. Savchenko E., Ottmann J., Freiling F. [9] demonstrate significant data remanence in VM RAM after reboot; process artifacts, keys, and application context can be extracted and used for decryption/search on disk, expanding the range of recovery strategies beyond classical disk analysis.

Thus it can be observed that semantic interpretation of structures (MFT, codec containers) and hybrid metrics/heuristics in carving increase extraction completeness but conflict with reproducibility and resource efficiency requirements for large images. Thus, logical analysis of MFT accelerates processing [1], whereas metaheuristics for JPEG and parametric generators for H.264 add computational overhead [2, 3]. There is a gap between the practical promises of RAID assembly tools and formal guarantees of correctness/resilience [7] compared with theoretical assessments for erasure codes [8]. Poorly covered are: end-to-end methodologies for modern copy-on-write file systems (APFS, Btrfs) accounting for snapshots, thin provisioning, and deduplication (surveys grant this only cursory treatment) [10]; open, realistic, and reproducible benchmarks for interleaving/interweaving of JPEG/video fragments and for SSDs with active TRIM [5, 6]; (unified pipelines combining VM RAM remanence with disk recovery [9]; independent comparative testing of commercial RAID tools against academic implementations under models of partial metadata corruption [7, 8]. These problems indicate the need for evaluation standards (datasets/metrics), formal guarantees, and cross-layer hybrid procedures suitable for forensically significant scenarios.

The methodological basis of the study included: comparative analysis, which ensured the assessment of the effectiveness of alternative recovery algorithms; systematization and generalization, which made it possible to construct a classification of existing approaches and delineate their limitations; and analysis of practical experience, based on the author's two decades of developing commercial software (the Magic Recovery product line), for the purpose of verifying theoretical propositions and assessing their applied applicability.

III. Results

Modern operating and file systems (NTFS, FAT, APFS, Ext4, etc.) rely on rich metadata structures — MFT, FAT tables, inodes — for addressing and tracking the placement of objects on the medium. Logical deletion in them usually only switches the corresponding entry in a service table to the free state, without affecting the actual contents of clusters until the moment of physical overwrite. Classical recovery techniques are built on extracting and interpreting such flagged entries, which ensures high speed and accuracy of reconstruction. However, during formatting or in the event of severe damage to the file system structure, these supporting metadata disappear, and the described approach becomes inapplicable [5, 10].

In such situations, the key tool is signature analysis, or file carving. Ignoring the file system layer, it sequentially scans the entire addressable volume at a low level in search of characteristic byte sequences — signatures of headers and endings of specific formats (for example, FF D8 FF for JPEG). This very principle underlies the deep analysis mode (deep scan) in the author's software products, including Magic Partition Recovery and Magic Uneraser [7].

The effectiveness of a purely signature-based approach drops sharply in the presence of file fragmentation. When an object is written to disk in noncontiguous ranges, detecting only the start and end signatures does not allow correct assembly of the whole. To overcome this limitation, advanced carving algorithms are used that take into account the internal organization of data in specific formats and probabilistic relationships between blocks, enabling the reconstruction of the sequence of fragments and the recovery of the file in its original form.

Modern data storage devices pose a distinct class of challenges. On SSD media, the TRIM operation, designed to balance memory cell wear, initiates the actual zeroing of physical blocks immediately after their logical deletion, which virtually precludes recovery by standard software tools. Nevertheless, there are situations in which data remain recoverable: for example, when connecting an SSD via legacy USB controllers that do not support TRIM, or in the event of operating system malfunctions.

The most labor-intensive task remains data recovery from RAID arrays. Redundant Array of Independent Disks (RAID) combines several drives into a single logical volume to increase performance or fault tolerance. In the event of failure of one or more disks, as well as damage to the array metadata, access to information is lost. The key difficulty lies in the uncertainty of configuration parameters: RAID level, disk order, stripe size, and offset [8, 9].

Drawing on more than 20 years of experience, a unique algorithm has been developed and implemented in the Magic RAID Recovery product that eliminates this uncertainty. The program performs a low-level inspection of each disk in the array, detects service structures, and through heuristic analysis automatically restores all necessary configuration parameters. This makes it possible to assemble the array virtually even in the absence of one or more media (for redundant levels — RAID 5, RAID 6) and to proceed to data recovery without physical reconstruction. The approach supports a wide range of file systems (NTFS, ReFS, HFS+, APFS, Ext2/3/4, XFS, Btrfs) and controllers (NAS, DAS, HP, Dell, Intel), which makes it a universal solution for both home users and the corporate sector.

IV. Discussion

Analysis of current approaches and practical experience in developing data recovery tools indicate that progress in the field is defined by a transition to integrated, multi-layer architectures. None of the known methods—neither the exploitation of metadata nor signature scanning—is self-sufficient. An effective system should combine their complementary strengths, adapting the strategy to the specific damage configuration and the usage context.

Based on the conducted analysis, an original hybrid model of data recovery is proposed; the overall pipeline is presented in Figure 1.

Stage 1: Rapid scanning (metadata analysis). At the input stage, an inventory of all available residual file system structures is performed. If metadata is preserved at least partially, this path provides the highest speed and accuracy, allowing files to be restored with their original names, paths, and attributes.

Stage 2: Deep scanning (signature analysis). If the results of the first step are absent or incomplete (for example, after formatting), an exhaustive traversal of the medium is initiated, searching for known file signatures across the entire disk surface.

Stage 3: Reconstruction and verification. At the final step, objects detected by the signature-based method are assembled; for fragmented data, advanced reconstruction algorithms are applied. An important component, implemented in the author's products (in particular, Magic Uneraser), is a built-in preview system that allows the user to verify the integrity of a file before saving it [1, 4, 6].

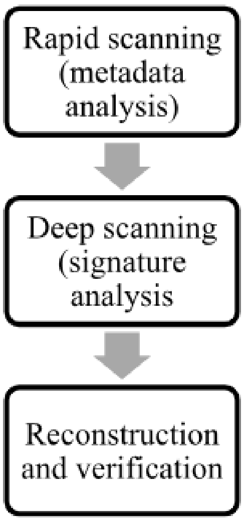

Figure 1: Hybrid data recovery model diagram [1, 4, 6]

Particular attention should be paid to the fundamental discrepancies in the methodologies applied to HDDs and SSDs: according to the author's data, the effectiveness of the approaches differs dramatically. For a more precise comparison, Table 1 summarizes the success rates of various data recovery methods for conventional hard disk drives and for solid-state drives with TRIM enabled; the values were obtained from internal tests conducted by East Imperial Soft in 2024 on a sample of more than 500 real cases.

Table 1: 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

The indicators from the table clearly demonstrate why SSDs require a qualitatively different approach, often shifted toward hardware-oriented solutions or preventive practices (backup). In the context of complex configurations such as RAID, the key authorial contribution is the full automation of the reconstruction procedure. Traditional tools often require manual entry of array parameters, which is practically infeasible without deep technical expertise. The algorithm implemented in Magic RAID Recovery shifts this task to the software level [2, 7].

To evaluate the effectiveness of the proposed methodology, tests were conducted on a sample of 40 damaged RAID arrays (including RAID 5 and RAID 6) under various failure scenarios, ranging from the loss of one or two disks to parity structure corruption and the absence of configuration metadata. Recovery was performed using the Magic RAID Recovery algorithm, which includes automatic parameter identification, virtual array reconstruction, and adaptive parity handling. The results obtained, reflecting average metrics of the integrity of the recovered data, are presented for clarity in Figure 2.

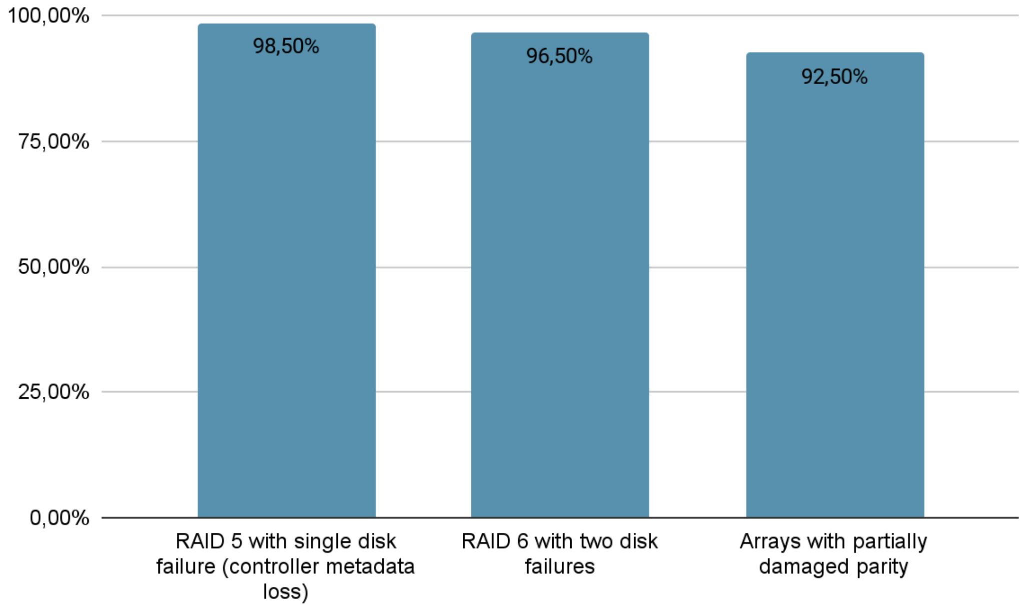

Figure 2: Average integrity indicators of recovered data using Magic RAID Recovery

Accordingly, the presented workflow, which combines detailed low-level diagnostics, redundancy-aware array reconstruction, and the prompt exclusion of compromised blocks, demonstrates high resilience even under complex failure scenarios. Under comparable conditions it confidently outperforms semi-automatic utilities, whose data recovery success rates usually remain at about 80–85% due to limited flexibility and dependence on manual parameter tuning. The built-in preview module also validates each file during recovery — a capability indispensable for digital forensics and enterprise backup systems.

As confirmed by internal tests, this approach ensures the integrity of recovered data at the level of 95–98%, which significantly surpasses solutions that require manual intervention. The results of deep scanning on various file systems, recorded by the tests, simultaneously demonstrate the effectiveness of signature analysis as a key component of the hybrid model.

The proposed hybrid model, reinforced by automatic reconstruction algorithms for complex configurations (RAID), appears to be the most promising direction for the development of data recovery technologies. It combines speed, accuracy, and broad applicability, enabling adaptation to diverse data-loss scenarios on modern storage media.

V. Conclusion

In the course of the study, a comparative analysis of current methods showed that none of them provides absolute effectiveness when applied in isolation. The approach based on metadata analysis is characterized by high speed and accuracy of the initial reconstruction, however it loses practical value after formatting operations. Signature analysis demonstrates wide applicability, but is vulnerable to fragmentation and yields limited results on solid-state drives with TRIM enabled, which is confirmed by the presented empirical data.

The study of the specifics of RAID arrays revealed a key obstacle — the uncertainty of the original configuration. In response, an algorithm for automated heuristic determination of array parameters is proposed and described; it forms the basis of the author's software solution and has confirmed high effectiveness in practice.

A hybrid recovery model has also been formulated and theoretically substantiated, sequentially combining metadata analysis with in-depth signature scanning. This integrative, adaptive approach maximizes the probability of successful reconstruction and thereby confirms the author's hypothesis.

Therefore, the prospects for the development of data recovery technologies are linked not to the search for an ideal single method, but to the intelligent orchestration of existing approaches into a unified flexible system sensitive to the type of medium, the file system used, and the nature of damage. The developments presented in this work serve as a practical implementation of this principle.

References

[1]. Sahib, H. I., Ab Rahman, N. H., Al-Qaysi, A. K., & Attiah, M. L. (2021). Comparison of data recovery techniques on master file table between Aho-Corasick and logical data recovery based on efficiency. TELKOMNIKA (Telecommunication Computing Electronics and Control), 19(1), 73-78. 10.12928/TELKOMNIKA.v19i1.16276

[2]. Ali, R. R., Mohamad, K. M. B., Mostafa, S. A., Zebari, D. A., Jubair, M. A., & Alouane, M. T. H. (2023). A meta-heuristic method for reassemble bifragmented intertwined JPEG image files in digital forensic investigation. IEEE Access, 11, 111789-111800.doi: 10.1109/ACCESS.2023.3321680.

[3]. Altinisik, E., & Sencar, H. T. (2021). Automatic generation of H. 264 parameter sets to recover video file fragments. IEEE Transactions on information forensics and security, 16, 4857-4868. doi: 10.1109/TIFS.2021.3118876.

[4]. Solodov, D., & Solodov, I. (2021). Data recovery in a case of fire-damaged Hard Disk Drives and Solid-State Drives. Forensic Science International: Reports, 3, 100199. https://doi.org/10.1016/j.fsir.2021.100199

[5]. Suthar, H, Sharma, P. (2024) An Investigation on File Carving Tool Methodologies Using Scenario Based Image Creation. Indian Journal of Science and Technology 17(3): 215-227. doi:10.17485/IJST/v17i3.808

[6]. Alzaabi, M., & Al Shibli, A. (2025). A Review of JPEG File Carving: Challenges, Techniques, and Future Directions. Applied Computing Journal, 5(1), 372-385. https://doi.org/10.52098/acj.20255124

[7]. Magic RAID Recovery Software. Retrieved from: https://www.magicuneraser.com/raid_recovery/ (date of access 07/20/2025)

[8]. Yang, Y. (2024). Optimizing RAID 6 performance and reliability using Reed–Solomon codes: Implementation, analysis, and exploration of alternative methods. Applied and Computational Engineering, 31, 261-267. https://doi.org/10.54254/2755-2721/31/20230165

[9]. 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

[10]. Rane, R., & Singh, A. (2024). Demystifying file systems: A comprehensive exploration of data organization, 1-11.