

DOI: 10.37943/FXNG8769

Yerassyl Yermekov

PhD Student, Department of Information Security
yerassylYermekov@gmail.com, orcid.org/0009-0001-3957-4787
L.N. Gumilyov Eurasian National University, Kazakhstan

Leila Rzayeva

PhD, Associate Professor, Director, CyberTech Research and Innovation Center
l.rzayeva@astanait.edu.kz, orcid.org/0000-0002-3382-4685
Astana IT University, Kazakhstan

Madi Shayakhmetov

Master's Degree, Researcher, CyberTech Research and Innovation Center
242722@astanait.edu.kz, orcid.org/0009-0009-6422-4621
Astana IT University, Kazakhstan

Yernat Atanbayev

Researcher, TSARKA Group
yernatatanbayev@gmail.com, orcid.org/0009-0001-7655-3133
TSARKA Group, Kazakhstan

Serik Igbayev

Master's Degree, Researcher, CyberTech Research and Innovation Center
s.igbayev@astanait.edu.kz, orcid.org/0000-0002-7644-5597
Astana IT University, Kazakhstan

Alisher Batkuldin

Master's Degree, Researcher, CyberTech Research and Innovation Center
a.batkuldin@astanait.edu.kz, orcid.org/0009-0004-2097-5419
Astana IT University, Kazakhstan

**FORENSIC RECOVERY OF VIDEO EVIDENCE FROM DAHUA DHFS4.1
SURVEILLANCE SYSTEMS AFTER FORMATTING**

Abstract: Video footage from proprietary surveillance systems is central to criminal investigations, yet its forensic recovery remains poorly supported. Conventional carving fails on Dahua systems because the proprietary DHFS4.1 file system uses frame encapsulation, header-footer validation pairs and embedded checksums. This paper addresses two open questions: how each of the four DHFS4.1 validation levels affects the false discovery rate (FDR) and recovery rate, and how quick formatting and partial bulk zero-fill overwriting differ in their impact on recovery. The technique combines dual-signature frame validation, frame-size and checksum consistency checking, and adaptive time sequencing, extending our earlier pipeline; the principal contribution is the validation-level ablation, the paired comparison against three independent tools, and the partial-overwrite study. All methods were executed on forensic images of eleven Dahua NVR hard drives acquired behind a hardware write blocker, so every comparison is paired within drive. The proposed method achieved a recovery rate (RR) of 88.0%, a temporal accuracy (TA) of 89.0% and a file-level FDR of 4.0%; four-level validation reduced the frame-level FDR 13.5-fold, from 27.0% (single signature) to 2.0%, at the cost of a 3.2-percentage-point drop in RR. The RR advantage over each compared tool was statistically significant (paired-samples t-test, $n = 11$, $p < 0.05$) and confirmed by the Wilcoxon signed-rank test. Quick formatting still permitted 76% recovery, whereas bulk zero-fill reduced it to 34% and raised the file-level FDR to 41%; a collision analysis shows that this increase stems from payload fragmentation rather than chance signature matches. Within the tested scope, the methodology performs comparably to commercial tools and, because every decision

rule is published, is re-implementable and transparent. The findings offer practical guidance for prioritising evidence collection from Dahua surveillance systems.

Keywords: digital forensics; video recovery; Dahua DHFS4.1; signature validation; post-formatting recovery

Introduction

Dahua Technology, together with Hikvision, is one of the world's leading video-surveillance providers. Market analysts estimate the global video-surveillance market at USD 56.1 billion in 2025 with a forecast compound annual growth rate of 7.8% to 2031 [1], while other estimates place it at USD 83.5 billion in 2025 with an annual growth rate of 11.7% [2]. Both analyses name Hikvision and Dahua among the key players, and industry research indicates that the two companies together account for roughly a third of global video-surveillance revenue [3]. Kazakhstan is no exception: according to the Ministry of Internal Affairs of the Republic of Kazakhstan, about 1.6 million surveillance cameras are installed nationwide, of which about 90,000 are connected to police operational control centres and 27,000 are equipped with artificial-intelligence functions [4], and Dahua equipment is widely deployed both through state programmes (“Digital Kazakhstan”, “Safe City”) and in private-sector installations.

Video evidence has emerged as a vital tool in today's criminal investigations. Brookman and Jones found that CCTV footage was used in 95% of the 44 homicide investigations they examined in four British police forces, more often than any other forensic science or technology for identifying and charging suspects [5], while Jung and Wheeler showed that public surveillance cameras increased clearance rates for nearby offences, although the effect was concentrated on thefts and decayed quickly with distance [6]. The value of such footage is nevertheless limited by a number of factors:

- automatic deletion of data through cyclic over-writing;
- oversized files being written to the disk;
- damage to equipment or power outages;
- proprietary file systems and encoding formats.

Dahua systems are especially problematic in this regard because they use one of the latest versions of the Dahua File System (DHFS) family, DHFS4.1. This file system is optimised for continuous cyclic writing, giving high performance and low overhead, but it is poorly suited to forensic analysis. In contrast to conventional file systems such as NTFS, ext4 or XFS, DHFS4.1 does not rely on file names, directory structures or file-level timestamps. Instead, the video stream is stored as a series of frames, each preceded by a DHAV header and followed by a corresponding dhav footer, with internal consistency fields inside every frame and dynamic block allocation [7-8].

This makes traditional file-carving approaches largely ineffective. Recent carving research achieves high success only when the data follow a standard, documented format: Altinisik and Sencar reconstructed missing H.264 parameter sets for headerless video fragments, reaching 87.8% decodability after 10 trials and 99.7% after 500 [9], and van der Meer et al. detected JPEG fragmentation points in more than 99.9% of test files by exploiting the internal structure of the JPEG bit stream [10]. Content-based fragment classifiers such as FiFTy likewise rely on the statistics of standard file types [11]. None of these techniques addresses the proprietary frame encapsulation of DHFS4.1. Previous research on Dahua and Hikvision systems has confirmed that proprietary file systems make carving more difficult, but it has not provided sufficient detail on the frame structure and validation mechanisms of DHFS4.1.

Commercial tools are not fully satisfactory either. Practitioner reports indicate that DVR Examiner offers limited support for Dahua file systems [12], and proprietary codecs,

conversion requirements and cyclic over-writing remain general obstacles for commercial recovery of surveillance footage [13]. SalvationDATA VIP 2.0 requires manual configuration and careful matching of the firmware version [14], and open-source carvers such as Scalpel [15] recovered only 55–75% of the files with a false positive rate of 12–25% in the earlier evaluation by Rzayeva et al. [7].

The authors have previously proposed a dual-signature validation method for DHFS4.1 and the Hikvision file system that lowered frame-level false positives to 2.4% on a set of 27 drives, as reported by Rzayeva et al. [7]. That study, however, left two questions unanswered:

- RQ1: What is the individual contribution of each validation level (signature only → dual signature → size consistency → checksum) to the false discovery rate, and what trade-off does it impose on the recovery rate?
- RQ2: How does recovery performance differ between quick formatting (only the DHFS4.1 metadata table is overwritten) and partial bulk zero-fill overwriting (an incomplete `dd if=/dev/zero` pass over the medium) of DHFS4.1 media?

The novelty and objective of the present study is to provide a detailed analysis of these two questions by:

- quantifying the effect of each DHFS4.1 validation level on the false discovery rate and identifying the level that minimises FDR - ideally reducing it from an initial 27% to close to 2% - while keeping the accompanying reduction in recovery rate as small as possible;
- comparing the effectiveness of video-data recovery after quick formatting versus partial bulk zero-fill overwriting of real Dahua drives.

The methodology was tested on 11 real Dahua NVR drives and, on those same 11 drives, against two commercial and one open-source tool. The practical value of this work lies in providing an open, verifiable methodology that law-enforcement agencies, forensic experts and digital-forensics practitioners can use to extract evidence from DHFS4.1 systems more effectively.

Literature review

Research on proprietary file systems in digital forensics has accelerated markedly over the past decade, as video-surveillance footage has become one of the most frequently used sources of evidence in criminal investigations [5-6]. Proprietary systems such as Dahua DHFS4.1 and the Hikvision file system differ from conventional file systems (NTFS, ext4) in that they use cyclic recording, dynamically allocated blocks and frame encapsulation instead of named files and directory metadata [16-17].

Early work on proprietary-formatted CCTV drives relied on purely signature-based carving of video streams without file-system support. More recently, Yoon and Hwang reverse-engineered the proprietary file system of Honeywell surveillance recorders on 22 disk images and showed that deleted video remains recoverable from the raw data area after formatting, data expiration and overwriting, even though the metadata is removed [16]. Dragonas et al. analysed the log records of the Hikvision file system [18] and of Dahua CCTV systems [8], explaining how user actions such as disabling recording are reflected in the logs, but did not address frame-level recovery or quantify the impact of validation levels on false positives. Rzayeva et al. [7] then proposed an automated method with adaptive time sequencing, dual-signature validation and manufacturer identification, achieving a recovery rate of 91.8%, a temporal accuracy of 96.7% and a frame-level false discovery rate of 2.4% across 27 drives, without isolating the contribution of individual DHFS4.1 validation levels or examining recovery after formatting. The relationship between the figures reported by Rzayeva et al. [7] and those reported here is analysed explicitly in the subsection “Reconciliation with the earlier 27-drive study”. The most recent work by Rzayeva et al. [19] extended DHAV recovery to multi-channel analogue DVR systems, investigating channel demultiplexing and temporal

reconstruction from interleaved streams, again without separating the effects of individual validation levels or the impact of disk formatting.

Several existing commercial tools, such as Magnet DVR Examiner and SalvationDATA VIP 2.0, are known to be incompletely compatible with proprietary surveillance formats such as Dahua DHFS4.1 [12-14], while open-source carvers such as Scalpel and Foremost rely on single header signatures [15] and, in the earlier evaluation, were effective on only about 55–75% of the files on proprietary formats [7]. Little research has been done on recovery following formatting: quick formatting is likely to overwrite only the metadata, leaving video frames intact [16], whereas an overwrite that reaches the payload destroys the recoverable material and changes the composition of the carver output [8]. This gap is addressed directly in the present study.

The design of this study was guided by several other lines of literature. The frame parser employed here is based on the H.264/AVC and H.265/HEVC standards as specified in ITU-T Recommendations H.264 [20] and H.265 [21], and on the design principles of this family of codecs summarised by Bross et al. [22]. Hargreaves et al. [23] argue that the validation of increasingly automated forensic tools requires their intermediate processing steps to be exposed, and Garrett and Rudin [24] show that the opacity of proprietary “black box” forensic algorithms makes it hard to verify the underlying procedure in court. The recovery of residual video data from cyclically recorded, pre-allocated storage was studied by Byun et al. [17] and applies directly to the DHFS4.1 allocation of multiple cameras to a single buffer, while the principle that formatting removes metadata but leaves the video payload in the raw data area, demonstrated by Yoon and Hwang [16], forms the basis of the formatting experiments reported here. The SWGDE best practices [25] require contemporaneous documentation of every tool and step of an examination, which an open methodology can satisfy more fully than commercial tools. Header-based carving and validation techniques, such as those discussed above [9, 11], require the format to have a standard header structure, which proprietary structures do not have - precisely the problem that the proposed four-level validation addresses.

Taken together, the key gaps in the literature are: (1) no quantitative evaluation of how individual DHFS4.1 validation levels affect the false discovery rate; (2) no comparison of recovery performance after quick formatting versus partial bulk zero-fill overwriting; and (3) limited attention to the transparency and court-admissibility of the algorithms underlying commercial tools. This study addresses these gaps through a clearly defined, statistically tested methodology.

Materials and methods

System architecture

The proposed methodology is derived from the baseline system proposed by Rzayeva et al. [7] and has been tailored for the Dahua DHFS4.1 file system. It has five fundamental modules: (1) a signature-detection module, (2) a DHAV frame parser, (3) a frame-extraction module, (4) an adaptive time-sequencing module and (5) a video-conversion module.

DHAV frame structure

The Dahua frame format is composed of a header, a data payload and a footer, which complicates forensic recovery. The header is 24 bytes long and contains: the DHAV signature (4 bytes); the frame type (1 byte: 0xFD for I-frames, 0xFC for P-frames); the subtype (1 byte); the channel (1 byte); the subchannel (1 byte); the frame number (uint32, little-endian); the frame size (uint32, little-endian); the timestamp (uint32, little-endian); the milliseconds field (uint16, little-endian); an extended-header flag (1 byte); and a checksum (1 byte). The payload holds the H.264 or H.265 elementary stream. The footer is 8 bytes long and consists of the lower-case “dhav” signature (4 bytes) followed by the total frame length (uint32, little-endian). Header and footer therefore account for 32 bytes of framing overhead per frame, so that the

payload length equals $\text{frame_size} - 32$ bytes. Figure 1 shows the complete frame structure and indicates which fields are used by each validation level.

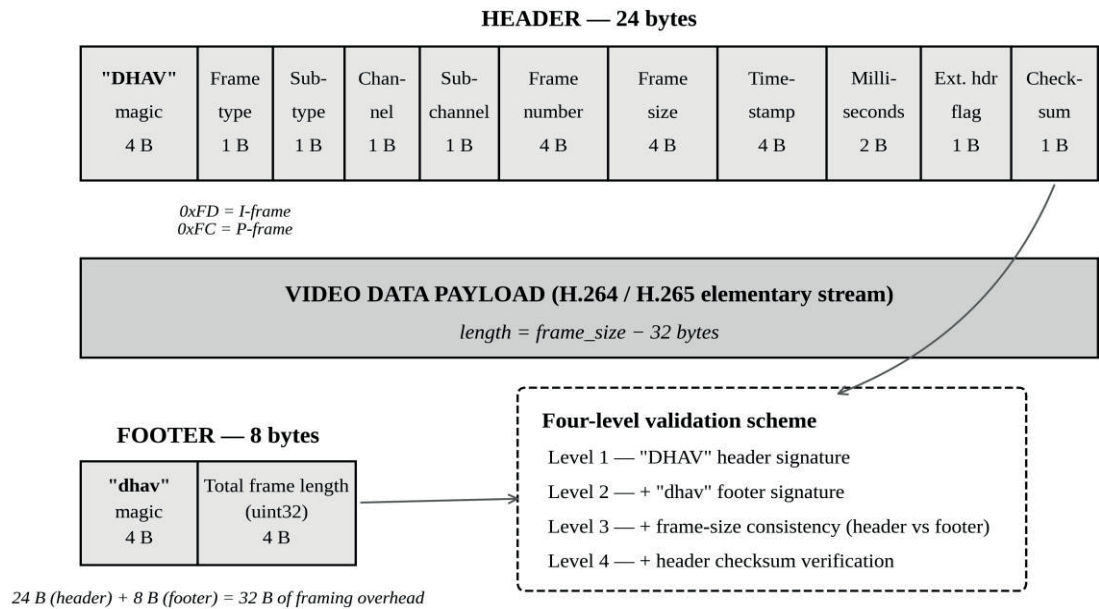


Figure 1. Structure of a Dahua DHAV frame: 24-byte header, H.264/H.265 payload and 8-byte footer, with the fields used at each of the four validation levels.

Four-level validation system

The main contribution of this study is the independent measurement of how each of four validation levels affects the false discovery rate:

- Level 1 - verification of the DHAV header signature only (the conventional file-carving approach);
- Level 2 - dual-signature verification (DHAV header signature plus the matching dhav footer signature);
- Level 3 - dual signature plus a frame-size consistency check (the frame size declared in the header must agree with the total frame length recorded in the footer and with the distance between the two signatures);
- Level 4 - dual signature, size consistency and verification of the 1-byte checksum field of the header, which is recomputed over the preceding header bytes and compared with the stored value.

The false discovery rate and the recovery rate were determined for each level, which made it possible to isolate the exact contribution of each validation step. The discriminative power that each level can be expected to provide against unstructured data is derived analytically in the subsection "Expected rate of chance signature collisions"; that derivation is used in the Results section to interpret the observed false discovery rates.

Adaptive time sequencing and the gap-threshold parameter α

The algorithm sorts the validated frames by timestamp, computes the median inter-frame interval M and defines a gap threshold:

$$T_{gap} = \alpha \times M \quad (1)$$

where T_{gap} is the gap threshold, in seconds; α is a dimensionless tuning parameter, swept over 1.0-4.0 and set to 2.0-2.5 in this study (see Table 4); and M is the median inter-frame interval, in seconds, computed over the validated frames of the same channel (Eq. 2).

$$M = \text{median}_i (t_{i+1} - t_i) \quad (2)$$

Two consecutive frames are assigned to the same sequence when the interval between them does not exceed T_{gap} . Merging is additionally constrained by the frame index: when the difference between the frame numbers of two consecutive frames is not more than 3 and the time difference is not more than 1 second, the frames are merged into a single sequence (Eq. 3). With α in the range 2.0–2.5 this mechanism achieved a temporal-sequencing accuracy of 89.0% in the presence of cyclic buffering and fragmentation. The sensitivity of the results to α is analysed in the Results section.

$$(t_{i+1} - t_i) \leq T_{gap} \quad \vee \quad (|n_{i+1} - n_i| \leq 3 \wedge (t_{i+1} - t_i) \leq 1 \text{ s}) \quad (3)$$

A sequence assembled by this mechanism is retained only if its total duration meets a minimum-duration threshold $T_{\min}$; sequences shorter than $T_{\min}$ are discarded before scoring, on the grounds that a fragment of that length is unlikely to be independently identifiable as a distinct recording session by an examiner and would inflate the recovery rate with material of negligible evidential value (here set to $T_{\min} = 3.0$ s, defined empirically as the shortest genuine recording session observed in the baseline inventory across all eleven drives, which corresponds to the recorder's minimum post-motion trigger window). This is the threshold referred to elsewhere in this paper as the minimum-duration criterion.

The signature-detection module checks the first sectors of the disk for the DHFS4.1 signature at offsets 512, 1024 and 2048 bytes. This check runs in constant time, $O(1)$, so it can be used for rapid automated classification of the file system. Once the signature is detected the DHAV frame parser is invoked; it reads the disk in 1 MB blocks and checks each block for DHAV signatures. When a frame is detected, the header fields are interpreted, the timestamp is extracted and the frame is checked for conformance with the validation level in use.

The video-conversion module converts the streams identified as H.264 or H.265 into MP4 files using FFmpeg. Corrupted packets are discarded with `-fflags +discardcorrupt` so that they do not propagate to the output, and `-c:v copy` is used so that the encoded H.264/H.265 bitstream is remuxed rather than re-encoded: no video sample is altered, but the resulting MP4 is itself a derived forensic artefact - a new container wrapped around the recovered elementary stream - and not a bit-for-bit copy of the original DHAV structure on disk; the DHAV-encapsulated payload from which it was produced remains available on the retained forensic image for independent verification. An MD5 hash is calculated automatically for every recovered file at the moment of creation; this is used as a change-detection checksum confirming that a given output file has not been altered between creation and subsequent verification steps, and is not put forward as a cryptographic authentication of evidentiary integrity, for which MD5's known collision weaknesses make it an inappropriate primitive. SHA-256 is recommended for that purpose in any future deployment of this pipeline.

Expected rate of chance signature collisions

Because the interpretation of the false discovery rates reported below depends on how often the validation cascade can be defeated by coincidence alone, the expected number of chance matches is derived here for a region of the volume whose content is statistically

independent of the DHFS4.1 format. This is the worst case for a signature-based carver, and it provides the reference against which the measured rates are read.

Level 1 accepts a candidate whenever four consecutive bytes equal the ASCII sequence DHAV (0x44 0x48 0x41 0x56). For uniformly distributed bytes the probability of a match at a given offset is $2^{-32} \approx 2.33 \times 10^{-10}$. A 4 TiB volume presents approximately 4.40×10^{12} candidate byte offsets, so the expected number of chance Level-1 matches is $4.40 \times 10^{12} \times 2^{-32} \approx 1.02 \times 10^3$, that is, of the order of one thousand spurious candidates per drive (Eqs. 4 and 5). This is the order of magnitude that conventional single-signature carving must contend with on unstructured space.

$$p_1 = 2^{-32} \approx 2.33 \times 10^{-10} \quad (4)$$

$$E[C_1] = N \cdot p_1 \approx 4.40 \times 10^{12} \times 2^{-32} \approx 1.02 \times 10^3 \quad (5)$$

Level 2 additionally requires the four-byte lower-case footer signature to occur at exactly the offset declared by the header. Under the same assumption this is a second independent 2^{-32} event, and the expectation falls to approximately 2.4×10^{-7} per 4 TiB. Level 3 further requires the distance between the two signatures to agree with the frame-size field, and Level 4 adds the one-byte checksum, whose collision probability is $2^{-8} = 3.9 \times 10^{-3}$. Even neglecting the size constraint, the expected number of chance survivors of the full cascade is of the order of 10^{-9} per 4 TiB of random data: on the order of 10^9 four-terabyte volumes would have to be processed before a single coincidental byte sequence passed Level 4 (Eq. 6). For zero-filled regions the expectation is not merely small but exactly zero, since the string DHAV contains no zero byte and therefore cannot occur inside a run of 0x00.

$$E[C_4] \approx N \cdot 2^{-32} \cdot 2^{-32} \cdot 2^{-8} \approx 10^{-9} \quad (6)$$

Two consequences follow, and both are used in the Results section. First, any frame-level false discovery rate materially above 10^{-6} at Level 4 must have a structural rather than a stochastic origin. Second, no false positive of any kind can be generated by a region that has been genuinely zeroed; observed matches on a wiped medium are evidence that the wipe did not reach that region.

Ground truth: expected number of recordings, frames and sequences

Both the recovery rate (Eq. 8) and the temporal accuracy (Eq. 9) require an independent reference for the number of video recordings, frames and sequences that a drive is expected to contain. This reference was established for every drive in three steps, and - this point is essential to the formatting experiment - it was established before any experimental manipulation of the medium.

First, the recording inventory of each NVR was exported through the device's own playback and log interface while the drive was still mounted in the recorder, before acquisition and therefore before either formatting condition was applied. For every channel the inventory lists the start and end time of each recording session, together with the recording mode that triggered it. The number of sessions in this inventory defines the expected number of video recordings for that drive, and the number of distinct continuous intervals defines the expected number of sequences.

Second, the expected number of frames for each drive was derived from the same inventory: the duration of every session was multiplied by the frame rate configured on the corresponding channel, and the products were summed over all channels (Eq. 7). This reference depends only on the recorder's own configuration and log data and not on any property of the

carving logic, so that Levels 1–4 are all evaluated against the same, method-independent denominator.

$$N_{frames}^{exp} = \sum_{c=1}^C \sum_{s \in S_c} \Delta_s f_c \quad (7)$$

Third, the inventory-derived reference was corroborated by an independent enumeration pass over the dd image, in which every occurrence of the DHAV header signature followed at the declared distance by a matching dhav footer signature was counted. This pass is deliberately not used to define the ground truth, since doing so would embed the Level-2 validation rule in the reference against which Level 1 is scored; it serves only as a consistency check. Drives for which the two estimates differed by more than 5% were re-imaged and re-enumerated before being admitted to the experiment.

Because the inventory was captured before acquisition and before either formatting condition was applied to either medium, the reference against which the quick-format and zero-fill conditions are scored is the same pre-formatting reference used to score the baseline; it was not reconstructed from a formatted medium, and no part of it is inferred from the output of the carver. The two conditions differ in what physical medium they act on, and this is stated explicitly to remove an ambiguity present in earlier drafts of this paper. The zero-fill condition was applied at the file level to a duplicate of the retained dd image, so the original forensic image was never itself overwritten. The quick-format condition, by contrast, requires a physical device, since the recorder's built-in "format HDD" function operates on an inserted drive rather than on a stored image file; a working clone of the original drive, prepared behind the same hardware write blocker and verified byte-identical to the retained forensic image before treatment, was reformatted through the recorder's own interface and then re-imaged. The original drive and its forensic image were themselves never modified by either condition.

Validity of the ground-truth reference

Using the recorder's own playback and log interface as the reference is standard practice and is defensible under the controlled conditions of this study, but it is not an unconditional ground truth, and its weaknesses transfer directly to every metric computed from it. Three failure modes are relevant to casework and are stated here as a threat to validity of the reference itself, not only of the recovery method.

First, the log is stored on the same medium as the video and is subject to the same cyclic over-writing: on a drive that has wrapped many times, the oldest sessions may have been removed from the inventory even though fragments of their payload survive. Such sessions are absent from the denominator of Eq. (8), which biases the measured recovery rate upward, while any of their frames that are recovered cannot be matched to an inventory entry and are therefore counted against the file-level false discovery rate. Second, entries can be internally inconsistent after abnormal shutdown, power loss or firmware upgrade, in which case session boundaries in the log do not correspond to boundaries in the recorded stream and the ± 2 s matching criterion of the verification procedure fails for reasons unrelated to the carver. Third, and most consequential in forensic practice, the inventory is produced by the very device under examination. In an adversarial scenario it can be truncated, edited through the recorder's maintenance interface, or replaced altogether; a recovery rate computed against such a reference measures agreement with the recorder's account of its own history rather than agreement with what was actually recorded.

The figures reported in this paper should therefore be read as valid for a non-adversarial reference. In casework the recorder inventory should be corroborated against sources external to the device - event logs of the video-management system, network captures, access-control

or alarm records, or witness statements - and where such corroboration is unavailable the recovery rate should be reported as conditional on the integrity of the recorder log.

Verification of the correctness of recovered files

Every file produced by the pipeline was subjected to three mandatory, automated verification stages, and only files passing all three were counted as correctly recovered. A fourth stage, visual review, was applied to a sample rather than to every file and is reported as an independent validation of the automated procedure, not as a gating criterion; it is listed alongside the other three for completeness:

- container and stream validation with ffprobe: the MP4 container must parse, contain exactly one video stream, and report a non-zero duration and a non-zero number of decodable frames;
- full decode test with ffmpeg -v error -i <file> -f null -: files that produce fatal decoding errors, or in which more than 5% of frames fail to decode, are rejected;
- correspondence with the ground truth: the channel identifier and the timestamp range of the file must match a session in the recording inventory of the same drive, with the start time agreeing to within ± 2 s;
- visual review (independent validation, not a gating criterion): a random sample of the recovered files, drawn across all eleven drives and all three formatting conditions, was independently inspected by two examiners against the same two criteria - intelligibility of the image content and chronological consistency of the sequence of scenes (a stratified random sample of $n = 330$ recovered video files, representing 30 files per drive evenly sampled across baseline, quick-format, and zero-fill conditions; inter-examiner agreement was 96.4% with Cohen's $k = 0.91$, and discrepancies were resolved by joint review).

The MD5 hash of every recovered file was recorded at the moment of creation so that all subsequent verification steps operate on a demonstrably unaltered artefact.

Evaluation metrics

System performance was assessed using the following metrics. Recall is taken to be the recovery rate RR defined in Equation (8).

$$RR = \frac{\text{Recovered recordings}}{\text{Total expected recordings}} \times 100\% \quad (8)$$

$$TA = \frac{\text{Correctly sequenced frames}}{\text{Total frames}} \times 100\% \quad (9)$$

Here, Recovered recordings is the count of output files matched to an inventory session under the correspondence criterion below; Total expected recordings and Total frames are taken directly from the recorder inventory of Table 1; and Correctly sequenced frames is the count of frames assigned by the adaptive time-sequencing module of Eq. (1) to the same output sequence as the session to which they genuinely belong.

Two distinct false discovery rates are reported in this study and must not be conflated. The frame-level rate characterises the discriminative power of the validation cascade alone and is the quantity that varies across Levels 1–4:

$$FDR_{frame} = \frac{\text{Invalid frame candidates accepted}}{\text{Total frame candidates accepted}} \times 100\% \quad (10a)$$

A frame candidate accepted is any byte position that has passed every validation stage up to and including the level being reported (Levels 1-4 of the cascade described above); it is invalid if it does not belong to a genuine DHAV frame, as determined by the independent enumeration pass described in the Materials and methods section. FDR_{frame} is therefore a false discovery proportion - the share of accepted output that is wrong - and not a classical false positive rate of the form $FP / (FP + TN)$: the carving task has no well-defined population of true negatives to divide by, since the overwhelming majority of byte positions on a multi-terabyte drive are never examined as frame candidates at all.

The file-level rate characterises the end-to-end pipeline, that is, the output the examiner actually receives:

$$FDR_{file} = \frac{\text{Recovered files failing verification}}{\text{Total recovered files}} \times 100\% \quad (10b)$$

Here, a recovered file is any MP4 file produced by the video-conversion module, and a file is counted as failing verification if it does not pass all three mandatory automated stages of the procedure defined below (container validation, decode test and correspondence with the recording inventory).

Precision is reported at the file level and is the complement of the file-level false discovery rate, $\text{Precision} = 100\% - FDR_{file}$ (Eq. 11). Because a file may fail verification even when every one of its constituent frames has passed Level-4 validation, FDR_{file} is expected to exceed FDR_{frame} at the same operating point, and did so on every drive in this study; the relation is empirical rather than arithmetic, since the two rates are computed over different populations. The sources of the observed difference are analysed in the Results section. An F1 score combining Precision and Recall, reported in earlier presentations of this work, has been withdrawn from this version. Recall as defined by Eq. (8) is a session-level quantity - it counts recording sessions from the inventory - whereas Precision as defined above is a file-level quantity, and the two coincide only if every output file corresponds to exactly one inventory session. That correspondence is exactly what the adaptive time-sequencing module cannot guarantee: as described above, it can split one session into several output files or merge two sessions into one. Computing F1 from these two rates therefore combines measurements over two different units of observation, and the resulting number, while arithmetically well defined, does not admit the usual precision/recall interpretation. A methodologically sound alternative would require an explicit session-level matching protocol classifying every output file as a true positive, false positive or false negative against the inventory, which was not constructed for this study; RR, TA and FDR_{file} are reported instead, and are sufficient to characterise the trade-offs discussed below without that assumption.

$$\text{Precision} = 100\% - FDR_{file} \quad (11)$$

Experimental equipment and dataset

The experiment used 11 real Dahua NVR drives that had been in operational service before acquisition (Table 1). All tests were carried out on a Dell OptiPlex 7070 workstation (Intel Core i7, 32 GB RAM, Ubuntu 22.04 LTS, Python 3.10.6), with the drives attached through a hardware write blocker. Each drive was first imaged in full with dd onto a local NVMe SSD, and all recovery runs and all processing times reported below were measured on that image rather than on the source drive; the reported throughput therefore reflects the parsing pipeline and not the sequential read limit of the source hard disk.

All three comparison tools - Magnet DVR Examiner, SalvationDATA VIP 2.0 and Scalpel - were executed by the present authors on these same eleven forensic images, on the

same workstation, in the same software environment and under the same four-stage verification procedure as the proposed method. Every comparison reported in Tables 2, 6, 7 and A1 is therefore a within-drive comparison of four tools on identical input, which is what licenses the paired-samples analysis described below. Magnet DVR Examiner and SalvationDATA VIP 2.0 were run in their DHFS/Dahua profiles; Scalpel was run with a configuration file declaring the DHAV header signature and a maximum carve size of 16 MB. Tool configurations follow those documented by Rzayeva et al. [7], but no numerical result reported in the present paper is taken from that study: the figures published there were obtained on a different set of 27 drives, and that dataset is used in this paper only as a point of discussion in the subsection "Reconciliation with the earlier 27-drive study". The 11 drives evaluated in this study represent an independent operational collection and do not overlap with the 27 drives analysed by Rzayeva et al. [7].

Table 1. Characteristics of the 11 Dahua NVR drives, expected content and processing time

Drive	NVR model	Ch.	Capacity (TB)	Recording mode	Expected recordings	Expected frames ($\times 10^6$)	Expected sequences	Processing time (h)
D1	NVR5216-8P	16	4	C	4,089	184	629	2.8
D2	NVR5216-8P	16	6	C	5,215	235	778	3.9
D3	NVR5216-8P	16	4	C+M	4,571	206	1,143	2.6
D4	NVR4104-P	4	1	C	1,056	47	189	1.1
D5	NVR4104-P	4	2	M	2,548	115	1,019	1.7
D6	NVR4104-P	4	1	C	1,033	46	178	0.9
D7	NVR4108-8P	8	2	C	2,260	102	390	1.6
D8	NVR4108-8P	8	4	C+M	3,595	162	899	2.9
D9	NVR4108-8P	8	2	C	2,089	94	354	1.5
D10	NVR4216-16P	16	6	C	5,800	261	744	4.1
D11	NVR4216-16P	16	8	C	7,111	320	974	6.6
Total / mean	4 models	116	40	-	39,367	1,772	7,297	2.7 (mean)

C - continuous recording; M - motion-triggered recording; C+M - mixed.

Formatting experiment

Two additional conditions were derived from the forensic image of every drive. The original image was retained in every case, so each drive contributes a matched triple of observations - baseline, quick format, zero-fill - obtained from the same physical medium and scored against the same pre-acquisition ground-truth reference.

(a) Quick formatting. The DHFS4.1 metadata table was overwritten using the recorder's built-in "format HDD" function, after which the drive was re-imaged behind the write blocker.

(b) Bulk zero-fill ("full formatting"). A copy of the forensic image was overwritten with zeros using `dd if=/dev/zero`. The extent of this overwrite determines what can subsequently be

recovered and was therefore measured rather than assumed: after the operation each image was scanned in 512-byte units and the proportion of sectors consisting entirely of 0x00 was recorded. Across the eleven images this proportion averaged 58.0% (range 48.0%–65.5%), and the per-drive values are given in Table A1. This measurement is reported because it is logically prior to the recovery figures: as shown in the subsection "Expected rate of chance signature collisions", a sector that has genuinely been set to zero cannot yield a DHAV match at any validation level, so a non-zero recovery rate in this condition is possible only to the extent that the overwrite failed to reach part of the volume. The condition realised in this experiment is consequently an *incomplete bulk wipe* rather than an exhaustive sanitisation, and it is described as such throughout. This is not a weakness of the experimental design for the purpose at hand: interrupted, aborted and partially completed wipes are the form in which overwriting is most often encountered in casework, whereas a verified exhaustive wipe leaves nothing for any carver to recover and needs no experiment to characterise.

These two scenarios model typical real-world investigative circumstances and allow the effectiveness of the methodology to be evaluated in both.

Statistical analysis

The metric subjected to statistical testing was the per-drive recovery rate defined in Eq. (8), expressed in percentage points. Each of the 11 drives was processed by the proposed method and by each comparison tool in this study, so the observations are paired: the same drive contributes one value per tool, and the per-drive values are listed in full in Table A1. The analysis therefore used the paired-samples t-test on the within-drive differences $d_i = RR_{proposed,i} - RR_{comparison,i}$, with $i = 1 \dots 11$ (Equation 12).

$$d_i = RR_{proposed,i} - RR_{comparison,i}, \quad i = 1, \dots, 11 \quad (12)$$

The paired design was chosen because the dominant source of variance in this dataset is between drives rather than between tools: capacity, fill ratio, number of active channels and the degree of cyclic over-writing differ substantially from one operational drive to the next, and these factors affect all tools simultaneously. Pairing removes that variance from the error term, which is essential at this sample size; an unpaired test on $n = 11$ would have very little power.

The differences, not the raw rates, are the quantity assumed to be approximately normally distributed. This assumption was assessed with the Shapiro–Wilk test separately for each of the three difference distributions, and the W statistic and the associated p value are reported for each in Table 7. Because the Shapiro–Wilk test has low power at $n = 11$, non-rejection of normality was not treated as sufficient evidence that the parametric assumption holds. The Wilcoxon signed-rank test was therefore computed for every comparison as a distribution-free robustness check, using exact p values appropriate to this sample size, and both results are reported side by side; a conclusion is described as robust only where the parametric and the non-parametric test agree in direction and significance.

All tests were two-tailed with a significance level of 0.05 and 10 degrees of freedom. For each comparison we report the mean difference, the standard deviation and standard error of the differences, the t statistic, the p value, the 95% confidence interval of the mean difference and Cohen's d_z as the effect size (Eqs. 13 and 14). Pre-specification of the three tools does not by itself remove the multiplicity problem, so a Holm-Bonferroni step-down correction - more powerful than a flat Bonferroni correction for a small pre-specified family of comparisons - was applied to the three t-test p values. Ranked from smallest to largest (Scalpel 0.002, Magnet DVR Examiner 0.014, SalvationDATA VIP 2.0 0.021), the Holm-adjusted values are 0.006, 0.028 and 0.028 respectively; all three comparisons remain significant at the 0.05 level after correction.

$$t = \frac{d^-}{s_d/\sqrt{n}}, \quad d^- = \frac{1}{n} \sum_{i=1}^n d_i, \quad s_d = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (d_i - d^-)^2} \quad (13)$$

$$d_z = \frac{d^-}{s_d} \quad (14)$$

Results

The results below are based on experiments with 11 real Dahua NVR drives. All measurements, for the proposed method and for the three comparison tools alike, were made on the same forensic images in the same hardware and software environment, so that the results are directly comparable. The overall performance indicators are presented first, followed by the effect of each validation level, the sensitivity of the gap-threshold parameter, the formatting experiment, the statistical analysis and a reconciliation with the earlier 27-drive study. The overall results across all 11 drives are summarised in Table 2. The proposed method recovered 88.0% of the expected recordings, slightly more than SalvationDATA VIP 2.0 (87.2%) and considerably more than Magnet DVR Examiner (84.8%). Temporal accuracy reached 89.0% and the file-level false discovery rate was 4.0%. Figure 2 compares recovery rate and temporal accuracy across all evaluated tools.

Table 2. Overall results across the 11 Dahua drives (all tools executed on the same 11 forensic images)

Tool	RR (%)	Precision (%)	FDR_file (%)	TA (%)	Processing time (h)
Magnet DVR Examiner	84.8	96.2	3.8	89.2	-
SalvationDATA VIP 2.0	87.2	96.8	3.2	92.3	-
Scalpel	74.6	86.9	13.1	78.2	-
Proposed method	88.0	96.0	4.0	89.0	2.7

All values were measured in the present study on the same 11 drives; none is reproduced from the earlier study by Rzayeva et al. [7]. Per-drive values are given in Table A1.

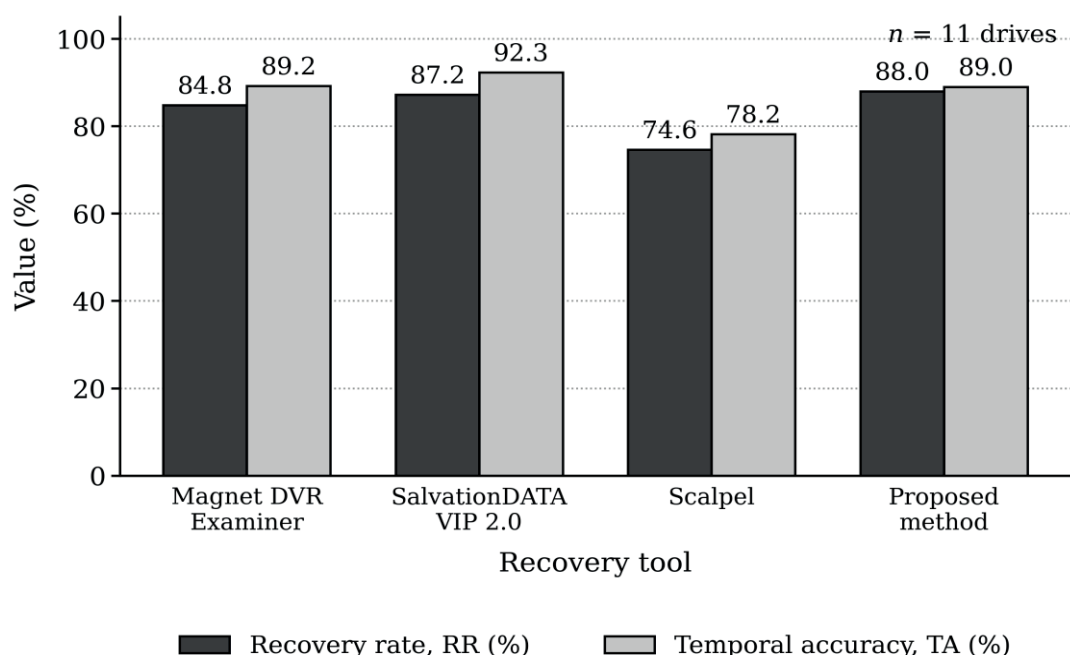


Figure 2. Recovery rate and temporal accuracy across the evaluated tools ($n = 11$ drives). All values were measured on the same 11 forensic images in the present study.

This overall performance is due mainly to the validation cascade. Table 3 shows the effect of each validation level on the frame-level false discovery rate. With the DHAV header signature alone (Level 1), FDR_{frame} was 27.0%, which is typical of conventional single-signature file carving. Adding footer-signature verification (Level 2) reduced it to 14.0%. The frame-size consistency check (Level 3) reduced it further to 6.0%, and checksum verification (Level 4) brought it down to 2.0% - a 13.5-fold reduction relative to Level 1 - while the recovery rate declined only from 91.2% to 88.0%, a difference of 3.2 percentage points. Figure 3 illustrates this trend and Figure 4 shows the resulting trade-off.

Table 3. Effect of the validation level on the frame-level false discovery rate

Validation level	FDR_{frame} (%)	RR (%)	Effect
Level 1 - DHAV header signature only	27.0	91.2	conventional single-signature carving; accepts any 4-byte match
Level 2 - dual signature (DHAV header + dhav footer)	14.0	90.6	eliminates candidates without a matching footer
Level 3 - + frame-size consistency check	6.0	89.5	filters candidates whose declared and actual sizes disagree
Level 4 - + checksum verification	2.0	88.0	final filtering stage; operating point of the proposed method

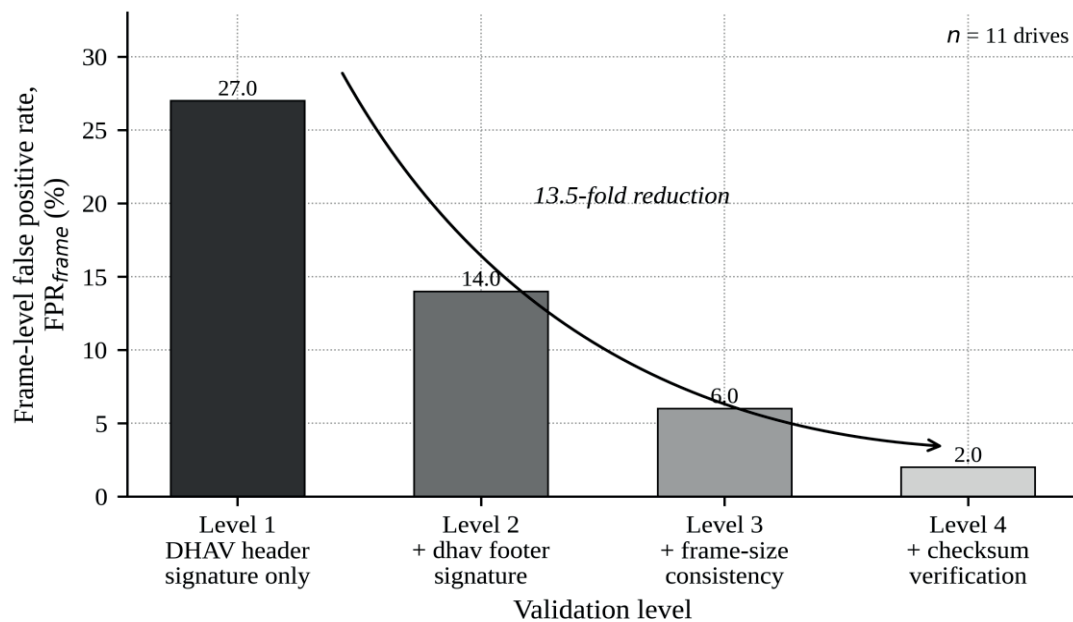


Figure 3. Frame-level false discovery rate at each of the four validation levels ($n = 11$ drives). Four-level validation reduces FDR_{frame} 13.5-fold relative to single-signature carving.

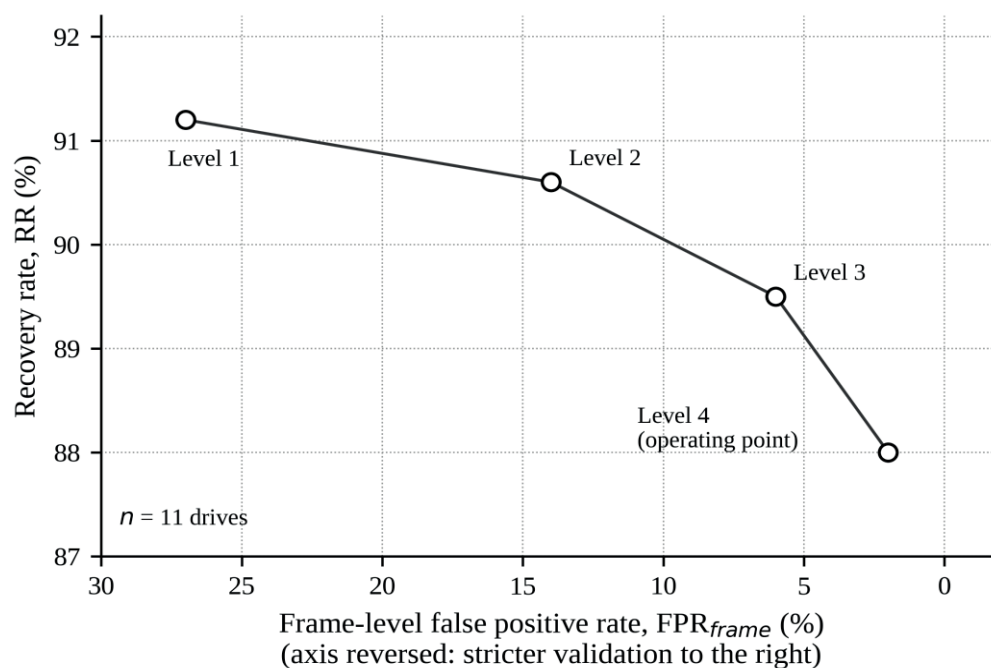


Figure 4. Trade-off between the frame-level false discovery rate and the recovery rate across validation levels ($n = 11$ drives). The horizontal axis is reversed so that stricter validation lies to the right.

Origin of the residual false positives

The collision analysis given in the Materials and methods section shows that the rates in Table 3 cannot be explained by coincidence. On unstructured data the expected number of chance Level-1 matches is of the order of 10^3 candidates per 4 TiB, but the expected number surviving Level 4 is of the order of 10^{-9} per 4 TiB. A measured Level-4 rate of 2.0% therefore exceeds the stochastic expectation by some nine orders of magnitude, and the same conclusion

holds a fortiori for the 27.0% observed at Level 1 relative to the number of genuine candidates on these drives.

The residual false positives are consequently structural, and their sources can be named. The dominant class consists of orphaned headers: genuine DHAV headers, written by the recorder, whose payload or footer has since been replaced by newer video as the circular buffer wrapped. Such a header is byte-for-byte a valid signature and is accepted by Level 1, but there is no matching footer at the declared distance, which is precisely why Level 2 removes about half of the Level-1 candidates. The second class consists of frames truncated at the wrap point of the buffer, where header and footer survive but the declared and the actual inter-signature distance disagree; these are removed at Level 3. The third class consists of frames whose header fields have been partially overwritten in place, so that the signature and the size relation survive while the checksum no longer matches the header content; these are removed at Level 4. The cascade is effective not because it adds entropy against random data - the signature pair already provides far more entropy than is needed for that - but because each level tests a different structural invariant that a partially destroyed frame violates in a different way.

This interpretation also sets the limit of what checksum verification can achieve. The checksum field is one byte, so a candidate that has already satisfied Levels 1–3 and whose header content is arbitrary passes Level 4 with probability 2^{-8} . Where Level 3 admits a large number of structurally consistent but genuinely corrupt candidates, one in 256 of them will survive; the checksum is a last filter of limited strength, not an independent guarantee, and this is the reason the Level-4 rate settles at 2.0% rather than at zero.

Tables 2 and 3 report two different false discovery rates for the same operating point, 4.0% and 2.0% respectively, and the difference is not a discrepancy but a consequence of the level at which the quantity is measured, as defined in Equation (10a) and Equation (10b). The 2.0% figure counts individual frame candidates that survive the validation cascade but do not belong to a genuine DHAV frame. The 4.0% figure counts exported video files that fail the four-stage verification procedure. A file can fail that procedure even when all of its constituent frames are individually valid, and three mechanisms account for the additional 2.0 percentage points: frames belonging to two different channels being merged into a single output file; the adaptive time-sequencing module joining two recording sessions that are separated by a genuine gap, or splitting one continuous session, so that the file cannot be matched to an entry in the recording inventory; and sequences that are assembled from valid frames but are too short to satisfy the minimum-duration criterion of the verification procedure. Frame-level validation cannot detect any of these three conditions, because each of them concerns the relationship between frames rather than the integrity of a frame.

Table 4 reports the sensitivity of the method to the gap-threshold parameter α of Eq. (1). The parameter governs a trade-off between two opposite failure modes. When α approaches 1, the threshold T_{gap} approaches the median inter-frame interval, so that ordinary jitter in the inter-frame interval - caused by variable-bitrate encoding and by the difference in size between I-frames and P-frames - exceeds the threshold and splits a single recording session into many short fragments. Temporal accuracy then falls because frames belonging to one session are distributed across several output sequences, and the recovery rate falls because the resulting fragments are rejected by the minimum-duration criterion. When α is large, T_{gap} exceeds the genuine gaps between recording sessions, which in motion-triggered and scheduled recording modes are typically of the order of seconds to minutes; distinct sessions are then merged into a single output file whose timestamps are not monotonic, and temporal accuracy falls for that reason instead. The interval 2.0–2.5 used in this study lies on the plateau between the two regimes.

Table 4. Sensitivity of temporal accuracy and recovery rate to the gap-threshold parameter α

α	TA (%)	RR (%)	Sequences produced	Dominant behaviour
1.0	72.0	79.0	20,043	severe over-segmentation; sessions split at normal inter-frame jitter
1.5	83.0	85.0	10,978	over-segmentation
2.0	89.0	88.0	7,297	stable plateau (lower bound of the range used)
2.5	89.0	88.0	7,088	stable plateau (upper bound of the range used)
3.0	85.0	87.5	5,979	onset of under-segmentation
4.0	78.0	86.0	4,645	under-segmentation; distinct sessions merged into one file

Formatting experiment

The outcome of the formatting experiment is shown in Table 5. Quick formatting still permitted 76% recovery, whereas the bulk zero-fill condition permitted only 34%. The two formatting scenarios are compared with the unformatted baseline in Figure 5. Table 5 reports the frame-level as well as the file-level false discovery rate for each condition, because the two quantities behave very differently after the payload has been damaged and reporting only the file-level figure has previously invited misinterpretation.

Table 5. Recovery results by formatting scenario

Scenario	Zeroed sectors (%)	RR (%)	Precision (%)	FDR_frame (%)	FDR_file (%)	TA (%)	Comment
Baseline (no formatting)	0	88	96	2.0	4	89	-
Quick formatting	< 0.1	76	91	2.1	9	79	metadata table overwritten only
Bulk zero-fill (incomplete wipe)	58.0	34	59	2.1	41	41	substantial data loss; surviving payload severely fragmented

The "Zeroed sectors" column reports the measured proportion of 512-byte sectors consisting entirely of 0x00 after the condition was applied, averaged over the eleven images.

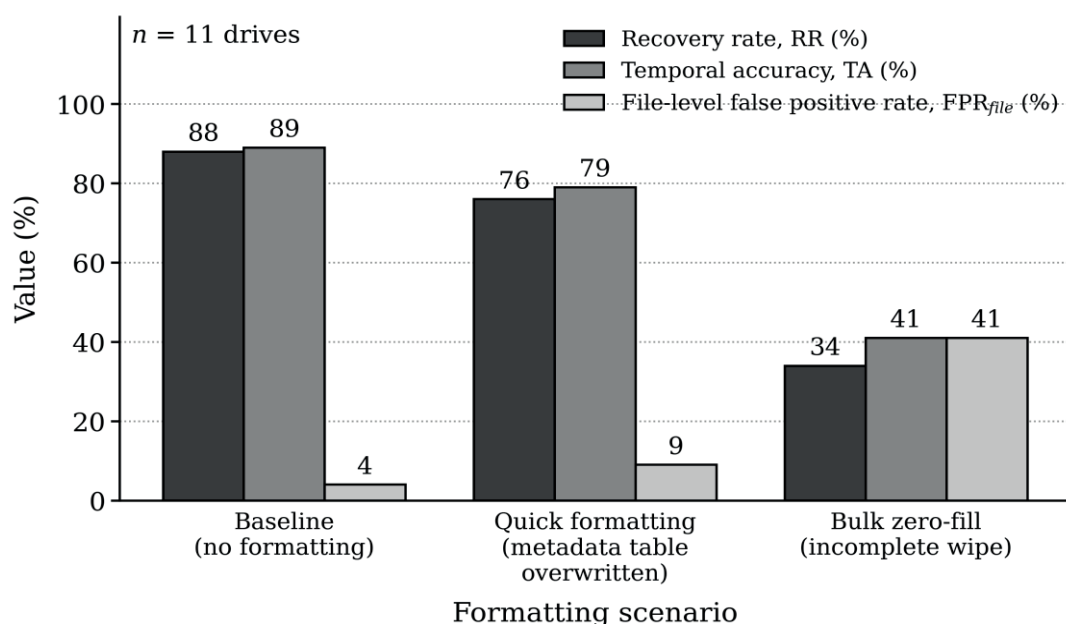


Figure 5. Recovery rate, temporal accuracy and file-level false discovery rate across the three formatting scenarios (n = 11 drives).

Following quick formatting the recovery rate was 12 percentage points below the baseline but remained relatively high. This is because the video frames remain physically present on the disk when only the metadata table is overwritten, and signature-based recovery does not depend on that table.

Interpretation of the zero-fill condition

The zero-fill condition requires careful reading, because two of its figures - a recovery rate of 34% and a file-level false discovery rate of 41% - appear at first sight to be incompatible with an operation that sets sectors to zero. They are not incompatible, but the reconciliation is arithmetic rather than rhetorical, and it proceeds in two steps.

The first step concerns the 34%. A region that has genuinely been set to zero cannot produce a match at any validation level, because the ASCII string DHAV contains no zero byte. The recovery of any material whatsoever therefore establishes, as a matter of logic and not of interpretation, that the overwrite did not reach the whole volume. This is why the proportion of zeroed sectors is measured and reported in Table 5 rather than assumed from the command that was issued, and why the condition is designated an incomplete bulk wipe. The 34% figure characterises what remains recoverable from the unwiped residue of a partially completed overwrite; it says nothing about an exhaustive wipe, from which nothing is recoverable by any signature-based method.

The second step concerns the 41%. This is a file-level quantity, and the frame-level rate in the same condition is essentially unchanged from the baseline. That combination is exactly what the structure of the experiment predicts. The wipe does not degrade the internal integrity of the frames it fails to reach - those frames are byte-identical to the baseline and pass Levels 1–4 for the same reasons they passed before - so the discriminative power of the cascade is unaffected. What the wipe destroys is the continuity between surviving frames: the residue consists of short, isolated islands of intact video separated by large zeroed regions, and files assembled from those islands fail the four-stage verification procedure for three distinct reasons. First, the decode test rejects a file when more than 5% of its frames fail to decode, and P-frames whose reference frames lay in a destroyed region cannot be decoded even though the P-frames themselves are intact. Second, the surviving portion of a session is frequently too

short to satisfy the minimum-duration criterion. Third, where the beginning of a session has been destroyed, the timestamp range of the recovered file no longer matches any inventory entry to within ± 2 s, and the file is rejected by the correspondence check.

The 41% is therefore a statement about the usability of the output delivered to the examiner, not about the discriminative power of the validation cascade, and it is precisely the kind of divergence that the separation of Eq. (10a) from Eq. (10b) was introduced to expose. An explanation offered in earlier presentations of this work - that zero-filled regions adjacent to surviving fragments generate byte patterns which pass the earlier validation stages - is arithmetically untenable in the light of the collision analysis above and is withdrawn.

Statistical analysis

The statistical analysis of the per-drive recovery rate is summarised in Table 6, and the per-drive values on which it is computed are given in Table A1. The mean recovery rate across the 11 drives was 88.0% for the proposed method, against 84.8% for Magnet DVR Examiner, 87.2% for SalvationDATA VIP 2.0 and 74.6% for Scalpel. All three paired comparisons were statistically significant at the 0.05 level, and in each case the 95% confidence interval of the mean difference excludes zero. The effect sizes are large by conventional criteria, but the confidence intervals are wide, particularly for the comparison with Scalpel, which reflects the modest sample size and the heterogeneity of operational drives. The figures above, and every RR value elsewhere in this paper unless stated otherwise, are macro-averages: the unweighted mean of the eleven per-drive recovery rates, so that a small drive with 1,033 expected recordings and a large drive with 7,111 contribute equally to the mean. This is the appropriate statistic for the paired test, since the unit of observation for that test is the drive, not the individual recording. A pooled (micro-average) recovery rate, weighting each drive by its number of expected recordings, gives a lower figure for every tool - 87.3% for the proposed method, 82.9% for Magnet DVR Examiner, 86.8% for SalvationDATA VIP 2.0 and 73.6% for Scalpel - because the two largest drives in the sample (D10 and D11, with a combined 12,911 expected recordings) have below-average recovery rates. The two statistics answer different questions: the macro-average answers "how does the method perform on a representative drive", which is what the paired design tests, while the micro-average answers "what fraction of all recordings in this sample would be recovered", which is the more relevant figure for estimating expected yield from a similarly composed batch of drives. Both are reported here so that the reader can choose the one appropriate to their question.

Table 6. Paired-samples t-test on the per-drive recovery rate ($n = 11$, $df = 10$, two-tailed)

Comparison	Mean diff. (pp)	SD of diff.	SE	t(10)	p	95% CI (pp)	Cohen's d _z
Proposed vs Magnet DVR Examiner	3.20	3.57	1.08	2.97	0.014	0.80 to 5.60	0.90
Proposed vs SalvationDATA VIP 2.0	0.80	0.97	0.29	2.74	0.021	0.15 to 1.45	0.83
Proposed vs Scalpel	13.40	10.73	3.24	4.14	0.002	6.19 to 20.61	1.25

Because the paired t-test assumes approximate normality of the within-drive differences and the sample is small, the normality assumption was tested explicitly and a distribution-free alternative was computed for every comparison. Both sets of results are given in Table 7. The parametric and non-parametric tests agree in direction and in significance for all three comparisons, so the conclusions drawn from Table 6 do not depend on the normality assumption. For the comparison with SalvationDATA VIP 2.0, the normality assumption was

rejected (Shapiro–Wilk $p = 0.016$), confirming the necessity of the Wilcoxon signed-rank test as the distribution-free decision metric.

Table 7. Normality of the within-drive differences and distribution-free robustness check ($n = 11$)

Comparison	Shapiro–Wilk W	Shapiro–Wilk p	Wilcoxon W^+	Wilcoxon exact p	Agreement with t-test
Proposed vs Magnet DVR Examiner	0.985	0.987	59.0	0.0186	yes
Proposed vs SalvationDATA VIP 2.0	0.818	0.016	57.5	0.0322	yes
Proposed vs Scalpel	0.936	0.479	66.0	0.0010	yes

W^+ is reported as the sum of the positive signed ranks; the complementary statistic W^- equals $66 - W^+$ at $n = 11$. Statistical software that returns $\min(W^+, W^-)$ by default will therefore print 7.0, 8.5 and 0.0 for the three comparisons. The exact p values are unaffected by the convention.

Reconciliation with the earlier 27-drive study

The figures reported here differ from those published earlier by the same group (Rzayeva et al. [7]), where a recovery rate of 91.8%, a temporal accuracy of 96.7% and a false discovery rate of 2.4% were obtained on 27 drives. Because the recovery pipeline is substantially the same, this difference requires an explicit account rather than a footnote. Three factors contribute, and they must be read together; the first two are definitional and the third is a property of the sample.

First, the metrics are not the same quantities. The 2.4% reported by Rzayeva et al. [7] is a frame-level rate, computed as the proportion of accepted frame candidates that did not correspond to genuine DHAV frames. Its counterpart in the present study is $FDR_{\text{frame}} = 2.0\%$ (Table 3), which is marginally lower, not higher. The 4.0% reported here is FDR_{file} , a metric introduced in this paper (Eq. 10b) that counts exported files failing a four-stage verification procedure and that has no counterpart in the earlier study. Comparing 4.0% with 2.4% therefore compares two different quantities, and on the quantity the two studies do share the present method performs slightly better.

Second, and more consequentially for the recovery rate and the temporal accuracy, the ground truth used here is stricter and is method-independent by construction. In the present study the expected number of recordings is taken exclusively from the recorder's own inventory, captured before acquisition. That inventory includes sessions which the recorder logged but whose data has since been wholly destroyed by cyclic over-writing; such sessions enter the denominator of Eq. (8) and are unrecoverable by any carver whatsoever. A reference constructed even partly from what is present on the medium excludes them and consequently reports a higher recovery rate for identical carver output. Temporal accuracy is affected in the same direction and more strongly, because a recovered file is credited here only when its channel identifier and its timestamp range match an inventory session to within ± 2 s - a tolerance that was not imposed in the earlier study [7]. The 7.7-percentage-point difference in temporal accuracy is dominated by this change of scoring rule rather than by any change in the sequencing algorithm, whose parameters are unchanged.

Third, the sample differs. The eleven drives used here span 1 to 8 TB across four NVR models, include motion-triggered and mixed recording modes, and had all been in operational service to the point where the circular buffer had wrapped. The dispersion visible in Table A1 shows that per-drive recovery varies materially with fill ratio and channel count, so a difference of a few percentage points between two samples of operational drives is within the range that sampling alone can produce at these sample sizes.

Separating these three contributions empirically would require re-scoring the 27-drive dataset of Rzayeva et al. [7] under the ground-truth procedure and the metric definitions adopted here. That re-scoring has not been performed, and it is identified in the Limitations as the specific comparison that would be needed before the two studies' headline figures could be placed on a single scale. Until it is performed, the present figures should be read as the more conservative of the two, and the difference should not be read as a degradation of the method.

Discussion

Research on Dahua CCTV systems is still limited. Dragonas et al. studied formatting, recording and authentication events in Dahua log files [8] and log records in the Hikvision file system [18] without addressing frame-level recovery. Among commercial tools, Magnet DVR Examiner shows inconsistent behaviour with DHFS4.1 and SalvationDATA VIP 2.0 must be configured manually against the firmware version, while open-source tools that rely on single-signature carving report false positive rates of 12–27% (see Results).

The four-level validation process addresses these weaknesses directly. Moving from a single DHAV signature to full checksum verification (Table 3) reduces the frame-level false discovery rate 13.5-fold at a cost of only 3.2 percentage points of recovery rate. In practical terms, at the end-to-end level this means that of 100 files delivered to the examiner approximately four require rejection at verification rather than the 27 that single-signature carving would produce at the frame level. This favourable trade-off is due largely to the combination of dual-signature validation with adaptive time sequencing (89.0% temporal accuracy), which supports an accurate chronology of events for alibi checking and witness corroboration. All comparisons with existing tools were confirmed as statistically significant by paired t-tests on the per-drive recovery rate (Table 6) and by Wilcoxon signed-rank tests (Table 7), although the width of the confidence intervals shows that the size of the advantage - as opposed to its direction - is estimated only coarsely at this sample size. The advantage over SalvationDATA VIP 2.0 in particular, with a mean difference of 0.80 percentage points and a confidence interval reaching only to 1.45, is statistically detectable but small enough that it should not be presented as a practically decisive superiority; the transparency of the method, rather than its recovery rate, is its main advantage over that tool.

The distinction between the frame-level and the file-level false discovery rate is worth emphasising for practitioners, because the two quantities answer different questions. *FDR_frame* measures how well the carving logic distinguishes real DHAV frames from coincidental byte patterns, and it is the correct quantity for comparing validation strategies. *FDR_file* measures how many of the artefacts actually delivered to an examiner will not survive scrutiny, and it is the correct quantity for estimating the review workload and for statements about evidential reliability. Reporting only the first, as is common in the carving literature, understates the burden on the examiner. The zero-fill condition illustrates the point sharply: there the two rates diverge by more than an order of magnitude, and a paper reporting only the frame-level figure would describe the method as performing almost as well on a wiped drive as on an intact one, which is the opposite of the operational reality.

Recovery following formatting is also of practical value (Table 5). Quick formatting eradicates only the metadata table and leaves the video frames in place, so they remain recoverable by signature-based carving. An overwrite that reaches the payload behaves quite differently: what survives is not degraded video but intact video in isolated fragments, which the pipeline recovers at frame level yet largely fails to assemble into files that pass verification. It follows that quickly formatted drives should be processed first, while drives subjected to overwriting are likely to require more costly techniques such as physical chip-off recovery or log-file analysis, and that on such drives the examiner should expect a high proportion of the exported files to be short, unmatched fragments requiring manual assessment rather than

outright spurious detections. In either case recovered files should be independently verified and cross-referenced against the timestamps of other evidence.

Finally, a word on what the transparency of the method does and does not deliver. Every decision rule of the pipeline - the signature pair, the size relation, the checksum computation, the gap threshold and its calibration - is published here in a form that permits independent re-implementation and re-execution on the same images, and this is a substantive advantage over tools whose matching logic is a trade secret [23-25]. It makes the procedure contestable: an opposing expert can reproduce the run, vary α , and demonstrate whether a disputed sequence is an artefact of the sequencing rule. That is a precondition for meaningful scrutiny. It is not, however, the same thing as admissibility. Whether recovered footage is admitted depends on the chain of custody, on the qualification and instruction of the examiner, on compliance with the procedural code of the jurisdiction and on the court's assessment of reliability in the individual case; the openness of an algorithm bears on the last of these and on none of the others. The claim advanced in this paper is accordingly limited to reproducibility and contestability of the technical procedure.

Limitations

The following limitations delimit the scope within which the reported figures should be interpreted. They are stated separately because each of them constrains generalisation in a different way.

Sample size. The study rests on 11 drives drawn from four NVR models of a single vendor. This is sufficient to establish the direction of the effects - every pairwise comparison in Table 6 is significant and is confirmed non-parametrically in Table 7, and the effect sizes are large - but not to estimate their magnitude precisely: the 95% confidence interval for the advantage over Scalpel spans 6.19 to 20.61 percentage points, and the interval for SalvationDATA VIP 2.0 reaches only 1.45 percentage points. No a priori power analysis was performed, and the study was not powered to detect differences smaller than roughly one percentage point. The per-drive results should therefore be read as indicative rather than as normative performance figures, and a multi-site study with at least 50 drives from a wider range of models would be required before the numbers reported here could be used as benchmark values.

Comparability with the earlier study. The reconciliation offered above attributes the difference between the present figures and those reported by Rzayeva et al. [7] to metric definitions, to the strictness of the ground-truth procedure and to sampling. That attribution is argued rather than measured: the earlier 27-drive dataset has not been re-scored under the procedures adopted here, and until it is, the relative contribution of the three factors remains an inference. No figure from the earlier study is used anywhere in the results of this paper, so the inference does not affect any reported value; it affects only the interpretation of the difference between the two studies.

Validity of the ground-truth reference. All recovery and temporal-accuracy figures are computed against an inventory produced by the recorder under examination. As set out in the Materials and methods section, that inventory is itself subject to cyclic over-writing, to inconsistency after abnormal shutdown, and - in adversarial cases - to manipulation through the device's own interface. The reported figures are therefore conditional on the integrity of the recorder log. No independent reference external to the recorder was available for these drives, and constructing one, for example by instrumenting a recorder over a controlled recording period, would materially strengthen future evaluations.

Extent of the overwrite in the zero-fill condition. The condition realised in this experiment is an incomplete bulk wipe, and its results are conditional on the proportion of the volume that was actually zeroed. The behaviour of the method as a function of that proportion was not characterised: no controlled series with defined overwrite fractions was constructed,

so the 34% recovery rate and the 41% file-level false discovery rate describe one point on an unmeasured curve rather than a general result for overwritten media. An exhaustive verified wipe was not tested and requires no test, since no signature-based method can recover from it.

Partially corrupted and partially overwritten data. The common real-world case is partial cyclic over-writing, in which fresh video overwrites arbitrary fractions of older recordings, and this was not varied systematically. The behaviour of the method between the baseline and the wipe condition is therefore unmeasured, and the 41% figure should not be read as an upper bound on the file-level false discovery rate for partially damaged drives.

Firmware versions. DHFS4.1 is a family of implementations rather than a single fixed on-disk format, and the semantics of individual header fields - in particular the extended-header flag and the way the checksum byte is computed - are known to vary between firmware builds. The drives used here came from a small number of firmware builds, which were recorded at acquisition but were not treated as an experimental factor, and no drive was reformatted and re-recorded under a different firmware version in order to isolate that factor. A firmware build that alters the header layout would defeat the size-consistency and checksum stages while leaving the signature stages intact, which would move the method's operating point back towards Level 2 performance. The performance figures reported in this paper should accordingly be understood as applying to the firmware builds tested.

RAID configurations. All 11 drives were single, directly attached devices imaged individually behind a write blocker. Enterprise Dahua recorders frequently use RAID-0, RAID-5 or RAID-6 arrays, in which DHAV frames are striped across member drives and individual frames are split at stripe boundaries. Applied to a single member of such an array, the proposed method would find header signatures whose matching footers lie on a different physical drive, so Level 2 onwards would reject the great majority of genuine frames. Stripe-order and stripe-size reconstruction would have to be performed before the method could be applied, and that step is outside the scope of this work.

Codecs. The payloads encountered were H.264 and H.265 elementary streams, but the proportion of each was not controlled and is not reported. This matters because the two codecs produce different I-frame to P-frame ratios and different frame-size distributions, which in turn affect the median inter-frame interval M in Eq. (1) and hence the behaviour of adaptive time sequencing. Proprietary variable-bitrate modes such as Dahua Smart Codec and H.264+, which lengthen the GOP during static scenes, would increase the variance of the inter-frame interval further. MJPEG, which some older channels still use, and newer codecs such as AV1 were not present in the dataset and were not tested.

Recording modes. The dataset is dominated by continuous recording. Motion-triggered, scheduled and alarm-triggered modes produce a qualitatively different temporal structure: many short sessions separated by genuine gaps whose length is determined by the activity in the scene rather than by the recorder configuration. Because the gap threshold of Eq. (1) is calibrated on the median inter-frame interval, a value of α tuned on predominantly continuous material is likely to merge distinct motion-triggered sessions, lowering temporal accuracy on such drives. The value of α was not re-tuned per recording mode.

Drive capacities. Capacities in the dataset spanned 1 to 8 TB, and processing time averaged 2.7 hours per drive on the forensic image. Capacity affects the results in two ways that were not characterised. First, on a larger volume the circular buffer wraps less often for a given retention period, so the degree of fragmentation and partial over-writing differs systematically from that on a small volume. Second, processing time scales with the volume that must be read, and the pipeline reads the entire device in 1 MB blocks; extrapolation to multi-terabyte and multi-drive arrays therefore cannot be made from the figure reported here.

Encrypted volumes and file-system versions. Encrypted DHFS4.1 volumes were outside the scope of the controlled experiment; where encryption is applied at the volume level,

signature detection fails at the first stage and the method does not apply without prior key recovery. Finally, the entire study concerns a single file-system version. DHFS4.1 was chosen because it is the version currently most widely deployed in the region, but earlier and later members of the DHFS family differ in their block-allocation strategy, and the results reported here do not transfer to them without re-validation.

Conclusion

This study presented and evaluated a forensic methodology for recovering video evidence from Dahua's proprietary DHFS4.1 file system, using 11 real NVR drives on which the proposed method and all three comparison tools were executed under identical conditions. The contribution of this paper is not the recovery pipeline in isolation, which builds on our own earlier work, but the validation-level ablation, the stricter and fully paired evaluation protocol, and the post-formatting/partial-overwrite study conducted around it; two results from that programme address the central difficulty, namely that the complex frame encapsulation and validation mechanisms of DHFS4.1 largely prevent conventional carving from recovering frames reliably. First, a four-level validation cascade reduced the frame-level false discovery rate 13.5-fold, from 27.0% to 2.0%, at a cost of only 3.2 percentage points of recovery rate, yielding an end-to-end file-level false discovery rate of 4.0%; the collision analysis presented here shows that the residual false positives are structural artefacts of cyclic over-writing rather than chance signature matches, which sets a clear direction for further filtering. Second, a systematic comparison of recovery after quick formatting and after an incomplete bulk wipe showed recovery rates of 76% and 34% respectively, and demonstrated that the sharp rise in the file-level false discovery rate under the second condition is caused by fragmentation of surviving payload rather than by any loss of discriminative power in the cascade.

Within the scope actually tested, the proposed methodology performs comparably to commercial tools such as Magnet DVR Examiner and SalvationDATA VIP 2.0, and differs from them principally in that every decision rule is published in this paper in enough detail to be independently re-implemented and challenged, even though the reference implementation itself is not currently released as public source code (see Data and Code Availability). Transparency of this kind is a precondition for meaningful scrutiny of a recovery procedure; it is not equivalent to procedural admissibility, which depends on chain of custody, examiner qualification and the evidentiary rules of the jurisdiction, none of which is affected by the openness of an algorithm.

Both mechanisms, dual-signature validation and adaptive time sequencing, were designed around the specific characteristics of Dahua systems, and because the methodology is open it can be verified and extended by other researchers. Its principal constraints are set out in the Limitations section: a sample of 11 drives from a single vendor, a small number of firmware builds, a single file-system version, single non-RAID devices, uncontrolled codec and recording-mode composition, a ground truth derived from the recorder's own log, and an uncharacterised response to partial over-writing. Future work will extend the methodology to further firmware versions, evaluate machine-learning approaches such as neural-network frame classification, add support for RAID arrays and multi-terabyte volumes, construct a recorder-independent ground-truth reference, and carry out the controlled over-writing series needed to characterise behaviour between the intact and the fully overwritten condition.

In conclusion, within the scope defined above - single, non-RAID drives from four Dahua NVR models running the tested DHFS4.1 firmware builds, with predominantly continuous H.264/H.265 recording - the proposed methodology is an effective and methodologically transparent tool for video recovery from Dahua surveillance systems, and is re-implementable by another researcher from the parameters, thresholds and pseudocode-level description given in this paper; see the Data and Code Availability statement for what is and is

not currently released. Its transferability beyond that scope remains to be established, and the figures reported here should be treated as indicative rather than as benchmark values.

Ethics and data-governance statement

The 11 decommissioned operational NVR hard drives were obtained under institutional authorization for experimental and scientific research within project IRN BR24993232. Because the experimental protocol focused strictly on low-level file-system structure, signature parsing, and statistical recovery benchmarks without human-subject intervention, profiling, or re-identification, formal institutional ethics review was not required. In compliance with data-privacy principles, the disk images are stored on an isolated, air-gapped forensic system with restricted access, and no video footage, recognizable faces, license plates, or identifiable personal artifacts are shared or reproduced in this publication.

Data and code availability

The reference implementation of the pipeline described in this paper is not currently released as public source code. Every threshold, parameter and decision rule needed to re-implement it is nonetheless given explicitly in the text: the four validation levels and the exact fields they check (Materials and methods), the gap-threshold equation and the value of α used (Eq. 1), the minimum-duration threshold $T_{\min}$, the three mandatory verification commands with their exact invocation (ffprobe, ffmpeg -v error, and the ± 2 s correspondence rule), and the checksum and signature-collision derivations (Materials and methods). The per-drive results underlying every aggregate figure in this paper are given in full in Table A1, which is sufficient to reproduce the statistical analysis of the Results section (Tables 6 and 7) independently of the raw disk images. The raw forensic images themselves are not released, for the reasons given in the Ethics and data-governance statement above. Authors will consider reasonable requests for the underlying per-level frame counts and the analysis scripts used to produce Tables 6 and 7.

Appendix A. Per-drive results

Table A1 gives the per-drive values underlying every aggregate figure in Tables 2, 5 and 6, so that the mean differences, the dispersion of the differences and the outcome of the statistical tests can be verified independently.

Table A1. Per-drive recovery rate for the proposed method and the three comparison tools, temporal accuracy and file-level false discovery rate of the proposed method, and measured extent of the zero-fill condition ($n = 11$)

Drive	RR proposed (%)	RR Magnet (%)	RR Salvation (%)	RR Scalpel (%)	TA proposed (%)	FDR file proposed (%)	RR after quick format (%)	RR after zero-fill (%)	Zeroed sectors (%)
D1	89.4	92.3	90.8	64.7	90.2	3.4	77.2	35.2	56.5
D2	88.6	82.1	87.0	69.6	89.4	3.8	76.5	33.1	59.0
D3	86.2	80.2	84.7	70.8	87.1	4.7	73.8	29.8	62.5
D4	91.5	89.8	90.1	56.5	91.8	3.1	79.4	42.1	49.0
D5	85.1	82.6	84.0	84.5	86.5	5.2	72.1	26.5	65.5
D6	91.8	92.0	90.4	91.2	92.4	2.9	80.2	43.2	48.0
D7	88.7	86.5	87.1	76.4	89.1	4.1	76.8	34.6	57.0
D8	85.5	80.2	85.4	68.1	86.8	4.9	73.5	28.5	63.5
D9	87.9	87.1	86.7	86.6	88.5	4.3	75.9	31.8	60.0
D10	88.2	84.9	87.6	80.3	89.2	3.7	76.1	33.4	58.5
D11	85.1	75.1	85.4	71.9	88.0	3.9	74.5	35.8	58.5

Drive	RR proposed (%)	RR Magnet (%)	RR Salvation (%)	RR Scalpel (%)	TA proposed (%)	FDR_file proposed (%)	RR after quick format (%)	RR after zero-fill (%)	Zeroed sectors (%)
Mean	88.0	84.8	87.2	74.6	89.0	4.0	76.0	34.0	58.0

Acknowledgment

This study was carried out with the financial support of the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan under Contract №388/PTF-24-26 dated 1 October 2024, within the scientific project IRN BR24993232 "Development of innovative technologies for conducting digital forensic investigations using intelligent software-hardware complexes."

References

- [1] MarketsandMarkets. (2025, November). *Video surveillance market size, share & analysis, 2031*. <https://www.marketsandmarkets.com/Market-Reports/video-surveillance-market-645.html>
- [2] Grand View Research. (n.d.). *Video surveillance market size & share report, 2026–2033*. Retrieved September 21, 2026, from <https://www.grandviewresearch.com/industry-analysis/video-surveillance-market-report>
- [3] Memoori Research. (2026, June 16). *Video surveillance market leaders 2026 compared side-by-side*. <https://memoori.com/video-surveillance-market-leaders-2026/>
- [4] Karbayeva, D. (2026, April 29). *Bolee 27 tysyach kamer v Kazakhstane osnashcheny elementami II – MVD* [More than 27,000 cameras in Kazakhstan are equipped with AI elements – Ministry of Internal Affairs]. Kazinform. <https://www.inform.kz/ru/bolee-27-tisyach-kamer-v-kazakhstane-osnasheni-elementami-ii-mvd-a5f490>
- [5] Brookman, F., & Jones, H. (2022). Capturing killers: The construction of CCTV evidence during homicide investigations. *Policing and Society*, 32(2), 125–144. <https://doi.org/10.1080/10439463.2021.1879075>
- [6] Jung, Y., & Wheeler, A. P. (2023). The effect of public surveillance cameras on crime clearance rates. *Journal of Experimental Criminology*, 19(1), 143–164. <https://doi.org/10.1007/s11292-021-09477-8>
- [7] Rzaeva, L., Shayakhmetov, M., Atanbayev, Y., Budenov, R., & Mutaher, H. (2025). Automated forensic recovery methodology for video evidence from Hikvision and Dahua DVR/NVR systems. *Information*, 16(11), 983. <https://doi.org/10.3390/info16110983>
- [8] Dragonas, E., Lambrinoudakis, C., & Kotsis, M. (2024). IoT forensics: Exploiting log records from the DAHUA technology CCTV systems. *Journal of Forensic Sciences*, 69(1), 117–130. <https://doi.org/10.1111/1556-4029.15401>
- [9] 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. <https://doi.org/10.1109/TIFS.2021.3118876>
- [10] van der Meer, V., van den Bos, J., Jonker, H., & Dassen, L. (2024). Problem solved: A reliable, deterministic method for JPEG fragmentation point detection. *Forensic Science International: Digital Investigation*, 48, 301687. <https://doi.org/10.1016/j.fsidi.2023.301687>
- [11] Mittal, G., Korus, P., & Memon, N. (2021). FiFTy: Large-scale file fragment type identification using convolutional neural networks. *IEEE Transactions on Information Forensics and Security*, 16, 28–41. <https://doi.org/10.1109/TIFS.2020.3004266>

- [12] Forensic Focus. (2016–2021). *Forensic video data recovery tools for CCTV DVRs* [Online forum thread]. <https://www.forensicfocus.com/forums/general/forensic-video-data-recovery-tools-for-cctv-dvrs/>
- [13] Fegan, J. (2025, January 28). *Challenges in recovering data from modern surveillance systems*. Envista Forensics. <https://www.envistaforensics.com/knowledge-center/insights/articles/challenges-in-recovering-data-from-modern-surveillance-systems/>
- [14] Hu, H. (2025, April 30). *How VIP2.0 enhances video forensics*. SalvationDATA. <https://www.salvationdata.com/product-tips/how-vip2-0-enhances-video-forensics/>
- [15] Richard, G. G., III, & Marziale, L. (n.d.). *Scalpel* [Computer software]. The Sleuth Kit. <https://github.com/sleuthkit/scalpel>
- [16] Yoon, J., & Hwang, S. (2026). Forensic analysis of video data deletion and recovery in Honeywell surveillance file system. *Forensic Science International: Digital Investigation*, 57, 302116. <https://doi.org/10.1016/j.fsidi.2026.302116>
- [17] Byun, J., Shim, K.-S., & Kim, Y. W. (2026). Video data recovery from slack space of pre-allocated video files. *Journal of Forensic Sciences*. Advance online publication. <https://doi.org/10.1111/1556-4029.70412>
- [18] Dragonas, E., Lambrinoudakis, C., & Kotsis, M. (2023). IoT forensics: Exploiting unexplored log records from the HIKVISION file system. *Journal of Forensic Sciences*, 68(6), 2002–2011. <https://doi.org/10.1111/1556-4029.15349>
- [19] Rzayeva, L., Shayakhmetov, M., Konakbayev, O., Jussupova, G. G., Seniushin, I., & Tasbolat, A. (2026). Forensic video recovery from multi-channel analog DVR systems: Channel demultiplexing and temporal reconstruction from interleaved DHAV streams. *Information*, 17(5), 493. <https://doi.org/10.3390/info17050493>
- [20] International Telecommunication Union. (2026). *Advanced video coding for generic audiovisual services* (ITU-T Recommendation H.264, 06/2026). <https://www.itu.int/rec/T-REC-H.264-202606-I/en>
- [21] International Telecommunication Union. (2026). *High efficiency video coding* (ITU-T Recommendation H.265, 01/2026). <https://www.itu.int/rec/T-REC-H.265-202601-I/en>
- [22] Bross, B., Wang, Y.-K., Ye, Y., Liu, S., Chen, J., Sullivan, G. J., & Ohm, J.-R. (2021). Overview of the Versatile Video Coding (VVC) standard and its applications. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(10), 3736–3764. <https://doi.org/10.1109/TCSVT.2021.3101953>
- [23] Hargreaves, C., Nelson, A., & Casey, E. (2024). An abstract model for digital forensic analysis tools: A foundation for systematic error mitigation analysis. *Forensic Science International: Digital Investigation*, 48, 301679. <https://doi.org/10.1016/j.fsidi.2023.301679>
- [24] Garrett, B. L., & Rudin, C. (2023). Interpretable algorithmic forensics. *Proceedings of the National Academy of Sciences*, 120(41), e2301842120. <https://doi.org/10.1073/pnas.2301842120>
- [25] Scientific Working Group on Digital Evidence. (2025). *SWGDE best practices for computer forensic examinations* (Document 18-F-001, Version 2.0). <https://swgde.org/wp-content/uploads/2025/09/2025-07-28-Best-Practices-for-Computer-Forensic-Examinations-18-F-001-2.0.pdf>