Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

RansomShield: Threat-score-driven adaptive backup with multi-modal behavioural ransomware detection

SEGUN VICTOR OLOJA
University of Medical Sciences Ondo
RANTIOLA FIDELIS FAMUTIMI
University of Medical Sciences Ondo
OLANREWAJU AYODEJI OLOJA
olanrewajuobolo@gmail.com

University of Medical Sciences Ondo
JOHN TOPE BALOGUN
University of Medical Sciences Ondo
EZEKIEL AYOMIDE SAMSON
University of Medical Sciences Ondo

Research Article

Keywords: ransomware detection, adaptive backup, ensemble learning, memory forensics, Shannon entropy, immutable snapshots

Posted Date: November 5th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11337939/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Ransomware defences usually treat detection and backup as separate systems: detectors raise alerts while backups run on fixed schedules, regardless of the current threat. This paper presents RansomShield, a proof-of-concept architecture in which a continuous threat score from a multi-modal detector directly governs backup frequency, snapshot immutability and isolation. The detector combines four signals accounting for a random forest on memory-forensic features, a rule-based API-call sequence scorer, a storage I/O anomaly analyser, and a file-type-aware Shannon entropy monitor. Their outputs are fused by confidence-weighted averaging with exponential smoothing. A four-level controller with hysteresis maps the smoothed score to a backup policy. The components were built using CIC-MalMem-2022, RanSMAP, NapierOne and the Mendeley Ransomware/Benignware System Calls dataset. On CIC-MalMem-2022, the memory detector achieved a cross-validated F1 score of 0.9999 ± 0.0002 for binary classification and 76.07% accuracy for 16-class classification. In simulated attacks modelled on seven ransomware families, all seven were detected after a mean of 13.4 file-system events and escalated to immutable, air-gapped snapshots at 30-second intervals, while a benign scenario produced a 5.6% event-level false-positive rate.

1. Introduction

Ransomware encrypts a victim's files or locks access to computing resources and demands payment, usually in cryptocurrency, for restoration. Unlike malware aimed at data theft or disruption, it monetises an attack directly by holding data hostage, which pressures victims to pay quickly (Connolly et al., 2020). It has disrupted hospitals, pipelines and supply chains. Nigeria's national computer emergency response team has warned of escalating ransomware attacks on IT, telecommunications, financial and healthcare service providers (ngCERT, 2024), and industry reporting names ransomware as the most significant cyber threat to Nigerian organisations (Check Point Software Technologies, 2024).

Defences fall into three broad categories: signature-based detection, behavioural detection and backup (Al-rimy et al., 2018). Signature-based systems match files against known threats and cannot reliably recognise new or polymorphic variants (Moussaileb et al., 2021; Gandotra et al., 2014). Behavioural systems monitor runtime activity, such as file-system operations, API calls and memory use, and can detect previously unseen ransomware, although they may produce false positives and add overhead (Ucci et al., 2019; Berrueta et al., 2019). Backup is the last line of defence, but modern ransomware actively targets it by deleting Volume Shadow Copies, encrypting network shares and tampering with synchronisation (Kharraz et al., 2016; McIntosh et al., 2021). Backups taken on a fixed schedule at predictable locations are therefore vulnerable. The 3-2-1 rule predates ransomware and does not assume an adversary who targets backups, and immutable, air-gapped storage resists tampering from a compromised host but is typically configured statically.

Within behavioural detection, several systems have become reference points. CryptoLock (Scaife et al., 2016) was among the first real-time detectors, combining file-type changes, write entropy and similarity

between original and modified files. ShieldFS (Continella et al., 2016) profiles process-level file-system activity with a random forest and restores files using a copy-on-write mechanism. UNVEIL (Kharraz et al., 2016) analyses ransomware in sandboxed environments populated with decoy files. RWGuard (Mehnaz et al., 2018) combines decoy files, process and file-system monitoring, and learning of a user's own encryption behaviour so that benign changes are not flagged. ElderRan (Sgandurra et al., 2016) applies machine learning to features from dynamic analysis, including API calls and file and registry operations. PayBreak (Kolodenker et al., 2017) takes a different approach: it intercepts cryptographic function calls to retain encryption keys so files can be decrypted without paying.

Surveys recommend combining signals because no single modality is reliable across all families (Berrueta et al., 2019; Oz et al., 2022). Ensemble methods reduce variance and can compensate for the weaknesses of individual classifiers (Dietterich, 2000), and weighted fusion relates to the stacking literature (Wolpert, 1992). Entropy-based detection exploits the near-maximal randomness of encrypted data (Shannon, 1948; Scaife et al., 2016), but it is confounded by compressed and media files and by legitimate encryption, so file-type baselines are needed (Davies et al., 2022).

Most of these systems rely on one or a small number of signal types, and their response to detection is either absent or binary: a recovery mechanism is triggered or it is not. Detectors with a recovery component, such as ShieldFS, use a binary trigger (Continella et al., 2016). In practice, the confidence of the detector does not shape how aggressively data is protected, although a low-confidence early warning and a high-confidence confirmed attack call for different responses. Table 1 summarises how the reviewed systems relate to this work. The comparison is qualitative since the systems were evaluated on different datasets and settings, so their reported accuracies are not directly comparable and are not tabulated here.

Table 1
Qualitative comparison of related systems.

System	Main detection signal(s)	Recovery / backup response
CryptoLock (Scaife et al., 2016)	File-type changes, write entropy, similarity	None
ShieldFS (Continella et al., 2016)	Process-level file-system I/O (random forest)	Copy-on-write; binary trigger
UNVEIL (Kharraz et al., 2016)	Decoy files, I/O behaviour in sandbox	None
ElderRan (Sgandurra et al., 2016)	API calls, file and registry operations	None
PayBreak (Kolodenker et al., 2017)	Interception of cryptographic calls	Key recovery
RWGuard (Mehnaz et al., 2018)	Decoy files, process and file-system monitoring, learned user encryption behaviour	Detection-focused
RansomShield (this work)	Memory, API sequence, storage I/O, entropy	Graduated four-level response driven by a continuous score

This paper addresses that gap with RansomShield, a proof-of-concept architecture that closes the loop between detection and backup. The contributions are:

1. A closed-loop architecture in which a continuous threat score drives backup behaviour (frequency, snapshot mode and isolation level).
2. A four-signal detection engine combining memory forensics, API-call sequences, storage I/O patterns and file entropy through confidence-weighted fusion with exponential smoothing.
3. A four-level adaptive backup controller with hysteresis that replaces the binary trigger with graduated, proportional protection.
4. A proof-of-concept evaluation that uses four public datasets for the components and simulated attacks modelled on seven ransomware families for the integrated system.

The remainder of the paper is organised as follows. Section 2 describes the method, Section 3 reports results, Section 4 discusses findings and limitations, and Section 5 concludes.

2. Method

2.1 Architecture

RansomShield has five layers: data sources, feature extraction, detection engine, threat assessment and response. Four telemetry sources (memory snapshots, API-call traces, storage I/O traces and file-write byte streams) each feed a dedicated feature extractor and detector. The threat-assessment layer fuses

the four outputs into a single smoothed score, which the response layer passes to the adaptive backup controller. Figure 1 shows the architecture.

2.2 Datasets

Four public datasets were used (Table 2), each to build one component.

Table 2
Datasets and their roles.

Dataset	Content	Role
CIC-MalMem-2022 (Carrier et al., 2022)	Memory-forensic records with 55 features from benign and malicious memory dumps (spyware, ransomware, trojan; 15 malware families). Official release: 58,596 records (29,298 per class); working copy used here: 58,058 records (29,227 benign, 28,831 malicious)	Memory detector
RanSMAP (Hirano and Kobayashi, 2025)	Open dataset of storage and memory access patterns captured by a hypervisor during executions of ransomware and benign applications	I/O analyser (34 features derived from a subset of 200 experiments, 140 ransomware)
NapierOne (Davies et al., 2022)	Mixed file corpus of about 100 data subsets covering 44 file types (over 500,000 files in total)	Per-file-type entropy baselines
Mendeley Ransomware/Benignware System Calls dataset (Mendeley Data)	System-call data from 270 ransomware and 270 benign samples run in a Windows 10 virtual machine	API sequence scorer

2.3 Detection components

A random forest (200 trees, Gini criterion, unrestricted depth, balanced class weights, fixed random seed) classifies memory-forensic features. Preprocessing converted features to float64 and removed three constant features and twelve features with pairwise correlation above 0.95, reducing 55 features to 40. Performance was evaluated on a stratified hold-out test set and by 10-fold stratified cross-validation on the training portion. A 16-class model (15 malware families plus the benign class) was also evaluated.

The scorer is rule-based; no model is trained. Over a sliding window of the last 500 API calls it computes the frequency of calls in six categories and combines them with fixed weights: cryptographic operations (0.35), file destruction (0.20), file enumeration (0.15), shadow-copy deletion (0.15), persistence (0.10) and network spread (0.05). When both the cryptographic and enumeration indicators exceed 0.3, the score is multiplied by 1.5, reflecting the typical enumerate-then-encrypt sequence.

From storage traces the analyser computes 34 features over a configurable window (30 s by default) and combines five anomaly components with fixed weights: write entropy (0.35), write-rate anomaly relative to a baseline (0.20), write-to-read ratio (0.15), delete rate (0.15) and file-type diversity (0.15).

For each write, Shannon entropy

$H = -\Sigma\, p(x_i) \cdot \log_2 p(x_i)$,

summed over the 256 possible byte values is computed over byte values and compared with a baseline for the file type derived from NapierOne. The monitor applies two criteria: an absolute threshold of 7.5 bits/byte and a deviation of at least 1.5 bits/byte from the baseline of the file type. Figure 2 shows the baselines. Compressed and media formats (ZIP, MP3, JPEG, PNG) already sit at or near the threshold, so an absolute threshold alone cannot separate them from encrypted data; this is why per-type baselines are used, and it also means the entropy signal is weakest for exactly these file types (Davies et al., 2022). Because NapierOne does not provide post-encryption samples for every file type, some thresholds were set from theoretical considerations.

2.4 Ensemble fusion

Let S_i be the score of component i and c_i its confidence. With fusion weights w = (0.35, 0.25, 0.25, 0.15) for the memory, API, I/O and entropy components, the raw threat score at time t is

$\text{ThreatScore}(t) = \Sigma_i\, (w_i \cdot S_i \cdot c_i) \,/\, \Sigma_i\, (w_i \cdot c_i)$

The score is smoothed with an exponential moving average,

$\tilde{S}(t) = \alpha \cdot \text{ThreatScore}(t) + (1 - \alpha) \cdot \tilde{S}(t - 1)$

with $\alpha = 0.3$, to prevent transient spikes from triggering unnecessary escalation. The fusion weights were set from component-level validation performance and were not optimised automatically.

2.5 Adaptive backup controller

The controller is a four-state machine driven by $\tilde{S}(t)$ (Table 3). Escalation thresholds are 0.30, 0.60 and 0.85; de-escalation thresholds are 0.05 lower (0.25, 0.55, 0.80), providing hysteresis so the posture is not relaxed on momentary dips during an active attack. Figure 3 shows the state machine.

Table 3
Threat levels and backup policy.

Level	Score range	Backup interval	Snapshot mode	Isolation	Additional actions
GREEN	0.00–0.29	30 min	Incremental	Standard	Normal operation
YELLOW	0.30–0.59	5 min	Full snapshot	Read-only	Monitoring alerts
ORANGE	0.60–0.84	30 s	Immutable	Air-gapped	Sandbox suspicious processes
RED	0.85–1.00	Continuous (1 s)	Immutable	Air-gapped	Suspend processes; initiate rollback

2.6 Evaluation protocol

The memory detector was evaluated with accuracy, precision, recall, F1, AUC-ROC, Matthews correlation coefficient and 10-fold cross-validated F1. The API scorer and I/O analyser were assessed descriptively on their datasets.

An attack simulator generated file-system events from behavioural profiles of seven families (WannaCry, Locky, Conti, Ryuk, Maze, REvil, Dharma). Each attack scenario had 100 benign events followed by 200 attack events (high-entropy writes, file enumeration and deletions). A benign scenario of 500 events measured false positives. An attack was counted as detected when the smoothed score first reached the YELLOW threshold (0.30). Detection latency is the number of events between attack onset and first detection. The simulation uses behavioural profiles rather than live ransomware binaries.

3. Results

3.1 Memory detector

On CIC-MalMem-2022 the random forest achieved perfect scores on the hold-out set (accuracy, precision, recall, F1, AUC and MCC all 1.0000), with a 10-fold cross-validated F1 of 0.9999 ± 0.0002 (see Table 4). The narrow variance indicates stable behaviour across partitions. Such near-ceiling scores are common on this dataset and should be read with caution (Section 5.2).

Table 4
Binary classification on CIC-MalMem-2022.

Metric	Value
Accuracy / Precision / Recall / F1 / AUC-ROC / MCC	1.0000
Cross-validated F1 (mean ± SD)	0.9999 ± 0.0002

Feature importance (Gini) was concentrated in service-scan and handle features (see Table 5). The three service-scan features (kernel drivers, number of services, shared-process services) together account for 42.80% of total importance, and the mutant (mutex) handle count ranks fourth, consistent with malware's use of mutexes for synchronisation (see Fig. 4).

Table 5
Top ten features by Gini importance.

Rank	Feature	Category	Importance
1	svcscan.kernel_drivers	Service scan	0.1576
2	svcscan.nservices	Service scan	0.1364
3	svcscan.shared_process_services	Service scan	0.1340
4	handles.nmutant	Handles	0.1203
5	dlllist.avg_dlls_per_proc	DLL list	0.0899
6	dlllist.ndlls	DLL list	0.0445
7	pslist.avg_handlers	Process list	0.0440
8	handles.nsection	Handles	0.0430
9	handles.nhandles	Handles	0.0380
10	handles.nevent	Handles	0.0320

For 16-class classification (15 malware families and benign), accuracy was 76.07% with a weighted F1 of 75.83%. The lower figure reflects the behavioural similarity between families, but shows that memory features retain information useful for attributing a detection to a family.

3.2 API sequence scorer and I/O analyser

Applied to the per-sample API-call data of the 540 Mendeley samples, the rule-based API scorer separated all 270 ransomware samples from all 270 benign samples. Because the scorer has fixed hand-set rules and was not evaluated on held-out data, this result describes how distinctive ransomware API patterns are in this dataset; it is not an estimate of generalisation performance.

For the I/O analyser, 34 features were extracted from 200 RanSMAP experiments (140 ransomware). Inspection of the feature distributions showed substantial overlap between ransomware and benign experiments for both ATA write entropy and memory write entropy, with only a few ransomware experiments at higher values; individual entropy features were therefore weakly discriminative on their own. Read- and write-related entropy features were strongly correlated, suggesting scope for dimensionality reduction. No standalone classification performance was computed for this component, and its output was used as the I/O component of the ensemble.

3.3 Integrated system under simulated attack

All seven simulated attacks were detected (see Table 6). Mean detection latency was 13.4 events, ranging from 11 (Locky) to 16 (Dharma). In each scenario the smoothed score reached 0.762, escalating the controller to ORANGE, and no scenario reached RED (0.85). Figure 5 shows the score trajectories. Benign activity fluctuates mostly between 0.1 and 0.4, whereas every attack scenario shows a rise at attack onset (event 100) and a plateau near 0.75; the seven attack trajectories are very similar in shape, which reflects the shared structure of the simulated profiles.

Table 6
Detection and backup escalation per family.

Family	Events to detect	Max score	Events to ORANGE	Final level	Backup actions
WannaCry	15	0.762	42	ORANGE	18
Locky	11	0.762	38	ORANGE	21
Conti	13	0.762	40	ORANGE	19
Ryuk	13	0.762	41	ORANGE	19
Maze	12	0.762	39	ORANGE	20
REvil	14	0.762	41	ORANGE	18
Dharma	16	0.762	43	ORANGE	17

In the 500-event benign scenario, 28 events exceeded the YELLOW threshold, an event-level false-positive rate of 5.6%. The maximum benign score was 0.438, below the ORANGE threshold (0.60), so benign activity triggered at most a temporary increase in backup frequency and never immutable or air-gapped snapshots. At the ORANGE level the controller issued between 17 and 21 backup actions per attack at 30-second intervals. These are counts of backup actions triggered; the study did not measure whether the resulting snapshots would have preserved clean file states.

4. Discussion

4.1 Findings

The results show that the proposed loop is operable. A continuous score from four signals can be turned into graduated backup behaviour, and the controller escalated through GREEN, YELLOW and ORANGE in every simulated attack while benign activity stayed within YELLOW. The memory detector is the strongest component, with service-configuration and mutex features dominating, this is in line with Carrier et al., (2022). The maximum score of 0.762 was identical across families, so the RED level was never reached. This is consistent with conservative fusion weights and EMA smoothing, but the near-identical trajectories in Fig. 5 also indicate that the simulated profiles drive the detectors in the same way. However, sensitivity to α and the weights should be examined before conclusions about family-specific behaviour are drawn. The 5.6% false-positive rate, though confined to YELLOW, would impose

backup overhead in production and indicates that dynamic or environment-specific threshold calibration is needed.

4.2 Limitations

- The system-level evaluation uses simulated behavioural profiles, not live ransomware in a sandbox, so latency and false-positive figures may differ in deployment.
- No ablation was performed, so the contribution of each component, and of fusion compared with single modalities, is not quantified.
- Recovery was not measured, so the benefit of adaptive over fixed-schedule backup is not demonstrated empirically.
- The I/O analyser's inputs overlapped substantially between ransomware and benign experiments in RanSMAP, and no standalone performance was measured, so its contribution to the ensemble is untested.
- The API scorer is rule-based and was evaluated on one dataset of 540 samples without a held-out split; it is not evidence of generalisation. The public dataset documentation describes per-API call counts for each sample, which limits what can be said about the temporal ordering of calls.
- CIC-MalMem-2022 is general malware data, not ransomware-only. Its official release is balanced, reportedly by oversampling the benign class (SMOTE), and near-duplicate records are common, so perfect scores may partly reflect dataset characteristics; deduplicated, family-held-out evaluation is needed.
- Fusion weights were set manually, and some entropy thresholds were derived theoretically because NapierOne lacks post-encryption samples for every file type.
- The simulation was run once per family; variation across repeated runs was not quantified.

5. Conclusion and future work

RansomShield demonstrates, as a proof of concept, that detection confidence can govern backup behaviour through a continuous threat score and a four-level controller with hysteresis. The memory detector reached a cross-validated F1 of 0.9999 on CIC-MalMem-2022, all seven simulated families were detected after a mean of 13.4 events and escalated to immutable, air-gapped snapshots, and benign activity stayed within the YELLOW level with a 5.6% event-level false-positive rate.

Future work should: (i) validate the system against live ransomware in a sandboxed network; (ii) run an ablation of each detector and of the fusion; (iii) measure recovery-point age and data loss against fixed-schedule backup; (iv) replace the rule-based API scorer with a model trained on larger trace sets and evaluated on held-out families; (v) optimise fusion weights and smoothing automatically and adapt thresholds per environment; (vi) integrate with SIEM and SOAR platforms; and (vii) conduct field evaluations in Nigerian organisations, where resource constraints make low-cost integrated protection especially relevant.

Declarations

Declaration of competing interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability: The datasets used are publicly available: CIC-MalMem-2022 (Carrier et al., 2022), RanSMAP (Hirano and Kobayashi, 2025), NapierOne (Davies et al., 2022) and the Mendeley Ransomware/Benignware System Calls dataset (Mendeley Data). The simulation code and result files are available on request.

Author Contribution

S.V, R.F and O.A conceived of the presented idea.S. V, J.T and E.A sourced for literature review and gathering of necessary material.S.V. R.F, O.A. J.T and E.A contributed to the design and implementation of the research, to the analysis of the results and to the writing of the manuscript.S.V. R.F, O.A. J.T and E.A discussed the results and contributed to the final manuscript.

References

1. Al-rimy, B.A.S., Maarof, M.A., Shaid, S.Z.M.: Ransomware threat success factors, taxonomy, and countermeasures: A survey and research directions. Computers Secur. **74**, 144–166 (2018). https://doi.org/10.1016/j.cose.2018.01.001
2. Berrueta, E., Morato, D., Magaña, E., Izal, M.: A survey on detection techniques for cryptographic ransomware. IEEE Access. **7**, 144925–144944 (2019). https://doi.org/10.1109/ACCESS.2019.2945839
3. Carrier, T., Victor, P., Tekeoglu, A., Lashkari, A.H.: Detecting obfuscated malware using memory feature engineering. *Proceedings of the 8th International Conference on Information Systems Security and Privacy (ICISSP)*, 177–188. (2022). https://doi.org/10.5220/0010908200003120
4. Check Point Software Technologies: Africa Cyber Security Outlook 2024. Check Point Software Technologies (2024)
5. Connolly, L.Y., Wall, D.S., Lang, M., Oddson, B.: An empirical study of ransomware attacks on organizations: An assessment of severity and salient factors affecting vulnerability. J. Cybersecur. **6**(1), tyaa023 (2020). https://doi.org/10.1093/cybsec/tyaa023
6. Continella, A., Guagnelli, A., Zingaro, G., De Pasquale, G., Barenghi, A., Zanero, S., Maggi, F.: ShieldFS: A self-healing, ransomware-aware filesystem. *Proceedings of the 32nd Annual Conference on Computer Security Applications (ACSAC)*, 336–347. (2016). https://doi.org/10.1145/2991079.2991110

7. Davies, S.R., Macfarlane, R., Buchanan, W.J.: NapierOne: A modern mixed file data set alternative to Govdocs1. Forensic Sci. International: Digit. Invest. **40**, 301330 (2022). https://doi.org/10.1016/j.fsidi.2021.301330
8. Dietterich, T.G.: Ensemble methods in machine learning. *Multiple Classifier Systems (MCS 2000)*, Lecture Notes in Computer Science 1857, 1–15. (2000). https://doi.org/10.1007/3-540-45014-9_1
9. Gandotra, E., Bansal, D., Sofat, S.: Malware analysis and classification: A survey. J. Inform. Secur. **5**(2), 56–64 (2014). https://doi.org/10.4236/jis.2014.52006
10. Hirano, M., Kobayashi, R.: RanSMAP: Open dataset of ransomware storage and memory access patterns for creating deep learning based ransomware detectors. (2025)
11. Kharraz, A., Arshad, S., Mulliner, C., Robertson, W., Kirda, E.: UNVEIL: A large-scale, automated approach to detecting ransomware. *Proceedings of the 25th USENIX Security Symposium*, 757–772. (2016)
12. Kolodenker, E., Koch, W., Stringhini, G., Egele, M.: PayBreak: Defense against cryptographic ransomware. *Proceedings of the 2017 ACM Asia Conference on Computer and Communications Security (ASIA CCS)*, 599–611. (2017). https://doi.org/10.1145/3052973.3053035
13. McIntosh, T., Kayes, A.S.M., Chen, Y.-P.P., Ng, A., Watters, P.: Ransomware mitigation in the modern era: A comprehensive review, research challenges, and future directions. ACM Comput. Surveys. **54**(9), 1–36 (2021). https://doi.org/10.1145/3479393
14. Mehnaz, S., Mudgerikar, A., Bertino, E.: RWGuard: A real-time detection system against cryptographic ransomware. *Research in Attacks, Intrusions, and Defenses (RAID 2018)*, 114–136. (2018). https://doi.org/10.1007/978-3-030-00470-5_6
15. Mendeley Data: *Ransomware/Benignware System Calls* [Data set]. https://data.mendeley.com/datasets/kbt8xt3678
16. Moussaileb, R., Cuppens, N., Lanet, J.-L., Le Bouder, H.: A survey on Windows-based ransomware taxonomy and detection mechanisms. ACM Comput. Surveys. **54**(6), 1–36 (2021). https://doi.org/10.1145/3453153
17. ngCERT: Escalation of ransomware attack in Nigeria. Nigeria Computer Emergency Response Team (2024). [Advisory]
18. Oz, H., Aris, A., Levi, A., Uluagac, A.S.: A survey on ransomware: Evolution, taxonomy, and defense solutions. ACM Comput. Surveys. **54**(11s), 1–37 (2022). https://doi.org/10.1145/3514229
19. Scaife, N., Carter, H., Traynor, P., Butler, K.R.B.: CryptoLock (and drop it): Stopping ransomware attacks on user data. *Proceedings of the 36th IEEE International Conference on Distributed Computing Systems (ICDCS)*, 303–312. (2016). https://doi.org/10.1109/ICDCS.2016.46
20. Sgandurra, D., Muñoz-González, L., Mohsen, R., Lupu, E.C.: Automated dynamic analysis of ransomware: Benefits, limitations and use for detection. arXiv:1609.03020. (2016). https://doi.org/10.48550/arXiv.1609.03020
21. Shannon, C.E.: A mathematical theory of communication. Bell Syst. Tech. J. **27**(3), 379–423 (1948). https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

22. Ucci, D., Aniello, L., Baldoni, R.: Survey of machine learning techniques for malware analysis. Computers Secur. **81**, 123–147 (2019). https://doi.org/10.1016/j.cose.2018.11.001

23. Wolpert, D.H.: Stacked generalization. Neural Netw. **5**(2), 241–259 (1992). https://doi.org/10.1016/S0893-6080(05)80023-1

Figures

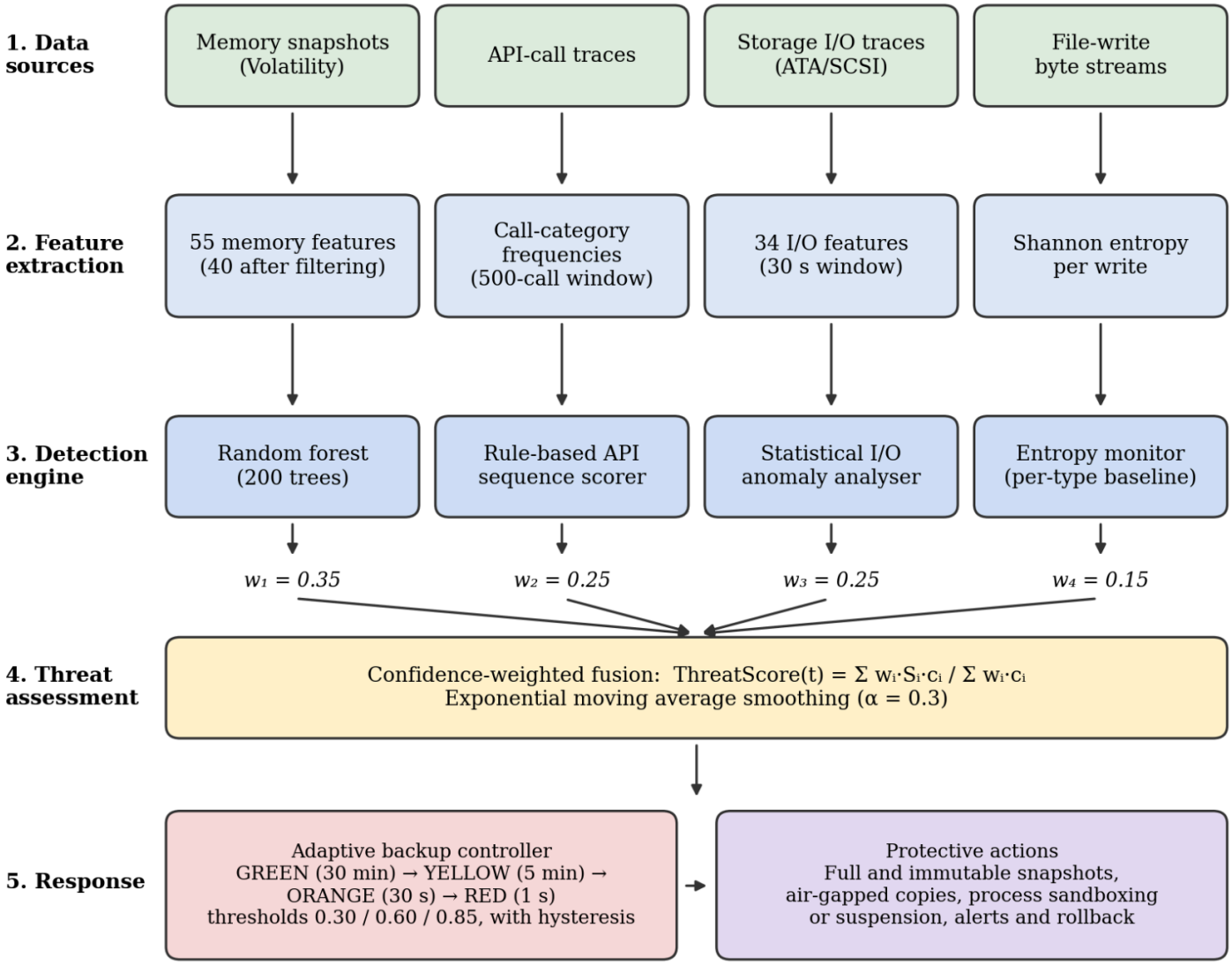


Figure 1

RansomShield architecture: four telemetry sources feed four detectors, whose outputs are fused into a smoothed threat score that drives the adaptive backup controller.

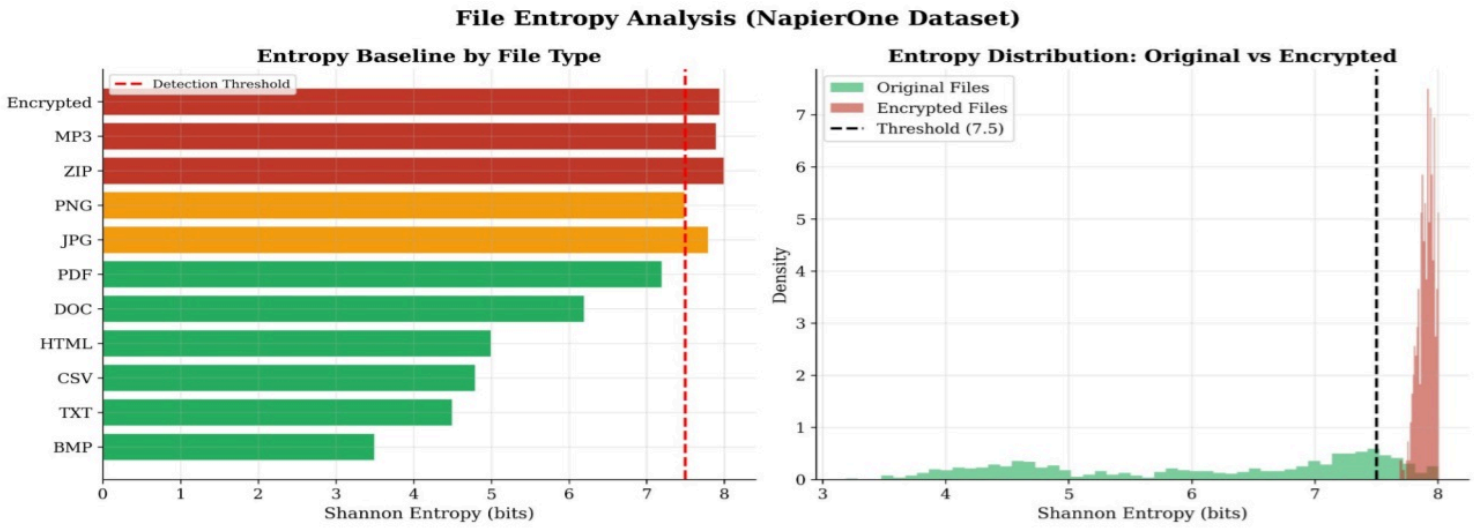


Figure 2

Shannon entropy baselines by file type (left) and the distribution of original versus encrypted files against the 7.5 bits/byte threshold (right).

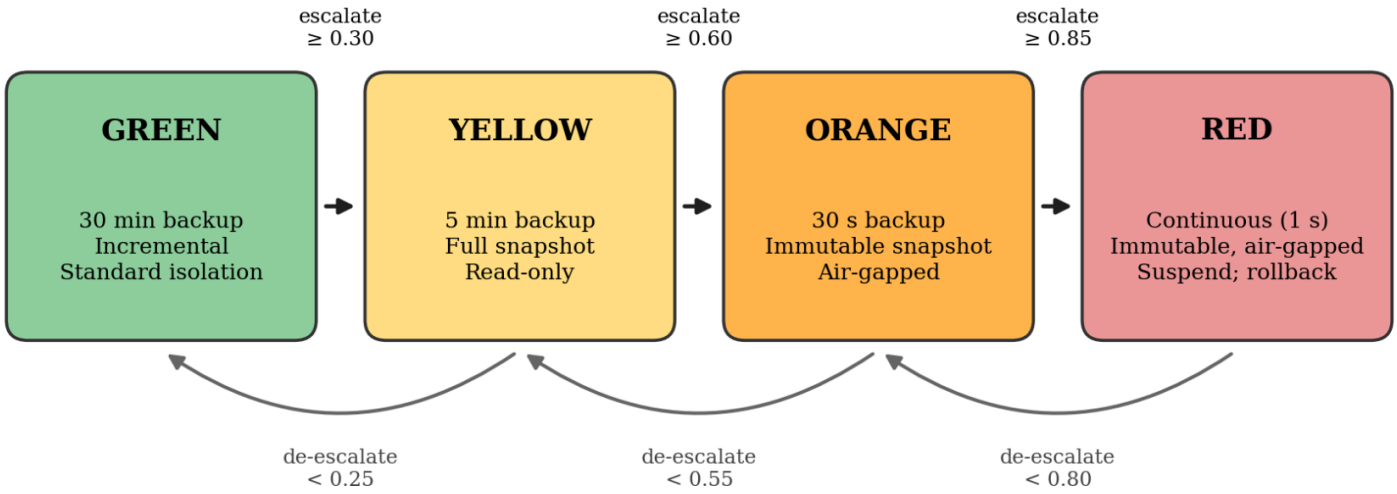


Figure 3

Adaptive threat-level state machine with escalation thresholds.

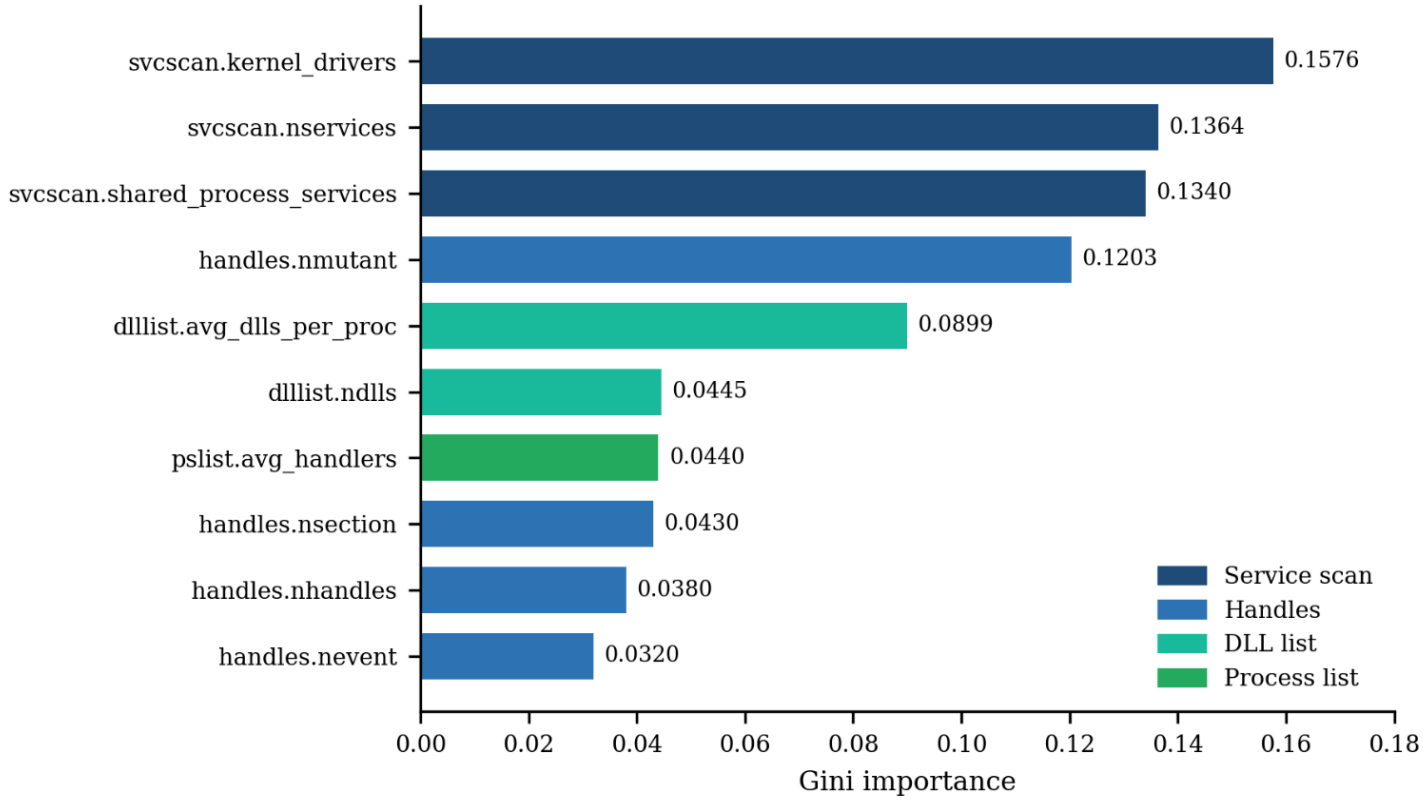


Figure 4

Ten most important random-forest features (Gini importance) on CIC-MalMem-2022.

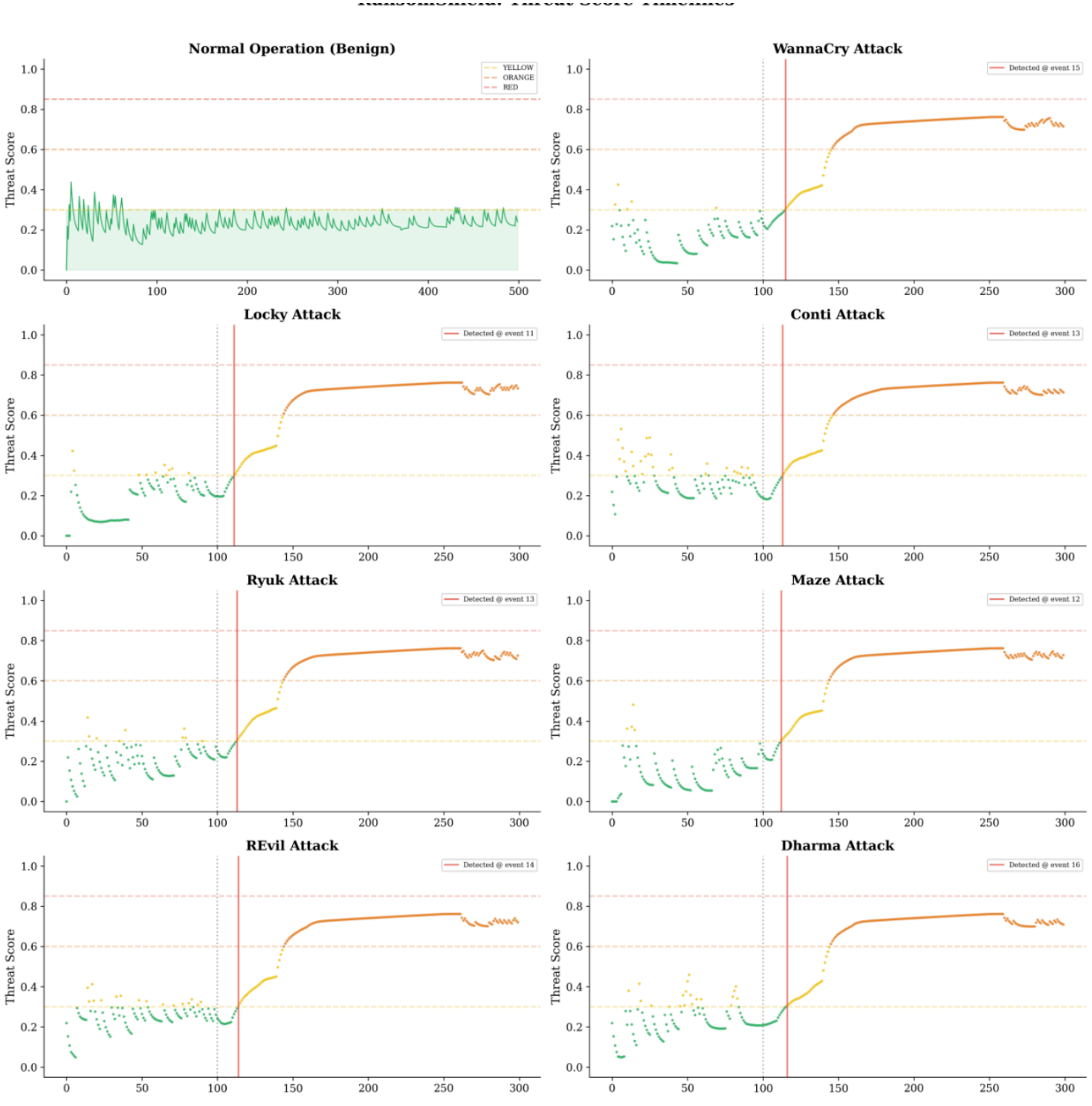


Figure 5

Threat-score timelines for the benign scenario and the seven simulated attacks. The vertical line marks first detection; dashed lines mark the YELLOW, ORANGE and RED thresholds.