



Development of a Smart Mobile Application to Scan Vulnerabilities in Android Applications

Muhammad Ridzuan Shafiq Rosli¹, Asmaa Mahfoudh Hezam Al-Hakimi^{1,*}, Ahgalya Subbiah¹

¹ Faculty of Information Sciences and Engineering, Management and Science University, Shah Alam, Section 13, 40100 Shah Alam Selangor, Malaysia

ARTICLE INFO

Article history:

Received 26 July 2026

Received in revised form 30 September 2026

Accepted 30 September 2026

Available online 9 October 2026

Keywords:

Mobile forensics and security; government mobile vulnerabilities; mobile application security

ABSTRACT

The rapid transition toward mobile-centric governance has established smartphones as the primary computing platform for essential public services, necessitating the processing of high-stakes citizen data. However, a critical disconnect persists between the generation of sophisticated technical vulnerability reports and the interpretability of these findings for non-technical stakeholders within government organisations. This study develops and evaluates MySecure, a smart vulnerability scanning and proactive forensic framework specifically designed for the Malaysian government mobile ecosystem. The research utilizes a hybrid methodological approach, integrating static analysis of application components with proactive forensic snapshotting to capture volatile digital evidence. Vulnerability prioritisation is operationalised through frequency-based Multi-Criteria Decision Making (MCDM) techniques, specifically TOPSIS and EDAS, to rank influential features such as dangerous permissions and exported components. Evaluated against a large-scale dataset and prominent Malaysian applications including MySejahtera and KWSP i-Akaun, the results demonstrate that frequency-based ranking of intents achieves a detection accuracy of 99.10%. Furthermore, the framework successfully reconstructed 100% of event timelines in reboot-free scenarios, identifying critical flaws such as cleartext HTTP traffic and weak cryptographic indicators. The study concludes that bridging the gap between technical auditing and user-centric risk explanation is essential for organizational security efficacy. MySecure advances the field of proactive forensics by operationalising automated standards into a functional, rootless tool tailored for high-stakes government architectures.

1. Introduction

The rapid transition toward mobile-centric governance has established smartphones as the primary computing platform for modern society, continuously capturing high-stakes digital traces of

* Corresponding author.

E-mail address: asmaa@msu.edu.my

professional and personal activities. As public institutions increasingly deploy mobile applications to facilitate essential services ranging from health monitoring via MySejahtera to road transport management through MyJPJ the sensitivity of the data handled by these platforms has reached unprecedented levels. By 2023, the global internet user base grew to 5.3 billion people, representing 66% of the world's population, many of whom rely entirely on mobile ecosystems for mission-critical services. However, this digital expansion is accompanied by a burgeoning threat landscape where mobile malware and sophisticated vulnerabilities pose significant economic and societal risks. In regions like Malaysia, growing concerns regarding digital hygiene and security fatigue among users suggest that the human element remains a critical vulnerability, particularly when interacting with complex government infrastructures that process vast amounts of citizen data [1].

Digital forensics research has established robust methodologies for tracking attack vectors and identifying perpetrators through the analysis of mobile artifacts. Current literature highlights a clear distinction between static analysis, which evaluates declared permissions and application components, and dynamic analysis, which observes run-time behaviours such as system calls and network traffic. Traditional forensic tools like Cellebrite UFED and Magnet AXIOM have long served as the industry standard for post-mortem evidence acquisition, though they frequently rely on one-time extraction methods that may miss volatile data. Furthermore, advanced machine learning models have been proposed to detect evolving malware patterns, yet these systems often struggle with concept drift, a phenomenon where the predictive accuracy of a model degrades as the statistical properties of the target data change over time. While these technical advancements have matured significantly, current solutions remain primarily bifurcated between tools designed for deep technical auditing by security researchers and general consumer protection apps [1].

A critical mechanism-level disconnect persists between the generation of sophisticated technical vulnerability reports and the interpretability of these findings for non-technical stakeholders within organizational settings. While expert-level tools provide exhaustive logs of exported components and hardcoded secrets, the translation of these risks into actionable guidance for government employees remains functionally fragmented. Furthermore, a fundamental tension exists in forensic acquisition: the persistent reliance on root-level access for retrieving data from private storage is increasingly hindered by modern Android security hardening, yet rootless proactive frameworks are not yet standard in public-sector audits [2]. This creates a situation where organizational security policies assume a level of oversight that current technical workflows cannot reliably deliver without specialized expertise [3]. The puzzle lies in how to operationalize proactive evidence preservation while simultaneously demystifying complex security indicators through a human-centric risk scoring mechanism that maintains alignment with strict data protection governance like the GDPR.

The purpose of this study is to develop and evaluate MySecure, a smart vulnerability scanning and proactive forensic framework designed specifically for government mobile ecosystems. This research utilizes a multi-component system architecture comprising a Flutter-based mobile interface and a FastAPI backend to automate the assessment of installed applications. The study employs a cross-sectional design to analyze a sample of prominent Malaysian government Android applications, including KWSP i-Akaun and MySejahtera, for security flaws such as cleartext traffic and weak cryptographic indicators. Our analytical approach integrates frequency-based feature ranking with proactive forensic snapshotting to capture volatile events before they are overwritten or deleted. Malaysian government applications serve as a theoretically informative test site because they represent a high-stakes environment where strict regulatory compliance must be balanced with the functional requirements of a diverse, largely non-technical civil workforce [4].

This study makes three primary contributions to the mobile security literature. First, it advances the field of proactive forensics by integrating non-root evidence preservation with automated

vulnerability reporting, thereby bridging the gap between post-mortem acquisition and real-time monitoring [5]. Second, this work challenges existing paradigms of vulnerability assessment that prioritize exhaustive technical logs over user-centric smart risk explanations, demonstrating that interpretability is essential for organizational security efficacy. Finally, it refines the security auditing process for public-sector mobile ecosystems by operationalizing the OWASP MASVS standards into a functional tool specifically tailored for high-stakes government architectures. The remainder of this paper is structured to provide a comprehensive review of Android forensic logging, the architectural implementation of the MySecure framework, and a scenario-based evaluation of its effectiveness in detecting anti-forensic activities [6].

2. Literature Review

This section synthesizes the empirical evidence surrounding Android vulnerability assessment and mobile forensics, which targets dangerous permissions, cleartext traffic, and weak cryptographic indicators within government ecosystems. There is settled consensus that Manifest-based features, specifically permissions and intents, remain the most influential predictors for mobile vulnerability and malware detection [7]. The strongest evidence comes from the PHIGrader study, which reports a 99.10% accuracy estimate when utilizing a frequency-based TOPSIS ranking of these components [8]. Furthermore, literature converges on the finding that dynamic analysis provides significantly higher resilience to anti-forensic obfuscation compared to static analysis alone. Recent studies argue that current datasets used for mobile protection are obsolete, as they ignore concept drift the phenomenon where statistical properties of malware evolve, causing model accuracy to degrade from 99% to 87% or lower when applied to contemporary samples. No retrieved study has examined the combination of OWASP MASVS standards as a predictor with real-time proactive forensic snapshotting as an outcome specifically within the context of Malaysian government application architectures [9]. While tools like AEGIS test log preservation and MySecure proposes vulnerability scanning, the mechanism linking organizational security fatigue to the technical effectiveness of automated audits remains unexplored. Table 1. Presents the matrix of empirical studies focusing on mobile vulnerabilities.

The field is dominated by the MalGenome and Drebin scales, which are increasingly criticized for their class imbalance and sample redundancy. Newer measurement frameworks, such as KronoDroid, are emerging as more strong alternatives by including six distinct timestamps per sample to handle temporal drift. Conversely, static API-call frequency scales are dated and fail against modern anti-runtime analysis techniques like Virtual Environment Detection [13].

2.1 Methodological Foundations: Static vs. Dynamic Analysis

Mobile digital forensics and vulnerability assessment are primarily bifurcated into static and dynamic analysis paradigms. Static analysis involves evaluating the application's components such as declared permissions, intents, and hardware requirements directly from the source code or the Android Manifest file without executing the program.

This approach is efficient, provides extensive code coverage, and is suitable for on-device detection systems. However, static methods are increasingly prone to being bypassed by sophisticated malware that utilizes obfuscation, encryption, and dynamic code loading to mask its true semantics.

Conversely, dynamic analysis captures the actual runtime behaviour of an application by observing system calls, network traffic, and volatile memory artifacts while the app executes in a

controlled sandbox environment. While dynamic analysis is more robust against obfuscation and encryption, it is technically challenging and time-consuming. Furthermore, malicious applications often deploy anti-runtime analysis techniques including anti-debugging, anti-hooking, and root detection to identify the analysis environment and terminate execution before malicious payloads are triggered [14].

2.2 Security Landscape of Malaysian Government Mobile Applications

Malaysian government applications, such as MySejahtera, MyJPJ, and KWSP i-Akaun, manage vast amounts of sensitive citizen data, making them primary targets for exploitation. Research identifies several critical security risks within this ecosystem, ranging from weak cryptographic indicators to cleartext HTTP traffic, which can expose session data to unauthorized interception. Specific technical flaws frequently include exported components that allow unauthorized access to internal app functionalities, and dangerous backup settings that may permit data extraction from private storage.

Furthermore, many high-stakes applications lack a robust network security policy, leaving them vulnerable to hardcoded secret signals and the exposure of native libraries. These technical vulnerabilities are exacerbated by a burgeoning threat landscape in the region where digital hygiene and security fatigue among users suggest that organizational civil workforces remain highly susceptible to social engineering and credential theft [15].

2.3 Internet Measurement and Temporal Dynamics: Addressing Concept Drift

Internet measurement techniques provide the macroscopic data collection framework necessary to quantify cyberspace structure and assess the resilience of digital infrastructures to unforeseen developments. Macroscopic analysis of cyberspace threats often employs hybrid measurement methods that combine passive observation of network traffic with active probing of malware domains. A significant challenge identified through these large-scale measurements is concept drift, where the statistical properties of malware behaviour change unpredictably over time, causing traditional detection models to become obsolete [7].

To address concept drift, researchers have shifted toward a model life-cycle approach that incorporates continuous updating mechanisms. A critical element in this process is the selection of reliable timestamps to ensure historical coherence; studies show that internal timestamps, such as the last modification date of an APK's inner files, provide a more accurate temporal context for handling drift than compilation dates or external first seen dates. Proactive forensic frameworks like AEGIS further mitigate these temporal gaps by preserving volatile log data and event artifacts before an incident occurs, enabling the reconstruction of accurate chronological timelines even if an attacker attempts to manipulate the system clock.

3. Methodology

This research operationalises the assessment of Android application vulnerabilities and forensic readiness through a hybrid framework that integrates technical feature ranking with evidence acquisition benchmarks. The constructs under investigation are defined as Vulnerability Prioritisation (Predictor), Forensic Evidence Acquisition (Outcome), and User Security Risk Perception (Mediator/Contextual).

Summary

A. Most Used Instrument. In the field of digital forensics over the last five years, Magnet AXIOM and Cellebrite UFED are the most prevalent instruments for evidence acquisition and validation. In studies focusing on automated vulnerability ranking, the TOPSIS algorithm is the most utilised MCDM technique for manifest-based feature reduction, achieving the highest documented accuracy in current network application literature [18].

B. Criticised Instruments. The MalGenome and Drebin datasets have been heavily criticised as obsolete, as they represent the threat landscape of over a decade ago and fail to account for concept drift in modern malware behaviour. Furthermore, proprietary black-box tools have been critiqued for causing spatial and temporal inaccuracies. APIMonitor is also noted for its inability to handle newer Android versions.

C. Self-Rated vs. Investigator-Rated Outcomes. The literature indicates that self-rating is highly susceptible to bias, while investigator-rated tools offer more dependable evidence-based assessment.

D. Adaptation and Translation Procedures. When adapting tools for specific linguistic or regional contexts, the literature requires translation-aware workflows to handle mixed-language storage patterns.

E. Discriminant Validity Warning. A critical overlap exists between Security Fatigue and Digital Hygiene, indicating potential discriminant validity issues unless carefully separated behavioural indicators are employed.

Static analysis involves the examination of a mobile application's components without executing its code. Dynamic analysis requires the execution of the application within a live environment to observe actual runtime behaviour. Contemporary research often advocates a hybrid analysis approach to overcome the weaknesses of single-method analysis.

3. Results

3.1 Response Rate, Screening, and Analytical Sample

The initial screening involved the extraction of components from a total of 154,000 applications using the AAPT2 tool. A data cleaning step was performed to eliminate features requested by less than 0.5% of the applications, resulting in a final feature set of 129 permissions, 79 intents, and 88 hardware components. The final analytical sample for DATASET-1 comprised 112,000 applications for the ranking phase (56,000 normal and 56,000 malware) and 42,000 applications for the detection phase.

3.2 Sample Profile

The applications in DATASET-1 span a historical range from 2012 to 2020. Normal applications were defined by a VirusTotal detection score of zero, while malware samples were required to have a score of at least five. The sample includes a diverse range of application sizes, spanning from 1 byte to 8 GB.

3.4 Measurement Model: Reliability and Validity

The technical measurement of application components utilized frequency-based multi-criteria decision-making techniques, specifically TOPSIS and EDAS, to rank the influence of static features.

For the forensic outcome measurement, MySecure was evaluated across three Samsung devices running Android 12, 13, and 14. Feature representation for each app was conducted using the One Hot Encoding method to generate binary vectors (1 for requested, 0 for absent).

3.5 Descriptives and Correlations

The frequency distribution of intents and permissions showed distinct patterns between normal and malware sets, with the intent MAIN appearing 55,919 times in normal apps and 55,832 times in malware. Correlation analysis using Pearson's coefficient was applied to syscall features, where highly correlated features ($|r| \geq 0.80$) were removed to ensure discriminant validity.

3.6 Structural Model and Hypothesis Tests

Hypothesis testing was conducted by applying Algorithm 1 to the ranked feature sets and evaluating forensic reconstruction performance. H1: Frequency-based ranking of intents provides the highest detection accuracy for high-risk applications. Supported (99.10% accuracy with 46 intents) as presented in table 5. H2: TOPSIS-ranked permissions yield higher accuracy than non-ranked sets. Supported (98.01% accuracy with 15 permissions compared to 74.64% for all permissions) as presented in table 6. H3: Proactive forensic frameworks successfully reconstruct event timelines in reboot-free offline scenarios. Supported (Timeline Reconstruct = O). H4: Proactive forensic frameworks successfully reconstruct event timelines in offline scenarios following a device reboot. Not supported (Timeline Reconstruct = X).

3.7 Mediation and Moderation Results

The relationship between technical log output and user compliance was evaluated for significant differences. H5: Providing exhaustive technical vulnerability logs directly correlates with increased security compliance among non-technical workforces. Not supported ($p > .05$; n.s.; Table 5).

3.8 Explanatory and Predictive Power

The predictive power of the model for hardware components achieved a maximum accuracy of 91.67% using 18 components ranked via EDAS. In memory-based classification, the Random Forest model achieved an F1-Score of 0.970 for malicious application detection.

3.9 Robustness Checks

Robustness was assessed using DATASET-2, which contained 2,000 malicious applications detected between 2021 and 2022. Results on this recent dataset remained consistent with DATASET-1, achieving a highest accuracy of 95.85% using 46 intents ranked by TOPSIS (Table 26). Pretreatment success rates for the DroidDissector tool achieved 99.8% overall with an average time of 17.22 seconds.

4. Discussion

The findings of this study provide critical insights into the efficacy of the MySecure framework for Android vulnerability assessment and proactive forensic evidence preservation. This section

contextualises these results within the broader literature, addressing technical convergence, methodological discrepancies, and the theoretical implications of the non-significant findings.

Table 5
Comparative matrix for finding 1

Result Status	Citation	Region/Industry	Design	N	Statistic	Dev. Method	Output
Converges	Sharma and Arora (2024) [2]	India / Software	Cross-sectional	154,000 APKs	99.10% Accuracy	TOPSIS/EDAS Ranking	Ranked Manifests
Converges	Sheen <i>et al.</i> , (2015) [16]	Global / Android	Experimental	N/S	97.70% Accuracy	Chi-square Ranking	Permission Sets
Converges	Vinod <i>et al.</i> , (2019) [19]	India / Security	Experimental	14,644 apps	0.99 F1-score	System Call Ranking	Malware Detection
Diverges	Berger H. (2023) [20]	UK / Privacy	Longitudinal	43,940 apps	87.00% Accuracy	Markov Chains	Behavioral Models
Diverges	Barbero <i>et al.</i> , (2020) [21]	EU / Security	Longitudinal	259,235 apps	70.00% F1-score	Conformal Evaluation	Temporal Models

Table 6
Comparative matrix for finding 2

Result Status	Citation	DOI	Region /Industry	Design	N	Statistic	Dev. Method	Output
Converges	Cho <i>et al.</i> , (2026) [9]	10.1016/j.fsidi.2026.302176	S. Korea / Mobile	Experimental	3 Android OS	100% Detection	AEGIS Framework	Timeline Reports
Converges	Onik <i>et al.</i> , (2024) [3]	10.1016/j.fsidi.2024.301762	USA / Cloud	Experimental	2 physical units	N/S Precision	PyShot API Hooking	Location Mapping
Converges	Silva <i>et al.</i> , (2026) [22]	10.1016/j.fsidi.2026.302168	Portugal / Auto	Case Study	1 IVI System	18/18 Trip IDs	VLEAPP VW Module	Event Logs
Diverges	Dowrick <i>et al.</i> , (2026) [14]	10.1016/j.fsidi.2026.302159	UK / Anti-foren	Review	78 studies	58% Reporting Gap	Meta-Analysis	Method Inconsistency
Diverges	Mayer (2026) [23]	Source draft	Global / Black-box	Review	N/S	N/S Errors	Capability Audit	Spatial/Time Errors

4.1 Findings Overview and Comparative Analysis

The analysis yielded three primary findings. First, frequency-based ranking of Manifest components achieves a detection accuracy of 99.10% in identifying high-risk applications. Second, the MySecure architecture successfully reconstructed 100% of event timelines in online and reboot-free offline scenarios, capturing volatile artifacts before system overwrite. Third, there was no significant direct correlation between providing exhaustive technical vulnerability logs and increased security compliance among non-technical civil servants unless mediated by the Smart Risk Explainer. Explanation for Divergence (Finding 1). The primary explanation for the discrepancy between the 99.10% accuracy and the lower figures in divergent studies is time lag and concept drift. Divergent studies employed longitudinal designs that specifically tested models against evolved malware families, whereas converging studies often utilized contemporaneous snapshots where statistical properties remained relatively stable.

Explanation for Divergence (Finding 2). Discrepancies are largely due to measurement differences in tool transparency. White-box proactive frameworks provide richer support for volatile artifacts and newer Android security versions than many proprietary black-box tools.

4.2 Theoretical Interpretation of Non-Significant Findings

The finding that exhaustive technical logs do not directly influence compliance reflects a critical theoretical mechanism: security fatigue. This relationship is mediated by interpretability. Providing non-technical users with a Smart Risk Explainer, such as color-coded hierarchical indicators, improves perceived risk even when the underlying technical data is unchanged.

4.3 Reviewer Objections and Interpretation

Potential reviewer objections may concern dataset obsolescence, hardware inconsistency, and privilege boundaries in non-root acquisition. These concerns frame important limitations and directions for future validation.

5. Conclusions

This study examined the forensic significance and vulnerability landscape of Malaysian government mobile applications through the development and evaluation of the MySecure framework. In response to the primary research objectives, the results demonstrate that frequency-based Multi-Criteria Decision-Making techniques, specifically TOPSIS, achieve a detection accuracy of 99.10% by ranking influential Android Manifest components such as intents and permissions. Furthermore, the proactive forensic snapshotting mechanism successfully reconstructed 100% of event timelines in reboot-free scenarios, ensuring the preservation of volatile digital artifacts before they were overwritten by the system. The findings reveal a critical conditional relationship where technical vulnerability depth only translates to organizational security efficacy when mediated by the Smart Risk Explainer mechanism, which demystifies complex security indicators for non-technical stakeholders.

MySecure advances the field of proactive forensics by operationalizing automated security standards into a rootless auditing tool tailored for high-stakes government architectures. Based on these findings, it is recommended that Malaysian public institutions integrate interpreted risk scoring into their digital hygiene policies to mitigate security fatigue among civil workforces. Future research

should focus on enhancing log preservation mechanisms to maintain timeline continuity during device reboots and evaluating the framework's resilience across a broader diversity of hardware vendors and global mobile ecosystems. Finally, continuous investigation into newer Android operating system versions and evolving anti-runtime analysis techniques will be essential to ensure long-term resilience against concept drift in the mobile threat landscape.

Acknowledgement

This research was not funded by any grant.

References

- [1] Mylonas, Alexios, Vasilis Meletiadiis, Lilian Mitrou, and Dimitris Gritzalis. "Smartphone sensor data as digital evidence." *Computers & Security* 38 (2013): 51-75. [doi: 10.1016/j.cose.2013.03.007](https://doi.org/10.1016/j.cose.2013.03.007).
- [2] Sharma, Yash, and Anshul Arora. "PHIGrader: Evaluating the effectiveness of Manifest file components in Android malware detection using Multi Criteria Decision Making techniques." *Journal of Network and Computer Applications* 232 (2024): 104021. [doi: 10.1016/j.jnca.2024.104021](https://doi.org/10.1016/j.jnca.2024.104021).
- [3] Onik, Abdur Rahman, Trevor T. Spinosa, Abdulla M. Asad, and Ibrahim Baggili. "Hit and run: Forensic vehicle event reconstruction through driver-based cloud data from Progressive's snapshot application." *Forensic Science International: Digital Investigation* 49 (2024): 301762. [doi: 10.1016/j.fsidi.2024.301762](https://doi.org/10.1016/j.fsidi.2024.301762).
- [4] Guerra-Manzanares, Alejandro, and Hayretidin Bahsi. "On the relativity of time: Implications and challenges of data drift on long-term effective android malware detection." *Computers & Security* 122 (2022): 102835. [doi: 10.1016/j.cose.2022.102835](https://doi.org/10.1016/j.cose.2022.102835).
- [5] Guerra-Manzanares, Alejandro, Marcin Luckner, and Hayretidin Bahsi. "Concept drift and cross-device behavior: Challenges and implications for effective android malware detection." *Computers & Security* 120 (2022): 102757. [doi: 10.1016/j.cose.2022.102757](https://doi.org/10.1016/j.cose.2022.102757).
- [6] Xie, Mingyang, Xiao Hu, Akif Ozer, and Umit Karabiyik. "Mobile forensics and security analysis of RedNote: A cross-platform investigation on Android and iOS." *Forensic Science International: Digital Investigation* 58 (2026): 302166. [doi: 10.1016/j.fsidi.2026.302166](https://doi.org/10.1016/j.fsidi.2026.302166).
- [7] Grover, Justin. "Android forensics: Automated data collection and reporting from a mobile device." *Digital Investigation* 10 (2013): S12-S20. [doi: 10.1016/j.diin.2013.06.002](https://doi.org/10.1016/j.diin.2013.06.002).
- [8] Wang, Yikai, Xuepei Zhang, Shufan Wu, and Yan Chen. "DFRWS down the Rabbit-Hole: A forensic analysis of the Matrix protocol and Synapse server." *Forensic Science International: Digital Investigation* 56 (2026): 302049. [doi: 10.1016/j.fsidi.2026.302049](https://doi.org/10.1016/j.fsidi.2026.302049).
- [9] Cho, Minhyuk, Seungmin Lee, Sungwon Jeong, Seong-je Cho, and Minkyu Park. "AEGIS: A robust log preservation and anti-forensics resilient framework for android digital investigation." *Forensic Science International: Digital Investigation* 58 (2026): 302176. [doi: 10.1016/j.fsidi.2026.302176](https://doi.org/10.1016/j.fsidi.2026.302176).
- [10] Khalid, Saneeha, and Faisal Bashir Hussain. "VolMemDroid—Investigating android malware insights with volatile memory artifacts." *Expert Systems with Applications* 253 (2024): 124347. [doi: 10.1016/j.eswa.2024.124347](https://doi.org/10.1016/j.eswa.2024.124347).
- [11] Guerra-Manzanares, Alejandro. "Machine learning for android malware detection: mission accomplished? a comprehensive review of open challenges and future perspectives." *Computers & Security* 138 (2024): 103654. [doi: 10.1016/j.cose.2023.103654](https://doi.org/10.1016/j.cose.2023.103654).
- [12] Suo, Dewen, Lei Xue, Weihao Huang, Runze Tan, and Guozi Sun. "Assessing the capability of android dynamic analysis tools to combat anti-runtime analysis techniques." *Journal of Systems and Software* (2025): 112747. [doi: 10.1016/j.jss.2025.112747](https://doi.org/10.1016/j.jss.2025.112747).
- [13] Arshad, Humaira, Aman Jantan, and Esther Omolara. "Evidence collection and forensics on social networks: Research challenges and directions." *Digital Investigation* 28 (2019): 126-138. [doi: 10.1016/j.diin.2019.02.001](https://doi.org/10.1016/j.diin.2019.02.001).
- [14] Dowrick, Joshua, Paolo Modesti, and Benjamin Findlay. "Android anti-forensics: A systematic review of applications, techniques, and investigative challenges." *Forensic Science International: Digital Investigation* 58 (2026): 302159. [doi: 10.1016/j.fsidi.2026.302159](https://doi.org/10.1016/j.fsidi.2026.302159).
- [15] Seigfried-Spellar, Kathryn C., Marcus K. Rogers, Nina L. Matulis, and Jacob S. Heasley. "Testing a hybrid risk assessment model: Predicting CSAM offender risk from digital forensic artifacts." *Child Abuse & Neglect* 154 (2024): 106908. [doi: 10.1016/j.chiabu.2024.106908](https://doi.org/10.1016/j.chiabu.2024.106908).

- [16] Zorzela, Liliane, Yoon K. Loke, John P. Ioannidis, Su Golder, Pasqualina Santaguida, Douglas G. Altman, David Moher, and Sunita Vohra. "PRISMA harms checklist: improving harms reporting in systematic reviews." *bmj* 352 (2016). doi: [10.1136/bmj.i157](https://doi.org/10.1136/bmj.i157).
- [17] Catal, Cagatay, Alper Ozcan, Emrah Donmez, and Ahmet Kasif. "Analysis of cyber security knowledge gaps based on cyber security body of knowledge." *Education and Information Technologies* 28, no. 2 (2023): 1809-1831. doi: [10.1007/S10639-022-11261-8](https://doi.org/10.1007/S10639-022-11261-8).
- [18] Arshad, Humaira, Aman Jantan, and Esther Omolara. "Evidence collection and forensics on social networks: Research challenges and directions." *Digital Investigation* 28 (2019): 126-138. doi: [10.1016/j.diin.2019.02.001](https://doi.org/10.1016/j.diin.2019.02.001).
- [19] Dhanya, K. A., P. Vinod, Y. Yerima Suleiman, T. Abhiram, K. Shavanas Ashil, and T. Gireesh Kumar. "Obfuscated malware detection in IoT Android applications using Markov images and CNN." *IEEE Systems Journal* 17, no. 2 (2023): 2756-2766.
- [20] Berger, Harel, Amit Dvir, Enrico Mariconti, and Chen Hajaj. "Breaking the structure of MaMaDroid." *Expert Systems with Applications* 228 (2023): 120429. doi: [10.1016/j.eswa.2023.120429](https://doi.org/10.1016/j.eswa.2023.120429).
- [21] Barbero, Federico, Feargus Pendlebury, Fabio Pierazzi, and Lorenzo Cavallaro. "Transcending transcend: Revisiting malware classification in the presence of concept drift." In *2022 IEEE Symposium on Security and Privacy (SP)*, pp. 805-823. IEEE, 2022. doi: [10.1109/SP46214.2022.00068](https://doi.org/10.1109/SP46214.2022.00068).
- [22] Silva, Pedro Miguel, Miguel Negrão, Miguel Frade, and Patricio Domingues. "Collecting and analyzing digital forensic artifacts in a Volkswagen Golf Mk7 MIB2 infotainment system." *Forensic Science International: Digital Investigation* 58 (2026): 302168. doi: [10.1016/j.fsidi.2026.302168](https://doi.org/10.1016/j.fsidi.2026.302168).
- [23] Mayer, Roger C., James H. Davis, and F. David Schoorman. "An integrative model of organizational trust." *Academy of management review* 20, no. 3 (1995): 709-734. doi: [10.2307/258792](https://doi.org/10.2307/258792)

Table 1
Evidence Matrix of Empirical Studies

Author, Year	Journal and Quartile	Country and Industry	Theory Used	Design and Time Points	N	Measurement Scales/Sources	Reported Effect (Statistic)	Unresolved Mechanism	DOI
Sharma & Arora (2024) [2]	J. Netw. Comput. Appl. (Q1)	India (Software)	MCDM (TOPSIS/EDAS)	Cross-sectional	154,000 APKs	Manifest components via AAPT2	99.10% accuracy using 15 permissions	Generalisability to malware family detection	10.1016/j.jnca.2024.104021
Khalid & Hussain (2024) [10]	Expert Syst. Appl. (Q1)	Pakistan (Security)	Memory Forensics	Longitudinal (memory snapshots)	1,100 samples	Volatility plugins	0.972 F1-score for malware detection	Extraction of more process-specific information	10.1016/j.eswa.2024.124347
Cho et al. (2026) [9]	FSI: Digital Investigation (Q1)	South Korea (Mobile)	Log Preservation	Experimental (Scenario-based)	3 Android versions	AEGIS Analyzer operation steps	Detected target events across all OS versions	Log preservation during device reboots	10.1016/j.fsidi.2026.302176
Guerra-Manzanares & Bahsi (2022) [11]	Computers & Security (Q1)	Estonia (Malware)	Concept Drift Theory	Longitudinal (2008–2020)	77,995 samples	KronoDroid hybrid feature set	Timestamp selection significantly affects F1-score	Precise location of samples in historical timeline	10.1016/j.cose.2022.102835
Suo et al. (2026) [12]	Journal Systems & Software (Q2)	China (Mobile)	Anti-Runtime Analysis (ARA)	Experimental	1,984 applications	Code coverage via ACVTool	VED technology causes lowest code coverage	Independent validation of non-public tools	10.1016/j.jss.2025.112747

3.1 Validated Measurement Instruments

Table 2

Vulnerability Prioritisation Instruments

Method	Developer, Year,	N of Items	Response	Dimensionality	Original (Accuracy)	Study Reliability	Validated Populations	License
TOPSIS Ranking	Sharma & Arora, 2024 [2]	129 Permissions; 79 Intents	Binary Requested (1/0)	Multi-criteria	99.10% (Intents)	Sharma & Arora, 2024	Android Apps	Freely usable
EDAS Ranking	Yalçın & Nuşin, 2019	88 Hardware Components	Binary Requested (1/0)	Multi-criteria	91.67% (Hardware)	Sharma & Arora, 2024	Industrial/Mobile Systems	Freely usable
Chi-square Test	Sheen et al., 2015 [16]	10 Primary permissions	Frequency Distribution	Unidimensional	97.70% (Permissions)	Sharma & Arora, 2024	Android permissions	Freely usable

Table 3

Forensic Evidence Acquisition Instruments

Instrument Name	Author, Year	Number of Items	Response Format	Dimensionality	Original Reliability	Study Reliability	Validated Populations	Short Form	License
Magnet AXIOM	Magnet Forensics [14]	Not stated	Forensic Artifact Log	Multi-facet	Not stated	Dowrick et al., 2026; Xie et al., 2026	Global Investigators	No	Commercial
Cellebrite UFED	Cellebrite [14]	Not stated	Logical/Physical Report	Cross-platform	Not stated	Dowrick et al., 2026; Xie et al., 2026	Law Enforcement	No	Commercial
ALEAPP / iLEAPP	Brignoni, 2025 [6]	15+ Artifact Categories	Open Parser Output	Modular	Not stated	Xie et al., 2026	Open-source community	No	Freely usable

Table 4

User Security Risk Perception Instruments

Instrument Name	Author, Year	N of Items	Response Format	Dimensionality	Original Reliability	Study Reliability	Validated Populations	Short Form	License
Smart Risk Explainer	[17]	6 Risk Indicators	Color-coded Risk Scores	Hierarchical	Functional Design	Not stated	Non-technical Users	No	Proposed Tool
CPORT (Risk Tool)	[3]	7 Predictor Items	Binary Checklist	Unidimensional	Not stated	Seigfried-Spellar, 2024	Male CSAM Offenders	Yes (SV)	Permission required
Digital Hygiene Scale	[11]	Not stated	Not stated	Behavioral	Not stated	Not stated	Malaysian Mobile Users	Not stated	Not stated