A Systematic Review on Hybrid Analysis using Machine Learning for Android Malware Detection

Asadullah Hill Galib
Institute of Information Technology
University of Dhaka
bsse0712@iit.du.ac.bd

B M Mainul Hossain
Institute of Information Technology
University of Dhaka
mainul@iit.du.ac.bd

***Abstract*—Android is the most ubiquitous mobile operating system nowadays. It's prevalence provokes humongous growth of android malware. Researchers seek to sort out an effective strategy to defend against those malware authors. Primarily they have focused on static and dynamic analysis using machine learning to detect android malware. But, multifarious evasion techniques by the shrewd malware authors have made those approaches ineffective. Yet researchers consistently aim at discovering an effectual strategy to fight against. Hybrid analysis: a fusion of static and dynamic analysis would be a good candidate for that as it prevails over the individual drawbacks of static and dynamic analysis with the cost of complexity. Recently researchers have put emphasis on this regard and revealed a lot of challenges and opportunities. This work aims at presenting a thorough and systematic review on hybrid analysis using machine learning techniques for android malware detection. It encompasses the leading researches on hybrid analysis: their contributions, strengths and weaknesses. This work also discusses the challenges, limitations and future directions of hybrid analysis for android malware detection.**

***Index Terms*—Android Malware Detection, Hybrid Analysis, Machine Learning**

I. Introduction

Android is the most prevalent mobile operating system (os) currently: 72.23% of total mobile os is android [1]. With the enormous growth of the android system [2], android malware also has grown significantly as well as upgraded its nature and activities [2]. On average 12,000 new malware instances are found per day [3]. To defend against that malware phenomena, researchers have put emphasis on android malare detection to ensure android mobile application security.

To detect android malware, there are three malware detection approaches: Static Analysis, Dynamic Analysis and Hybrid Analysis. Static analysis approach uses the static features of the android application such as Permissions, API Calls, Intents, Call-Graph, Opcode, Hardware Usage Analysis, Meta-data etc. Dynamic analysis approach investigates the dynamic behavior of the application running on an emulated environment or on a real device. These dynamic features/behaviors include System Calls, Network Traffic, File Operations, Running Services, Network Operations etc. Hybrid analysis tends to incorporate both the static and the dynamic approaches into a common ground.

Static and dynamic analysis have their own limitations. Currently malware authors are too smart to evade these detection techniques. They use many evasion techniques to evade the analysis. For static analysis, commonly used evasion techniques by the malware authors are data obfuscation, control flow obfuscation, encryption, reflection, dynamically loaded code, repackaging etc. [4]. For dynamic analysis, anti-analysis, mimicry, data obfuscation, misleading information flows and function in-directions etc. are used as evasion techniques [4]. Besides, limited code coverage lessens the effectiveness of the dynamic analysis.

As static and dynamic analysis have their weaknesses individually, combining both analysis into a common ground would be helpful in this regard. Hybrid analysis approach integrates both static and dynamic analyses to take advantage of their strengths and to mitigate their weaknesses. Thats why hybrid analysis is so crucial to concentrate. Though hybrid analysis is complex enough, it is effective and feasible according to related research. Previously, researchers have been more focused on static and dynamic analysis. As a result, a lot of research works are carried out in those domains, but comparatively a few works have been performed in hybrid analysis. Researchers nowadays give more focuses on it because of its effectiveness and potential.

Though there exists some reviews on android malware detection, none of them focused on hybrid analysis with machine learning. Tam et al. [4] depicted the evolution of android malware and analysis techniques, but they did not give too much focuses on hybrid analysis. Qamar et al. [5] presented an all-inclusive review on mobile malware, though they nearly overlooked hybrid analysis approach. Baskaran et al. [6] covered hybrid analysis in their android malware detection review in parallel with static and dynamic analysis. Naway et al. [7] focused on deep learning techniques and Feizollah et al. [8] investigated feature selection for analysis in their reviews. None of them provided an in-depth investigation of hybrid analysis.

Due to hybrid analysis approach's huge potential and importance in android malware detection, a brief review of the existing researches on hybrid analysis is necessary. In this work, we provide a comprehensive and systematic review of hybrid analysis approach in android malware detection, analyzed the existing works: their strengths and weaknesses and discussed about challenges, limitations and future directions

in this regard.

To be specific, this work makes the following contributions:

1) To the best of our knowledge, this is the first review on hybrid analysis approach in android malware detection.
2) This work presents the importance of hybrid analysis over static analysis and dynamic analysis by analyzing their weaknesses.
3) This work provides a thorough and systematic review of the existing works on hybrid analysis approach and an analysis of their pros and cons.
4) This work provokes a discussion about the challenges, limitations and future directions regarding hybrid analysis for android malware detection.

II. Background

A. Android Malware

Android malware is an application running on android operating system that implicitly or explicitly performs malicious activities. It includes viruses, worms, Trojan horses, ransomware, spyware and other malicious applications. Android malware tends to cause many malevolent things such as - disrupting normal functioning, taking access controls, leaking information, root exploitation, manipulating data, private content exposed, phishing, policy misconfiguration, disruption of services, getting control of device without users knowledge etc. [5].

Moreover, malware is growing exceedingly to keep pace with the immense growth of android applications. In each month, on average almost 10 million new malwares are introduced [9]. According to report, new malware is found in every 10 seconds [10]. Most alarming fact is that, nowadays noxious malware authors also aware of the malware detection system and they use many novel and crafty evasion techniques to avoid detection. To fight against these cunning black hats, incorporating the most up-to-date and comprehensive detection technique is compulsory.

B. Detection Techniques

Researchers generally analyze android malware with the following three approaches:

(I) Static Analysis
(II) Dynamic Analysis
(III) Hybrid Analysis

In static analysis, various static features are extracted from source code and meta-data. If the source code is not available, reverse engineering is applied to reproduce the source code. According to the static features, a detection model is built using machine learning techniques to classify android malware. Researchers used Androguard, ApkTool, Appknox, DroidMat etc. tools for static analysis. According to the existing research [11]–[15] in static analysis, the most used static features are as follows:

(i) Permissions
(ii) Intents
(iii) Opcode
(iv) Hardware Usage Analysis
(v) Meta-data
(vi) Intents
(vii) API Calls
(viii) Intents
(ix) Suspicious Files
(x) Potentially Dangerous Functions and Methods

Dynamic analysis deals with the dynamic features/behaviors of an application. To track the dynamic behaviors of an application, the application is to be run/executed on an emulated environment or on a real device. According to the dynamic features, a detection model is built using machine learning techniques to classify android malware. Researchers used Droidbox, Marvin, Cuckoo Sandbox, AppsPlayground, DroidLogger etc. tools for dynamic analysis. According to research [16]–[19] in dynamic analysis, the most used static features are as follows:

(i) System Calls
(ii) Network Traffic
(iii) Running Services
(iv) File Operations
(v) Network Operations
(vi) Phone Events

Hybrid Analysis incorporates both static and dynamic features for detecting android malware. As it deals with both static and dynamic features at the same time, it is more complex and costly with respect to time and effort. Maybe this is the reason for the fact that there is not a good number of researches on hybrid analysis in comparison with the static or dynamic analysis. Andrubis, AndroData etc. are used by the researchers for hybrid analysis.

C. Limitations of Static and Dynamic Analysis

Static Analysis faces many troubles such as data obfuscation, control flow obfuscation, encryption, reflection, dynamically loaded code, repackaging etc. [4] by the shrewd malware authors.

On the other hand, Dynamic Analysis also have some drawbacks. To evade dynamic analysis, anti-analysis technique is used frequently by malware authors to detect virtual machines or emulated environments. If the application detects emulated environments in advance, they will act like benign application. By doing so, dynamic analysis might fail to detect android malware. Besides, malware authors use mimicry, data obfuscation, misleading information flows and function in-directions etc. to evade dynamic analysis [4]. The biggest weakness of dynamic analysis is limited code coverage: covering all path is not feasible when investigating dynamic behavior of an application.

III. Hybrid Analysis using Machine Learning

Hybrid analysis integrates both static and dynamic features for more effectiveness than the static or dynamic analysis. Firstly, this hybrid approach seeks to extract the static features and dynamic features of android applications. After that, those extracted static and dynamic features are combined to build a

detection model. Extracting and combining static and dynamic features is the most challenging task of the hybrid analysis. Finally, according to the combined static and dynamic features, a detection model is built using machine learning techniques to classify android malware. Figure 1 depicts the overview of hybrid analysis approach using machine learning.

By incorporating static and dynamic approach into a common ground, hybrid analysis leads to more complexity in android malware detection. The detection process is more likely to take more time and effort. Though hybrid approach might be more effective for android malware detection than the static or dynamic approach, accomplishing a viable malware detection technique is difficult.

As hybrid approach is the combination of static and dynamic approach, this approach can overcome the individual weakness as well as can accumulate the advantages of them. Thereby, hybrid approach strengthens the detection process with the cost of time and complexity. Hybrid methods can also increase robustness, monitor edited apps, increase code coverage and find vulnerabilities [4].

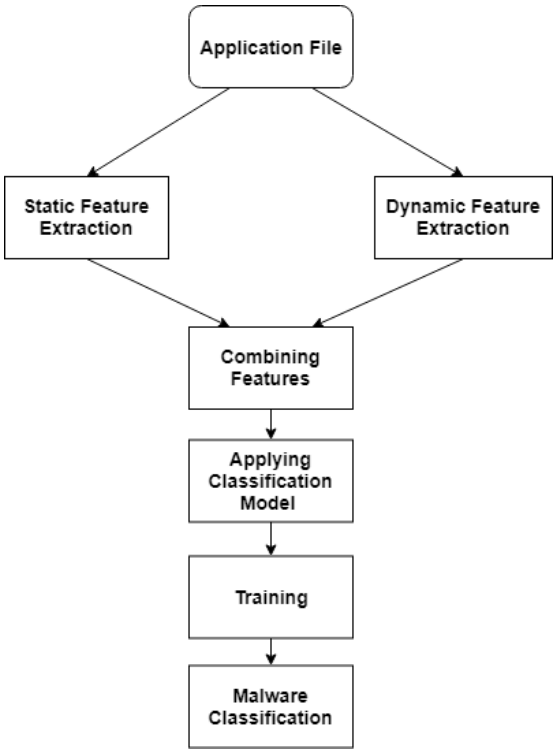

Fig. 1. Hybrid Analysis using Machine Learning

IV. Methodology

To build up a systematic literature review, we have followed a state-of-the-art guideline presented by Kitchenham and Stuart [20]. According to the guideline, Developing a review protocol is compulsory to shape a systematic review. The review protocol includes:

- The rationale for the review
- Research questions
- Search strategy
- Study selection criteria
- Study selection procedures
- Study quality assessment
- Data extraction
- Data synthesis

It is not mandatory to follow all the steps given by the protocol. Only relevant steps should be done, other steps could be overlooked. Details of each step taken in our work is described in the following subsections.

A. The Rationale for the Review

Hybrid analysis using machine learning for android malware detection is a promising research domain because the weaknesses of static and dynamic analysis approach have lessened their effectiveness. Though there is a few researches so far in this domain, the potentiality of this domain needs a brief review of the existing literature. To identify the overview, methodology, contributions, challenges, pros and cons of the existing literature and to present a guideline for the future researchers, an inclusive review of hybrid analysis is essential.

B. Research Questions

Identifying the research questions is the most key part of a systematic review. To present a systematic review of hybrid analysis using machine learning, we have identified the following research questions:

1) What are the static and dynamic features used in hybrid analysis using machine learning?
2) What are the most common dataset sources of the existing literature?
3) Which machine learning algorithms are most frequently used in the existing researches?
4) Which evaluation metrics are most widely used in the existing literature?
5) What are the evaluation results of the existing researches?
6) What are the limitations of the existing literature?

C. Search Strategy

To achieve most relevant relevant result, we have identified keywords and terms regarding hybrid analysis. The keywords and terms are kept relevant with the research question. We have also employed alternative spelling and synonyms of the keywords. We have used 'AND' and 'OR' operators to filter our expected search results. The resulting search strings are as follows:

- (hybrid OR static and dynamic) AND (analysis OR approach OR technique) AND (android* OR mobile) AND (malware OR malicious*) AND (detection OR classification) AND (machine learning OR deep learning)
- (android malware detection) AND (hybrid static and dynamic) AND (analysis OR approach OR technique)

D. Study Selection Criteria

From the search results, we have to include and exclude researches for our primary study according to a predefined criteria. That inclusion and exclusion criteria play a vital role to select relevant researches.

1) Inclusion Criteria:

- Journal, Conference Proceedings related to hybrid analysis using machine learning for android malware detection
- Date (year) of publication: 2012-2019
- Most recent version of the research paper

2) Exclusion Criteria:

- Research that uses - hybrid keyword, but not directing to the hybrid analysis (combination of static and dynamic analysis) we have mentioned
- Research that incorporates hybrid analysis, but not employing machine learning techniques
- Research that lacks of a well-defined methodology and unambiguous contributions

E. Study Selection Procedures

We have searched relevant researches in many research databases and electronic sources according to search strategy. We have focused on the prominent sources with regard to our research domain. We have performed our search in the following electronic sources and search engines:

- Google scholar
- ACM Digital library
- IEEExplore
- Inspec
- Springer
- Elsevier

Besides, we examined the related research or equivalent section of the selected research papers for further references of relevant papers. We also looked over the researches that cites our selected research papers.

F. Study Quality Assessment

We have scrutinized the selected papers for bias, internal validity and external validity. Though there is no consensus about the interpretation of - quality, the CRD Guidelines [21] and the Cochrane Reviewers Handbook [22] suggest that quality correlates insofar as the study minimizes bias and maximizes internal and external validity [20]. We also considers the study comparison method and dataset size for quality assessment.

G. Data Extraction

To keep track of the extracted data/information, a data extraction form have been maintained. Table I depicts the contents of the form.

TABLE I
DATA EXTRACTION FORM

Data Item	Value	Remarks
Paper ID		
Paper Name		
Author Name		
Publishing Year		
Publisher		
Journal/Conference Name		
Research Question 1		
Research Question 2		
Research Question 3		
Research Question 4		
Research Question 5		

V. SYSTEMATIC LITERATURE REVIEW

A systematic review on hybrid analysis can be accomplished by addressing and answering the research questions mentioned before. In the following section, we have resolved the research questions and presented an inclusive systematic review of the consequential researches in hybrid analysis. At first, the first four research questions are resolved according to the existing researches. Then each of the researches on hybrid analysis are discussed briefly; the last two research questions are sorted out in that part.

Permissions and API Calls as static features and System Calls as dynamic features are most frequently used in the existing researches.

The most common dataset according to the existing researches are Drebin and Android Malware Genome Project. Besides, most researches use Google Play Store and local app stores to collect benign applications. ContagioDump, VirusTotal, VirusShare etc. sources are also used for malware samples.

Support Vector Machine (SVM) is the most frequently used machine learning algorithm in the existing research. Besides, Naive Bayes, Random Forest, J48, Logistic Regression etc. are also common in the existing researches.

Accuracy, True Positive Rate (TPR), False Positive Rate are the most common evaluation metrics according to the existing researches.

One of the state-of-the-art work in hybrid analysis, Marvin [23] employed a lot of static and dynamic features to detect android malware. It extracted Permissions, Intents, Suspicious Files, API Calls, Developers Certificate etc. as static features and File Operations, Network Operations, Phone Events, Dynamically Loaded Code etc. as dynamic features. It used SVM and Linear Classifier (Regularized Logistic Regression) to build detection model where Linear Classifier can detect more accurately but SVM is faster comparatively. For labeled test data, Marvins performance is sound enough as its accuracy to detect malware is 98.24 % with less than 0.04% false positive rate. But for previously unseen malware, its accuracy is close to 90%. Besides, to avoid obsolescence of its classification model in future, it presented a retraining strategy. Though Marvin considers a lot of features, it overlooked system-level events such as System Calls : an integral part of the behavioral aspects (dynamic features).

Mobile-SandBox [24] used Permissions, Services, Receivers, Intents, Potentially dangerous functions and methods as static features and investigated Native Code (Native API Calls) and Network Traffic as dynamic features to classify malware. It lacks in performance as it did not provide any solid performance metrics.
Samadroid [25] presented an on-device malware detection architecture which ensures the resource efficiency by reducing memory overhead of local devices. It used a subset of Drebin's [11] features (6 out of 8) as static features and 10 predefined System Calls as dynamic features. Its accuracy is almost 98% with a false positive rate of 0.1%. Though it incorporated System Call into its feature space and outperform Drebin [11], but it used old dataset. Thereby it might fail to fight against recent malware as malware behavior changes frequently over time. It also overlooked any additional dynamic features.
Kapratwar et al. [26] used Permissions and System Calls for hybrid analysis. Its performance (AUC) is significantly better for static features in comparison with dynamic features. But it used a small (200 apps) and old dataset and overlooked other static and dynamic features.
Hadm [27] incorporated Deep Neural Network for feature extraction from a set of static and dynamic features. It exhibited that combining advanced features derived by deep learning with the original static and dynamic features provides consequential returns. It achieved 94.7% accuracy with a false positive rate of 1.8% where with the original features the best accuracy is 93.5%. An improvement of 1.2% with the cost of complexity.

Dhanya et al. [28] used Permissions as static and API Calls as dynamic feature. Separability assessment Criteria is used for feature selection in this research. Using the 77 selected features and four different machine learning algorithms (Naive Bayes, SVM, J48 & Random Forest), they evaluated their work. Their performance regarding F-measure, precision and recall is dubitable as they used Drebin, an outdated and limited dataset. Besides they did not consider any other features except Permissions and API Calls.
Liu et al. [29] proposed a hybrid malware detecting scheme for android where Permissions and API Calls are used as static features and System Calls used as dynamic features. Their scheme's detection accuracy is from 93.33% to 99.28% according to experimental results. Though they considered only a small feature-set and their dataset is also limited.
Table II depicts the literature overview of hybrid analysis using machine learning.

VI. Discussion

In this section, we have pointed out the opportunities, challenges, limitations and future directions of hybrid analysis using machine learning for android malware detection.

A. Lack of Research

As mentioned before, there is not enough research in hybrid analysis, though it is a promising and effective approach in android malware detection. So, researchers have to put more emphasize in this regard. A lot of opportunities and research directions are available right now. Researchers' enthusiastic focus on this field would have been beneficial from android application security perspective as well as from academia perspective. To fight against the escalating malware authors community, more research work is essential.

B. Dataset Inadequacy

Malware is growing enormously in every second, but there does not exist any up-to-date dataset for the researchers. Previously stated, almost 10 million new malwares are found in each month [18]. But most of the dataset used in research is dated and obsolete nowadays. Thereby, their performance in android malware detection is doubtful considering the vast population of the new malware. Dataset inadequacy is a vital factor as dataset is responsible for the evaluation of any research. So, android malware dataset has to be updated on a regular basis to assure the effectiveness of the new research and to justify the feasibility of the existing research.

C. Better Performance

Despite there is not so many researches are carried out in android malware detection using hybrid analysis, those few researches in this domain exhibit better performance on average than the typical static and dynamic approaches and engender a lot of opportunities. By grabbing those opportunities and overcoming the challenges ahead, hybrid analysis using machine learning would be a vanguard for android malware detection in future.

D. New Malware Family

As existing malware behavior is decoded by the existing tool or research outcome, malware authors update existing malware families and create new malware families frequently to evade detection. New malware families is a concerning issue. Because their behaviors are mostly unfamiliar to the typical detection system. They try to trick existing detection systems by introducing new behavior as well as exhibiting benign behavior. So researchers should consider this issue carefully to ensure security. *How do we detect new malware families effectively* - is a promising research domain in that regard.

E. Exploring New Feature

Most of the existing research dealt with some common features such as Permissions, API Calls, Intents, App Components, System Calls, File Operations, Network Operations, Phone Events, Dynamically Loaded Code etc. But it would be possible that there exists more distinguishable features to detect android malware. In this regard, Talha et al. [30] revealed many unknown characteristics of android malware, however it did not integrate any machine learning technique to detect android malware. They revealed that over-privileged permissions is one of the characteristics of malware. Besides they uncovered that malware's average number of incoming and outgoing connections, average size of download and

TABLE II
SYSTEMATIC LITERATURE OVERVIEW OF HYBRID ANALYSIS USING MACHINE LEARNING

Ref.	Publishing Year	Static Features	Dynamic Features	Dataset Source	Dataset Size	Algorithms	Metrics	Values	Limitation
Marvin [23]	2015	Permissions, Intents, Suspicious Files, API Calls, Developer's Certificate etc.	File Operations, Network Operations, Phone Events, Dynamically Loaded Code etc.	Google Play Store, VirusTotal, Genome Project, Contagio	150,000 apps (135,000 benign, 15,000 malware)	SVM and Linear Classifier (Regularized Logistic Regression)	Accuracy, FPR	98.24%, $<$0.04%	Overlooking system-level events such as System Calls
MobileSandBox [24]	2013	Permissions, Services, Receivers, Intents, Potentially Dangerous Functions and Methods	Native Code (Native API Calls) and Network Traffic	Asian markets and Google Play Store	40,000 apps				Lacking in performance as no solid performance metrics given
Samadroid [25]	2018	Permissions, API Calls, Intents, App Components	System Calls (10)	Drebin	5,560 apps	SVM, Naive Bayes, Decision Tree and Random Forest	Accuracy, TPR, FPR	91.6%$\sim$ 98.97%, 81.1%$\sim$ 98.5%, 0.03%$\sim$ 7.8%	Overlooking many dynamic features; using limited and old dataset
Kapratwar et al. [26]	2017	Permissions	System Calls	Google Play Store, VirusTotal, Drebin	200 apps (103 benign, 97 malware)	Nave Bayes, J48 & Random Forest, Simple Logistic, IBk	AUC	0.5844$\sim$ 0.9660	Overlooking many static and dynamic features; using small and old dataset
Hadm [27]	2016	Permissions, API Calls, Intents	System Call Sequences	Google Play and VirusShare	5888 apps (4002 benign, 1886 malware)	Deep Neural Network, SVM, Hierarchical Multiple Kernel Learning	Accuracy, FPR	94.7%, 1.8%	Higher complexity with respect to accuracy gains
Dhanya et al. [28]	2019	Permissions	API Calls	Drebin	400 apps (200 benign, 200 malware)	Nave Bayes, SVM, J48 & Random Forest	F-measure, Precision, Recall	0.71%$\sim$ 0.975, 74.7%$\sim$ 97.6%, 72.5%$\sim$ 97.5%	Using limited and old dataset; considering few features
Liu et al. [29]	2016	Permissions	System Calls	Gnome Project, Wandoujia App Market	1000 apps (1000 benign, 1000 malware)	SVM, KNN	ACC, TPR, FPR	93.33%$\sim$ 99.28%, 94.59%$\sim$ 99.47%, 0.20%$\sim$ 11.01%	Using limited dataset, considering few features

upload, average number of INTERNET_CLOSE action are distinguishable features with respect to benign applications. Looking for more discernible features might create new opportunities in android malware detection.

F. Reducing Complexity

Since hybrid approach combines static and dynamic approach, its overall complexity is higher with respect to time, cost and effort. *How do we reduce the complexity of hybrid analysis* - would be a potential direction for future researchers. The outcome of such an effort would present a more viable and effective approach.

VII. Conclusion

Android malware is the key factor for the most security breaches in android operating system. Currently malware authors are sharp-witted enough to evade the typical anti-viruses or obsolete approaches of malware detection. As android malware generally tries to preserve the facade of a benign application using multifarious evasion techniques, it is worthy and necessary to take a perceptive approach to defend them.

Detecting android malware effectively and feasibly in advance is the biggest challenge of this fast-growing digital world. Hybrid analysis approach has the capability and can offer a sound direction on this subject. By exploring this field, researchers have published some remarkable research already. This work tends to highlight those researches by providing a comprehensive and systematic review of them. Besides it points out the specific challenges, limitations and future directions in hybrid analysis. By doing so, this research seeks to contribute to academia as well as raise concern for android mobile application security.

References

[1] N. G., "Android: Market share other stats [infographic]," Jul 2019.

[2] B. Popper, "Google announces over 2 billion monthly active devices on android," May 2017.

[3] C. Lueg, "Cyber attacks on android devices on the rise," Nov 2018.

[4] K. Tam, A. Feizollah, N. B. Anuar, R. Salleh, and L. Cavallaro, "The evolution of android malware and android analysis techniques," *ACM Computing Surveys (CSUR)*, vol. 49, no. 4, p. 76, 2017.

[5] A. Qamar, A. Karim, and V. Chang, "Mobile malware attacks: Review, taxonomy future directions," *Future Generation Computer Systems*, vol. 97, 03 2019.

[6] B. Baskaran and A. Ralescu, "A study of android malware detection techniques and machine learning," 2016.

[7] A. Naway and Y. Li, "A review on the use of deep learning in android malware detection," *arXiv preprint arXiv:1812.10360*, 2018.

[8] A. Feizollah, N. B. Anuar, R. Salleh, and A. W. A. Wahab, "A review on feature selection in mobile malware detection," *Digital investigation*, vol. 13, pp. 22–37, 2015.

[9] "Malware statistics trends report: Av-test," Apr 2019.

[10] "Android malware threat profile."

[11] D. Arp, M. Spreitzenbarth, M. Hubner, H. Gascon, K. Rieck, and C. Siemens, "Drebin: Effective and explainable detection of android malware in your pocket.," in *Ndss*, vol. 14, pp. 23–26, 2014.

[12] S. Y. Yerima, S. Sezer, G. McWilliams, and I. Muttik, "A new android malware detection approach using bayesian classification," in *2013 IEEE 27th international conference on advanced information networking and applications (AINA)*, pp. 121–128, IEEE, 2013.

[13] E. Mariconti, L. Onwuzurike, P. Andriotis, E. De Cristofaro, G. Ross, and G. Stringhini, "Mamadroid: Detecting android malware by building markov chains of behavioral models," *arXiv preprint arXiv:1612.04433*, 2016.

[14] M. Ghorbanzadeh, Y. Chen, Z. Ma, T. C. Clancy, and R. McGwier, "A neural network approach to category validation of android applications," in *2013 International Conference on Computing, Networking and Communications (ICNC)*, pp. 740–744, IEEE, 2013.

[15] Q. Jerome, K. Allix, R. State, and T. Engel, "Using opcode-sequences to detect malicious android applications," in *2014 IEEE International Conference on Communications (ICC)*, pp. 914–919, IEEE, 2014.

[16] L. K. Yan and H. Yin, "Droidscope: Seamlessly reconstructing the {OS} and dalvik semantic views for dynamic android malware analysis," in *Presented as part of the 21st {USENIX} Security Symposium ({USENIX} Security 12)*, pp. 569–584, 2012.

[17] B. Amos, H. Turner, and J. White, "Applying machine learning classifiers to dynamic android malware detection at scale," in *2013 9th international wireless communications and mobile computing conference (IWCMC)*, pp. 1666–1671, IEEE, 2013.

[18] W.-C. Wu and S.-H. Hung, "Droiddolphin: a dynamic android malware detection framework using big data and machine learning," in *Proceedings of the 2014 Conference on Research in Adaptive and Convergent Systems*, pp. 247–252, ACM, 2014.

[19] W. Enck, P. Gilbert, S. Han, V. Tendulkar, B.-G. Chun, L. P. Cox, J. Jung, P. McDaniel, and A. N. Sheth, "Taintdroid: an information-flow tracking system for realtime privacy monitoring on smartphones," *ACM Transactions on Computer Systems (TOCS)*, vol. 32, no. 2, p. 5, 2014.

[20] B. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering," 2007.

[21] K. S. Khan, G. Ter Riet, J. Glanville, A. J. Sowden, J. Kleijnen, *et al.*, *Undertaking systematic reviews of research on effectiveness: CRD's guidance for carrying out or commissioning reviews*. No. 4 (2n, NHS Centre for Reviews and Dissemination, 2001.

[22] J. P. Higgins and S. Green, *Cochrane handbook for systematic reviews of interventions*, vol. 4. John Wiley & Sons, 2011.

[23] M. Lindorfer, M. Neugschwandtner, and C. Platzer, "Marvin: Efficient and comprehensive mobile app classification through static and dynamic analysis," in *2015 IEEE 39th annual computer software and applications conference*, vol. 2, pp. 422–433, IEEE, 2015.

[24] M. Spreitzenbarth, F. Freiling, F. Echtler, T. Schreck, and J. Hoffmann, "Mobile-sandbox: having a deeper look into android applications," in *Proceedings of the 28th Annual ACM Symposium on Applied Computing*, pp. 1808–1815, ACM, 2013.

[25] S. Arshad, M. A. Shah, A. Wahid, A. Mehmood, H. Song, and H. Yu, "Samadroid: a novel 3-level hybrid malware detection model for android operating system," *IEEE Access*, vol. 6, pp. 4321–4339, 2018.

[26] A. Kapratwar, F. Di Troia, and M. Stamp, "Static and dynamic analysis of android malware.," in *ICISSP*, pp. 653–662, 2017.

[27] L. Xu, D. Zhang, N. Jayasena, and J. Cavazos, "Hadm: Hybrid analysis for detection of malware," in *Proceedings of SAI Intelligent Systems Conference*, pp. 702–724, Springer, 2016.

[28] K. D. T. Gireesh Kumar, "Efficient android malware scanner using hybrid analysis," *International Journal of Recent Technology and Engineering(TM)*, vol. 7, pp. 76–80, 2019.

[29] Y. Liu, Y. Zhang, H. Li, and X. Chen, "A hybrid malware detecting scheme for mobile android applications," in *2016 IEEE International Conference on Consumer Electronics (ICCE)*, pp. 155–156, IEEE, 2016.

[30] A. T. Kabakus and I. A. Dogru, "An in-depth analysis of android malware using hybrid techniques," *Digital Investigation*, vol. 24, pp. 25–33, 2018.