

PAPER • OPEN ACCESS

Web application scanning for malware attack detection with provide appropriate incident report by using hybrid method

To cite this article: Afiza Ariffin and Aina Razak 2020 *J. Phys.: Conf. Ser.* **1529** 052084

View the [article online](#) for updates and enhancements.

You may also like

- [An Analysis of Machine Learning-Based Android Malware Detection Approaches](#)
R. Srinivasan, S Karpagam, M. Kavitha et al.
- [PAM Clustering Aided Android Malicious Apps Detection](#)
Nibras Talib Mohammed, Mohsin Hasan Hussein and Abbas Jabber Rashid
- [Enhancing blockchain security: a novel approach to integrated malware defence mechanisms](#)
Aastha Sharma, Divya Upadhyay and Shanu Sharma



PRIME
PACIFIC RIM MEETING
ON ELECTROCHEMICAL
AND SOLID STATE SCIENCE

HONOLULU, HI
October 6-11, 2024

Joint International Meeting of
The Electrochemical Society of Japan (ECSJ)
The Korean Electrochemical Society (KECS)
The Electrochemical Society (ECS)

Early Registration Deadline:
September 3, 2024

**MAKE YOUR PLANS
NOW!**

Web application scanning for malware attack detection with provide appropriate incident report by using hybrid method

Afiza Ariffin¹, Aina Razak¹

¹Department of Computer Science, Faculty of Computer Science & Information Technology, University Putra Malaysia, 43400 Serdang, Selangor, Malaysia

Abstract. In the era of Internet of Things (IoT) technology, potential malware that has the ability to attack any computer software, server or network from various way are increasingly due to the common use of anti-malware software and some of malware, which is not detectable by antivirus. Seeing that the number of malware is increasing rapidly and each malware have different way to attack, it become more challenges to the existing malware detection to stave off extensively. Malware detection is a process to perform analysis of malware on examine the components and behaviours of malware. Malware detection involve two techniques, which consists of static and dynamic analysis techniques. Static analysis technique examine malware without running it or not viewing the actual code. It will employs different tools to identify malicious file, provide the functional information and collect technical indicator to produce signature. For dynamic analysis technique also called as behaviour analysis technique, it runs malware to observe its behaviour, understand the functionality and identify the technical indicator for signature detection. Although there are advantages to conducting static and dynamic separately, but there are some limitation. Implement static and dynamic analysis techniques together are more valuable for reverse engineering complex malware. It will helps to identify the true intent and capabilities of malware and can provide a technical indicator, which cannot be achievable individually. This research propose a hybrid technique, which integrates static and dynamic analysis technique, to examine malware and measure the effectiveness based on the detection time. The proposed scheme should able to run and examine a malware with a low detection time. Besides, this research will also come with some suggestion to be taken once the malware is detected. By using Java language, the web application to analysis malware by using hybrid technique will be presented.

1. Introduction

Nowadays, Internet is become an important part in people's life. People use the Internet in their daily activities such as online payments, online banking payments and communicate with other people using the social media like Facebook, Twitter, Instagram and so on. Therefore, the rate at which Internet users an exposed to security threat due to malware attacks is extremely high. Indirectly, the number of criminal program and illegal program also will grow quickly. Most of these programs are created to support the growth of organization computer crime [1]. Criminals will use malware to take over computers and steal confidential data, personal data or any information for profit [1]. Due to the increase in malware and computer crimes, it is forcing digital forensic investigators to go deeper in malware analysis and detection. Today, malware forensics has become part of computer forensics to identify and analysis unknown malware [13].

Malicious software also known as Malware is any malicious code in software that can be used to compromise computer operations, collect sensitive information, do illegitimate action on data, gain



access to private computer resources, host or networks. It will damage computer programs without user consent. Malware are able to exploit resource from various system platforms. Malware also refers to malicious programs or files and will have an adverse effect on computer users [2]. Malware comes with different types of threats such as virus, worms, Trojans, rootkits and so on. Malware considered as high-level issues because it had potential to attack the security goals, which are confidentiality, integrity and availability. Malware inflicts on various objectives such as stealing, encrypting or removing sensitive data, altering or retrieving the rights of core computing functionality as well as monitoring the activities of computer users without their knowledge and consent. Because of this, the number of everyday and existing emerging malware in this extremely high range will evolve in their structure as well as difficult to detect.

Malware analysis could be retrieved either by using static or dynamic method. Static analysis act to examine the malware without actually running it while dynamic analysis will execute malware in a controlled and monitor the environment to observe detailed particular process of malware detection that will analyse the whole process behaviour of malware. Each technique comes with different element to categorize either as basic or advanced. There consist of some advantage and limitation based on the method of malware analysis.[3].

Static malware analysis uses a signature-based approach for example that involves file fingerprints, virus scanning, reverse engineering the binary, file obfuscation, analyse memory artefact, packer detection and debugging. Signature based is identify the presence of malware that infect by match at least one-byte signature also known as blacklist. It is ineffective against the sophisticated malware programs and codes. Static analysis fails at different code by using obfuscation technique used by virus coders also polymorphic and metamorphic malware but there is advantage from binary code information that contains very useful information about malicious behaviour of program in term of code sequence and parameters [4].

Dynamic malware analysis uses a behaviour-based approach for malware detection that will analyse the suspicious activity. It involves the API call, by intrusion detection traces, any changes of registry, calls for network and system also the memory write. It is effective against all types of malware because it will execute the sample of malware. However dynamic also have some limitation for obfuscation techniques and polymorphic malware but it is necessary complement compare to static approach [5]. Hybrid is the combination of Static and dynamic technique. This project will use integrated static and dynamic method which known as hybrid to analyse the malware attack detection.

Hybrid malware analysis method will collect information about malware from static and dynamic analysis. By implement this method, it will reduce the limitation from both method either static or dynamic analysis. Therefore, improving the ability to detect the program intend to properly [6]. It will be observed by the code analysis by checking the signature of malware and running in the virtual environment to observe actual behaviour.

2. Related Work

This research proposed the integrated both malware analysis methods, which are static and dynamic analysis, named as hybrid malware analysis. However, some related works focus on static analysis and some of that using the dynamic analysis in order to detect the malware attack patterns.

2.1 Scalable malware classification based on integrated static and dynamic features [10]

This research goal is using the combination of both static and dynamic feature that being complete and robust solution to evasion techniques used by virus writer. The objective is to improve classification accuracy while also support with scalability. The accuracy is approving by the integrated static and dynamic features. As for scalability is measure based on reducing the feature space by select the most dominant static features and dynamic features. Based on this research, they aim to classify malware by reducing featured space not to analyse and identify of malware binary file.

2.2 Droid Detector: Android Malware Characterization and Detection Using Deep Learning [11]

This research is run on Android platform. This research highlights the issues unknown hidden in a large number of malware becomes highly serious threaten by Android security. This paper use deep

learning methodology to increase the attention in artificial intelligence. By using deep learning technique that can automatically detect whether an application is a malware or not. Based on this research, they aim to detect either malware is existing or not but not to find the accuracy feature based on static and dynamic features.

2.3 Manual Malware Analysis Using Static Method [12]

This research is focus to analyze malware using static method in operating system environment by using Anti-Virtualization avoidance technique. This is capable to do malware analysis more accurate compare with virtual machine environment. By using the VMware this research success to identify which malware may refuse or bypass to run in a virtual technology. Based on this research, they only proposed the static method analysis so that might be occur issue when it not goes to signature based approach.

Table 1. Existing Research.

Author, Year	Methodology	Finding
B.Tewfik,B. Zakaria A. Nemrat, B. Chafika,et al, 2016	Using Machine Learning for malware classification extract the static and dynamic analysis feature to get the integrated feature vector of malware	Based on this project, they aim to classify malware by reducing featured space not to analyze and identify of malware binary file.
Mohammad, A. Mamoun, A.Izzat,Z.Mohammad,et al. 2016	Imply the multi-faceted detection decisions to evaluate collective analysis and removal decision by backup the file	This proposed is beyond to API from various malware analysis website scanner, once the API is robust the result of collective analysis will change and accuracy is difficult to examine
P.V.Shijo, A.Salim,et al. 2015	Support vector machine learning technique to classify data using integrated static and dynamic analysis	This project only apply two algorithm to classify the unrated and rated data and not to reduce the features for each method.
V.Deepti, S.P.Choudary., et al, 2015	See the pattern of operation in specific location and each define different text word in dynamic report	This proposed comes to detect the pattern of malware by using dynamic analysis method using larger sample space and able to cluster and classify the class of malware.
Cao, Ying., et al, 2013	Malware Behavior Capturing System by using multi-virtual machine framework.	This research only provides original data behavior for behavior-based malware analysis.

3. Malware Analysis Detection Method

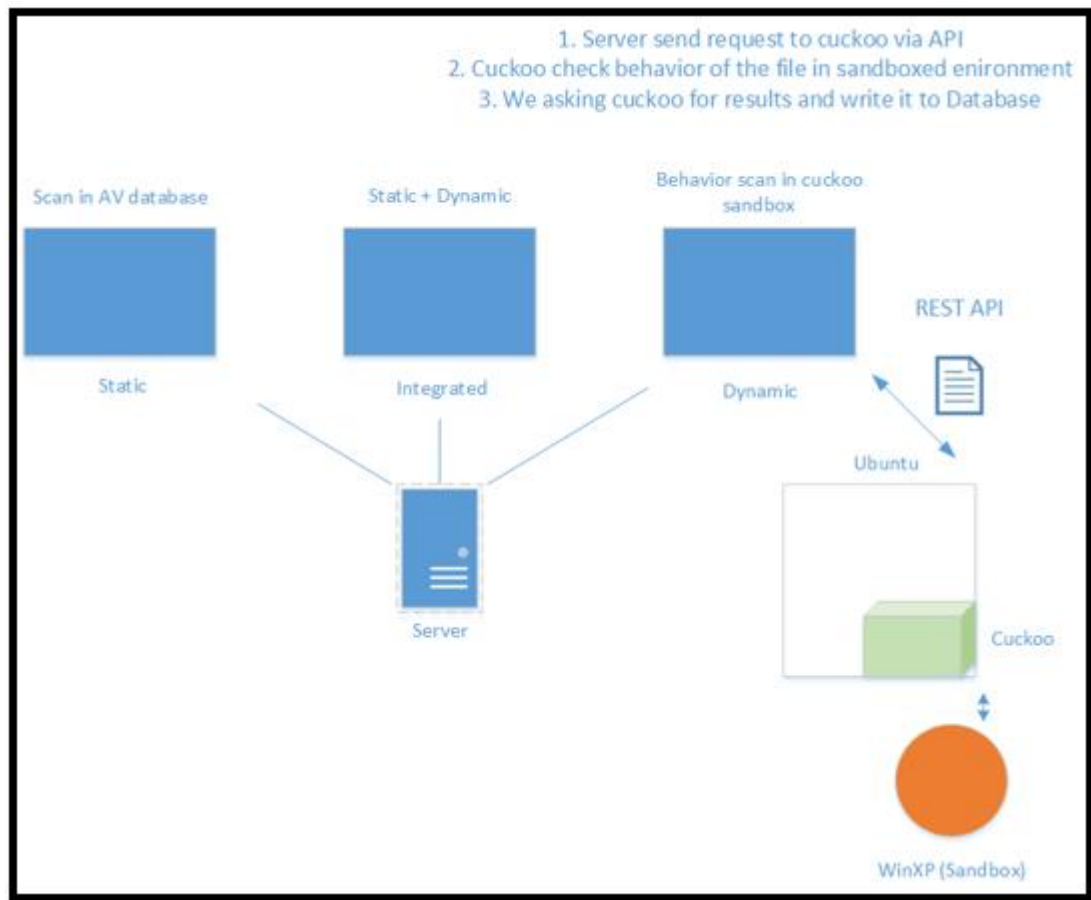


Figure 1. Framework of hybrid method

Above figure, show the framework of hybrid method of web application scanner for malware attack detection using static, dynamic and integrated also known as hybrid technique. For static method, the dataset will scan it in database at server only. However, for dynamic method behaviour scan will scan in Cuckoo sandbox that collaborate the result by Rest API from Ubuntu.

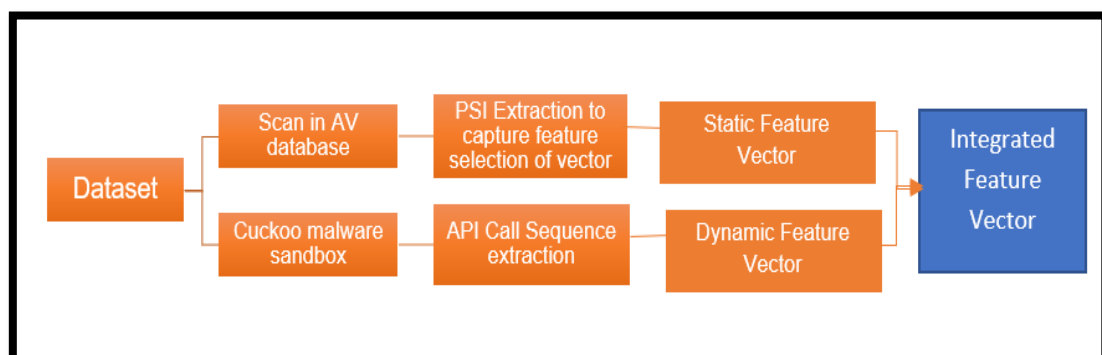


Figure 2. Architecture of hybrid method

Figure 2 shown the architecture of proposed hybrid method. Both analysis is performed by dataset that contain both malicious and benign files. From the training data static analysis will extracting the

PSI feature before selection feature vector creation process. Based on that, the static feature vector will be analyzed as a result. Dynamic analysis will involve the monitoring log files from malware analyzer. It will extract API call sequence.

- **First component** for static analysis and static features is classification task. Feature that gained by extracted malware binary files stated as dataset that contain malware and benign files. Code obfuscation may insert unwanted PSI to the binary files. Based on the PSI for each feature list will create a table of PSI sorted according to the frequency below threshold are eliminated.
- **Second component** for static analysis and static features is create binary feature vector with each PSI in the feature list as attributes. Each of malware and benign file will compared with the list before represent a binary vector denoting the string to classify its contains malware or not and record as a true or false binary value.
- **First component** for dynamic analysis and dynamic feature extracting the API calls gains by binary file while execution. This research examines using cuckoo malware analyzer with virtual machine as the secure environment. Malware files will generate the log file result of the behavior that contains API call, modification of registry and some memory and process address. The attackers use the same set of API call sequence for malicious code attack to find the similarity between files in the same class must be greater than the similarity between the files in different classes.
- **Second component** for dynamic analysis and dynamic feature is sort the global list of both API call grams with frequency of occurrence. N-gram based method proposed to analyses the call sequence called API-call grams. As the size of the n-gram increases the number of similar n-grams between two files within the same class itself is very low.
- **Third component** for dynamic analysis and dynamic feature is create feature vector with both 3 API and 4 API call grams as attribute and will eliminate some API-call-grams that classify as low frequency.

In this research, static and dynamic analysis and features will integrate as hybrid analysis by using machine learning techniques. This research will focus for random forest and support vector machine algorithm for malware classification.

4. Experiments

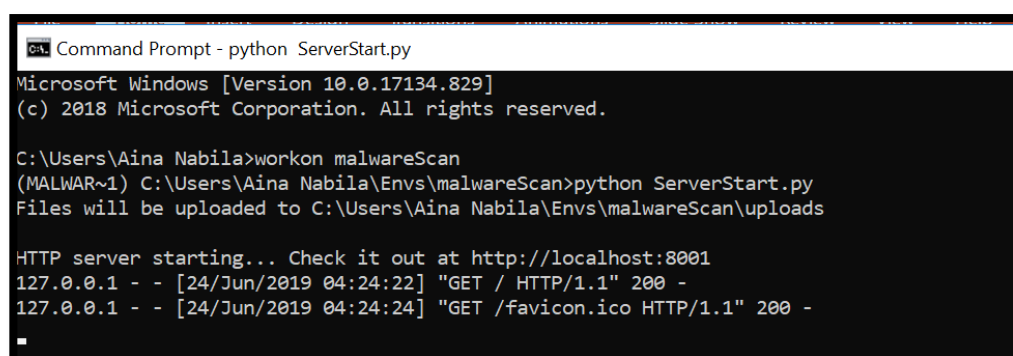
The experimental for training dataset is set up on Ubuntu. This research will used gradient boosting machine learning compare to anchor paper that evaluate the accuracy performance using the random forest technique of machine learning. Gradient boosting is anomaly detection in supervised learning especially in term of security of data. By the detection, it will produce the rank approach of machine learning for information retrieval system. Besides, it will optimize the ranking and regression which random forest is harder to achieve.

1) Set up the server on Command Prompt

Syntax:

workon malwareScan

python ServerStart.py



```
Command Prompt - python ServerStart.py
Microsoft Windows [Version 10.0.17134.829]
(c) 2018 Microsoft Corporation. All rights reserved.

C:\Users\Aina Nabila>workon malwareScan
(MALWAR~1) C:\Users\Aina Nabila\Env\malwareScan>python ServerStart.py
Files will be uploaded to C:\Users\Aina Nabila\Env\malwareScan\uploads

HTTP server starting... Check it out at http://localhost:8001
127.0.0.1 - - [24/Jun/2019 04:24:22] "GET / HTTP/1.1" 200 -
127.0.0.1 - - [24/Jun/2019 04:24:24] "GET /favicon.ico HTTP/1.1" 200 -
```

Figure 3. Server Start on Command Prompt

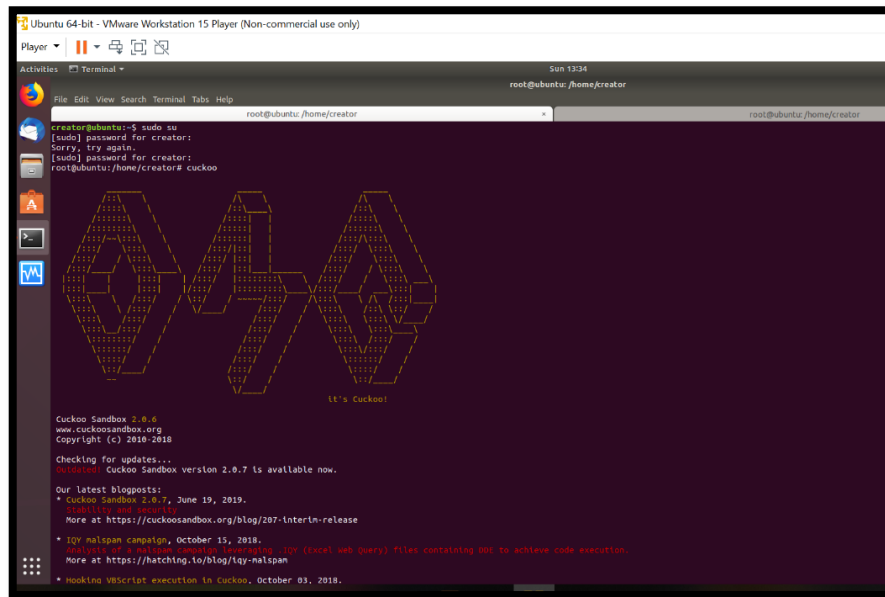
2) Open Ubuntu that already install on VMware

Syntax:

sudo su

First tab: cuckoo

Second tab : cuckoo api -H 192.168.47.128

**Figure 4.** First tab cuckoo malware analyser

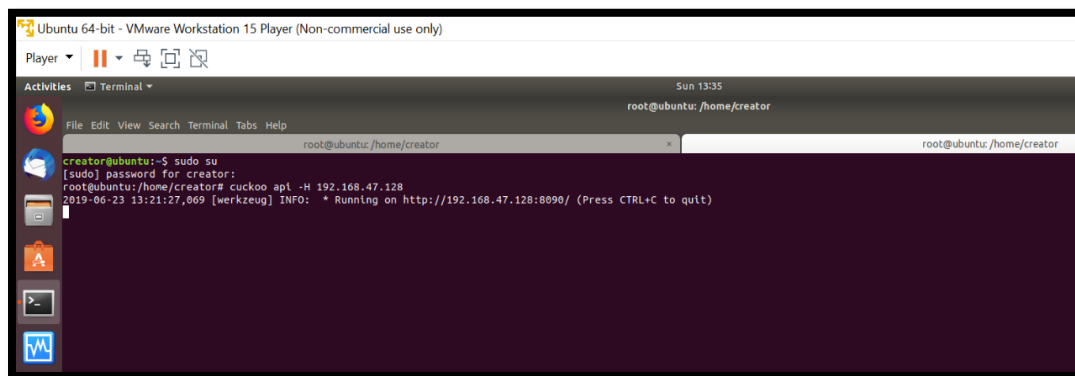


Figure 5. Second tab cuckoo malware analyser

3) Run the web malware scanner on <http://localhost:8001/>

Process 1: Upload and scan the file

Process 2: Wait until the scan finish

Id	FileName	Hash	Type	Time	Result
1	test	b255836a4d9ccc70b11def6f176895b1	static	10	Clear
2	config.json	e1f63416673762fcee27856f74ea7e3	static	69.2801673412323	Clear
3	test_file.txt	fa407356c44ed1f5acaba1573e55c9cf	Static	49.5105438232422	Clear
4	example.png	aae4af68295cdf264e7e5eeceaaafd649	Static	63	Clear
5	example.png	aae4af68295cdf264e7e5eeceaaafd649	Dynamic	193	Clear
6	example.png	aae4af68295cdf264e7e5eeceaaafd649	Integrated	256	Clear
7	output.png	ab2e94e5f0b34f4096566b5d87545932	Static	108	Clear
8	output-1.png	ab2e94e5f0b34f4096566b5d87545932	Static	125	Clear
9	output-1.png	ab2e94e5f0b34f4096566b5d87545932	Dynamic	192	Clear
10	output-1.png	ab2e94e5f0b34f4096566b5d87545932	Integrated	317	Clear
11	2.png	f83a72a9c02aa3d7394f44786fe242cd	Static	100	Clear
12	2.png	f83a72a9c02aa3d7394f44786fe242cd	Dynamic	189	Clear
13	2.png	f83a72a9c02aa3d7394f44786fe242cd	Integrated	259	Clear
14	1.png	e0b035324f70883b82e4a74c78757f72	Static	68	Clear
15	1.png	e0b035324f70883b82e4a74c78757f72	Dynamic	189	Clear
16	1.png	e0b035324f70883b82e4a74c78757f72	Integrated	257	Clear
17	asset.png	03e0392de533914753133df93e4ee9c3	Static	72	Clear
18	asset.png	03e0392de533914753133df93e4ee9c3	Dynamic	187	Clear
19	asset.png	03e0392de533914753133df93e4ee9c3	Integrated	259	Clear

Figure 6. Web scanner interface

5. Results

The result of this research is to ensure the proposed objective is achieved which is to analyse the performance based on time and accuracy taken for each method.

5.1 Result of performance on time for each method

Id	FileName	Hash	Type	Time	Result
1	test	b255836a4d9ccc70b11def6f176895b1	static	10	Clear
2	config.json	e1f63416673762fcee27856f74ea7e3	static	69.2801673412323	Clear
3	test_file.txt	fa407356c44ed1f5acaba1573e55c9cf	Static	49.5105438232422	Clear
4	example.png	aae4af68295cdf264e7e5eeceaaafd649	Static	63	Clear
5	example.png	aae4af68295cdf264e7e5eeceaaafd649	Dynamic	193	Clear
6	example.png	aae4af68295cdf264e7e5eeceaaafd649	Integrated	256	Clear
7	output.png	ab2e94e5f0b34f4096566b5d87545932	Static	108	Clear
8	output-1.png	ab2e94e5f0b34f4096566b5d87545932	Static	125	Clear
9	output-1.png	ab2e94e5f0b34f4096566b5d87545932	Dynamic	192	Clear
10	output-1.png	ab2e94e5f0b34f4096566b5d87545932	Integrated	317	Clear
11	2.png	f83a72a9c02aa3d7394f44786fe242cd	Static	100	Clear
12	2.png	f83a72a9c02aa3d7394f44786fe242cd	Dynamic	189	Clear
13	2.png	f83a72a9c02aa3d7394f44786fe242cd	Integrated	259	Clear
14	1.png	e0b035324f70883b82e4a74c78757f72	Static	68	Clear
15	1.png	e0b035324f70883b82e4a74c78757f72	Dynamic	189	Clear
16	1.png	e0b035324f70883b82e4a74c78757f72	Integrated	257	Clear
17	asset.png	03e0392de533914753133df93e4ee9c3	Static	72	Clear
18	asset.png	03e0392de533914753133df93e4ee9c3	Dynamic	187	Clear
19	asset.png	03e0392de533914753133df93e4ee9c3	Integrated	259	Clear

Figure 7. Performance time result

5.2 Appropriate Incident Report

To generate hybrid malware detection report, create another new tab in terminal on cuckoo web runserver 192.168.47.128:8000

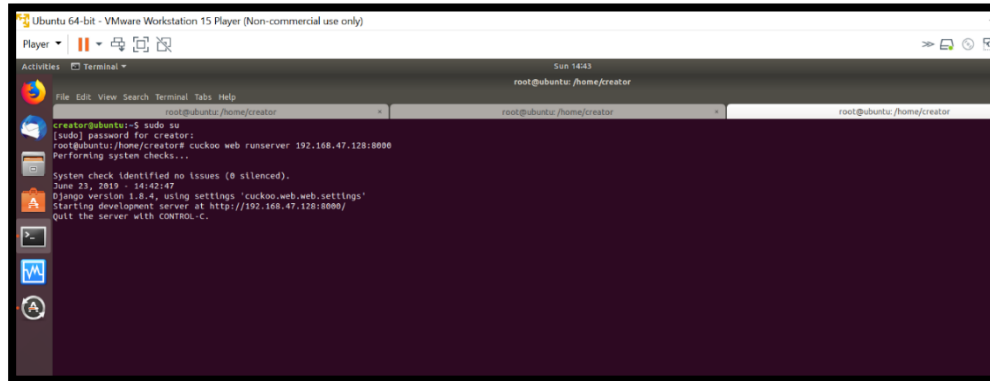


Figure 8. Incident Report

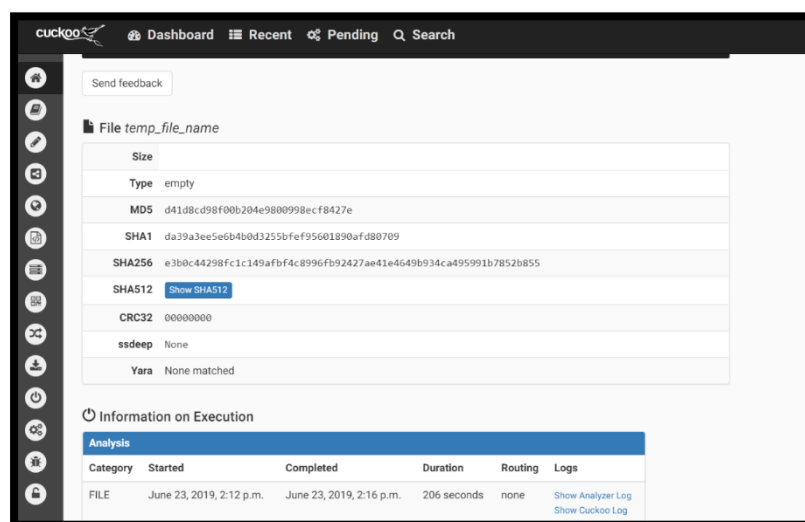


Figure 9. Reporting in Cuckoo sandbox

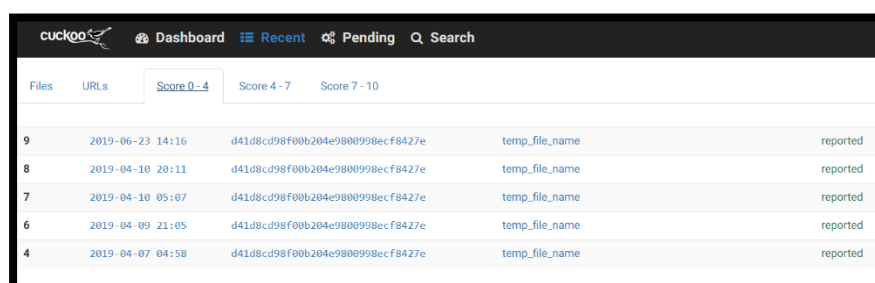


Figure 10. Rank approach

5.3 Result of accuracy performance for each method

The analyser system is configured with VMware inside which already install with windows XP operating system as host machine. The binary files are executed in binary files and the result will be gain from Cuckoo sandbox. In previous research (P. V. Shijoa, 2015), they use the random forest to detect the accuracy of malware detection but for this research we use the Gradient Boosting method to detect the accuracy performance for each method. Table 4 will show the classification and the summary that gain from this project.

Table 2. Classification and summary result for each method

Characteristics	(P. V. Shijoa, 2015)			This proposed project		
Machine learning	Random Forest			Gradient Boosting		
Result of Time taken	Not provided			Included with time taken for each method (Enhancement)		
Accuracy percentage (%)	Static	Dynamic	Hybrid	Static	Dynamic	Hybrid
	94.84%	96.65%	97.68%	94.87%	97.12%	98.26%

6. Conclusion

As technology nowadays keeps evolved and becomes more sophisticated, security technology also needs to grow together to ensure the goal of confidential, integrity and availability of system and data keep safe. Besides, malware also keep attacker never give up to improve their skills to exploit vulnerabilities. Based on that, the precaution like malware detection analysis will be one of the step to alert user to strengthen and improve their security.

Acknowledgments

Special thanks from authors for financial support from Universiti Putra Malaysia

References

- [1] S YusirwanS, Y Prayudi, I Riadi (2015) Implementation of Malware Analysis using Static and Dynamic Analysis Method. *International Journal of Computer Applications* (0975 – 8887) Volume **117** – No. 6, May 2015
- [2] P.V.Shijo, A.Salim,et al. (2015) Integrated static and dynamic analysis for malware detection, pp. 804-811. <https://doi.org/10.1016/j.procs.2015.02.149>
- [3] B.Tewfik,B. Zakaria A. Nemrat, B. Chafika,et al (2016) Scalable malware classification based on integrated static and dynamic feature pp. 113-124. DOI: 10.1007/978-3-319-51064-4_10
- [4] Mohammad, A. Mamoun, A.Izzat,Z.Mohammad,et al. (2016) The Malware Detection Challenge of Accuracy DOI: 10.1109/OSSCOM.2016.7863676
- [5] Y. Zhenlong, L. Yongqiang, A.Izzat,Z.Mohammad, et al (2016) Droiddetector: Android malware characterization and detection using deep learning. pp. 114-123. <https://doi.org/10.1109/TST.2016.7399288>
- [6] N. Awang, D. Yusof, S. Ariffin, et al (2015) Manual Malware Analysis Using Static Method pp. 324-328.
- [7] Difference Between Static Malware Analysis and Dynamic Malware Analysis | Difference Between | Static Malware Analysis vs Dynamic Malware Analysis. (2018).
- [8] A. Moser, C. Kruegel, and E. Kirda. Limits of static analysis for malware detection, in Computer Security Applications Conference, 2007. ACSAC 2007, pp. 421-430, Dec 2007.

- [9] I. You and K. Yim. Malware obfuscation techniques: A brief survey. In International Conference on Broadband, Wireless Computing, Communication and Applications (BWCCA), pp. 297-300, Nov 2010.
- [10] Christodorescu, M., Jha, S., Maughan, D., Song, D., Wang, C.: Malware Detection. In: Advances in Information Security. Verlag New York, Inc. Secaucus, NJ, USA: Springer (2007).
- [11] Weka 3: Data Mining open source Software. Accessed 2014. www.cs.waikato.ac.nz/ml/weka/.
- [12] V.Deepti, S.P.Choudary., et al, (2015) A Simple Method for Detection of Metamorphic Malware using Dynamic Analysis and Text Mining pp.265-270. DOI: 10.1016/j.procs.2015.06.031
- [13] Cao, Ying., et al, (2013) A Malware Behavior Capturing System Implemented at Virtual Machine Monitor Layer DOI: 10.1109/CIS.2012.126
- [14] The Cuckoo sandbox. Accessed 2014. <http://www.cuckoosandbox.org/>