BKH1925010F

by Iftekhar Efat

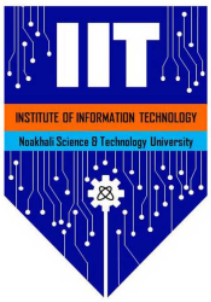
Submission date: 13-Mar-2022 03:07AM (UTC-0400)
Submission ID: 1782995741
File name: BKH1925010F.pdf (284.61K)
Word count: 1836
Character count: 9583

Machine Learning used in Computer viruses

Sultana Marjan
BKH1925010F

March 13, 2022

Report submitted for **SE2206: Information Security** under BSc. in Software Engineering Program, Institute of Information Technology (IIT), Noakhali Science and Technology University



Project Area: **Information Security .....**
Project Supervisor: **MD. IFTEKHARUL ALAM EFAT**
Assistant Professor
Institute of Information Technology (IIT)
Noakhali Science and Technology University

In submitting this work I am indicating that I have read the University's Academic Integrity Policy. I declare that all material in this assessment is my own work except where there is clear acknowledgement and reference to the work of others.

I give permission for this work to be reproduced and submitted to other academic staff for educational purposes.

OPTIONAL: I give permission this work to be reproduced and provided to future students as an exemplar report.

Abstract

Computer viruses are basically programmed in such a way that they destroy other files and other documents on our device.It is a malicious piece of computer code that is designed to spread from device to device.On the other hand ,machine learning (ML) is a set of program which gives computers the capability to learn without being explicitly programmed.The combination of different features of a malicious file that could be checked quickly by creating a reliable fingerprint by using machine learning . This paper proposes a new and more polished malware detector that can detect potential viruses.

1 Introduction

1.1 Sub Intro

Three types of malware detection procedures are Signature Based Malware Detection ,Behavioral Based Malware detection and Heuristic Based Malware Detection. The most representative sequence of bytes of a malware as a feature and uses a pattern matching method to detect malware are selected by the signature Based Detection and it is the most common method .Behavioral Based Malware Detection techniques observe the behavior of a program to detect whether it is malicious or not .Heuristic analysis is a technique for detecting viruses that involves looking for suspicious aspects in code.Signature detection is a method of detecting malware that involves comparing code in a program to the code of known virus kinds that have already been encountered, studied, and documented in a database.Signature detection methods, while effective and still in use, have become increasingly constrained as a result of the creation of new threats that emerged around the turn of the century and continue to appear all the time.To address this issue, the heuristic model was created with the goal of detecting suspicious traits in unknown, new viruses, modified versions of current threats, and known malware samples.[1]Cybercriminals are continually generating new dangers, and heuristic analysis is one of the few approaches for dealing with the massive volume of new threats that are noticed on a regular basis.

2 Background

2.1 Applying Machine Learning

Algorithms to Android Malware Detection: For billions of people, mobile devices are increasingly becoming their major mode of communication and information access.As a result, they've been the target of a slew of assailants.To acquire access to these devices, you must first gain access to them. In 2014, 16 million mobile phones were sold.Malware had infected the gadgets. These illnesses have the potential to spread.Identity theft, extortion, and robbery are all possibilities. As a result, the security of mobile devices is more vital than ever.Due to a lack of processing resources and a variety of execution methods,mobile security presents unique issues that are not present on PCs. Computer technology has recently progressed.Improvements in science and hardware have made it possible for machine learning to be used efficiently to retrieve important information data gleaned from big databases. Types of Android Malware: Malware comes in a variety of forms. Malware can be injected into legitimate apps and operate in the background while the program is doing its intended function. The user may never be aware that malware is present . Malware is a type of computer virus that can infect can also be downloaded in the middle of a game. This is referred to as a "update." attack". The user may be encouraged to download at other times.Malware that was advertised for a different purpose. This is referred to as a "Drive-by download" is a term used to describe a method of downloading files.

Malware is written to take advantage of a range of flaws and serve a variety of purposes. Malware can be used by attackers to gather information such as contacts and passwords. They can also make use of to take control of the device using malware Taking control of the situation is a huge step forward.They can then use premium-rate numbers to send text messages to.Rob the phone or join it to a botnet that can be managed.

Machine Learning Strategies: There is no "silver bullet" machine learning algorithm that works for every situation, as there is with anything in Computer Science. Training time and accuracy are two significant criteria to consider in general. For the study described in a basic neural network using gradient descent is used in this paper.It makes use of the optimizer function.

3 Methods

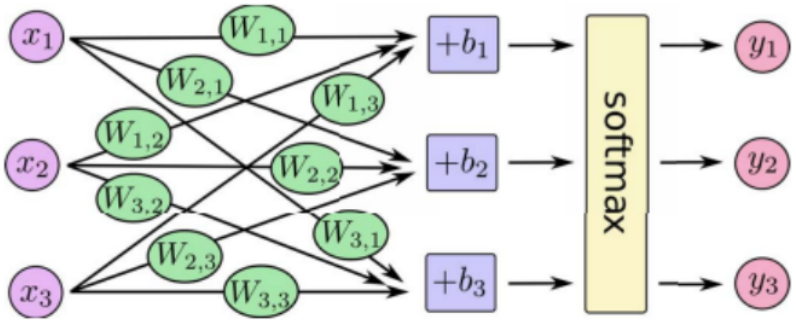
Collecting malware samples: In general, there are two ways to collect malware samples for research: "directly" using a honeypot or purchasing them from black hat hackers, or "directly" using a honeypot or acquiring them from black hat hackers.One such repository for researchers is Andrototal.After obtaining the information, the following command was used to download it:examples from 2016's first three months.[2]

Figure 1: Python code for malware detection

```
# the first chunk of each will be used for testing (and the rest for training)
for i in range(int(sys.argv[1])):
    j = random.randrange(1, len(malicious_app_name_chunks))
    k = random.randrange(1, len(benign_app_name_chunks))
    app_names_chunk = malicious_app_name_chunks[j] + benign_app_name_chunks[k]
    batch_xs = [dataset['apps'][app]['vector'] for app in app_names_chunk]
    batch_ys = [dataset['apps'][app]['malicious'] for app in app_names_chunk]
    sess.run(train_step, feed_dict={x: batch_xs, y_: batch_ys})
```

To demonstrate that the model generated using the TensorFlow approach described above is effective, it must be tested on a set of samples that are distinct from the ones that were used to create it,were put to use in the training. If the information in the database is incorrect,weights are valuable for things other than training data since they generalize.Even never-before-seen malware, such as 0-days, can be classified.Malicious and benign apps were mixed in equal amounts.The model's guess classification was compared to the actual classification.The right response, resulting in a value between 0 and one that reflects the percentage of samples properly categorised ,The Python code that was used to evaluate the problem is shown in Figure the efficiency of the model The section below titled "Results" will give you the information you need.The values, as well as an explanation of the outcomes.

Figure 2: A set of weights and biases

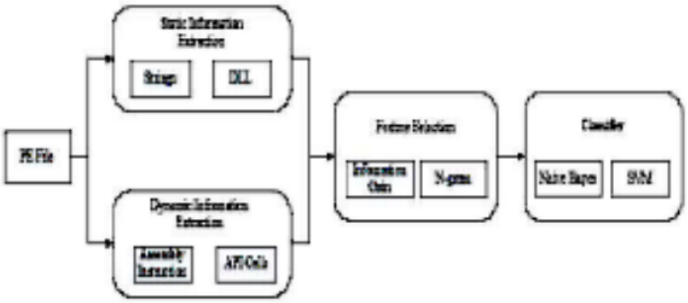


followed by a softmax function, can be used to classify a set of input values, as shown in this diagram. Each weight symbolizes a neuron, and each y value represents a category (in this case).

In our scenario, y1=1 denotes malevolent behavior while y2=1 denotes benign behavior).

Machine learning approaches for online malware detection: The number of services and features offered by various cloud service providers (CSP) has recently increased dramatically. Using such services has opened up several opportunities for businesses to migrate their infrastructure to the cloud, allowing them to more easily and flexibly provide services to their consumers.Infrastructure as a Service is the practice of renting out access to servers to clients for compute and storage reasons (IaaS). The rise in popularity of IaaS has sparked major and pressing concerns about cyber security and privacy. Malicious actors frequently use malware against cloud services in order to corrupt sensitive data or hinder their performance.As a result of this expanding threat, malware detection for cloud settings has become a hotly debated topic, with a variety of approaches being developed and implemented. We present online malware detection based on process level performance metrics in this paper, and compare the effectiveness of various baseline machine learning models such as Support Vector Classifier (SVC), Random Forest Classifier (RFC), KNearest Neighbor (KNN), Gradient Boosted Classifier (GBC), Gaussian Naive Bayes (GNB), and Convolutional Neural Networks (CNN). According to our findings, neural network models are best equipped to identify malware because they can accurately detect the influence malware has on the process level properties of virtual machines in the cloud.[3]

Figure 3: figure for malware detection



4 Results

It's a good idea to run the same data through the model several times to strengthen the weights that arise. The categorization accuracy for the results is shown in Figure 5 below.The number of training steps employed varies. The training procedures were as follows:The number of steps has been raised to 80. As can be observed, the outcomes have reached a halt.After n = n = n = n = n = n = This signifies that the maximum value should be used for this data.Nothing further can be done now that the learning threshold has been reached,can be gleaned from the information. Because the samples were divided,less than 20 training steps would not be sufficient to divide the population into 20 subsets.The model will be able to learn from all of the data given.

5 Conclusion

While the research presented in this study is limited to permission requests made by Android apps, future work will expand the amount of features utilized to train the dataset.In a virtual machine, a dynamic analysis of the app's activity could be useful.Add even more information (system calls, for example).Because computational resources aren't an issue, a larger system can be used.More training steps and a larger dataset could also assist to improve the results.. We would, however, like to be able to construct an app.It might be downloaded and run in real time on an Android device inspections. As a result, the algorithm would have to be modified.Finally, advanced machine learning algorithms may be implemented.

References

[1] Zahra Bazrafshan, Hashem Hashemi, Seyed Mehdi Hazrati Fard, and Ali Hamzeh. A survey on heuristic malware detection techniques. In *The 5th Conference on Information and Knowledge Technology*, pages 113–120. IEEE, 2013.

[2] Matthew Leeds and Travis Atkison. Preliminary results of applying machine learning algorithms to android malware detection. In *2016 International Conference on Computational Science and Computational Intelligence (CSCI)*, pages 1070–1073. IEEE, 2016.

[3] Jingling Zhao, Suoxing Zhang, Bohan Liu, and Baojiang Cui. Malware detection using machine learning based on the combination of dynamic and static features. In *2018 27th International Conference on Computer Communication and Networks (ICCCN)*, pages 1–6. IEEE, 2018.

BKH1925010F

ORIGINALITY REPORT

38%	20%	24%	18%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Matthew Leeds, Travis Atkison. "Preliminary Results of Applying Machine Learning Algorithms to Android Malware Detection", 2016 International Conference on Computational Science and Computational Intelligence (CSCI), 2016 Publication	13%
2	Submitted to University of Adelaide Student Paper	5%
3	Submitted to Southern Illinois University Student Paper	3%
4	Submitted to Shinas College of Technology Student Paper	3%
5	usa.kaspersky.com Internet Source	3%
6	publish.tntech.edu Internet Source	3%
7	global.oup.com Internet Source	2%

8	Zahra Bazrafshan, Hashem Hashemi, Seyed Mehdi Hazrati Fard, Ali Hamzeh. "A survey on heuristic malware detection techniques", The 5th Conference on Information and Knowledge Technology, 2013 Publication	1 %
9	Jingling Zhao, Suoxing Zhang, Bohan Liu, Baojiang Cui. "Malware Detection Using Machine Learning Based on the Combination of Dynamic and Static Features", 2018 27th International Conference on Computer Communication and Networks (ICCCN), 2018 Publication	1 %
10	mylibrary.sutd.edu.sg Internet Source	1 %
11	11811.co Internet Source	1 %
12	Submitted to University of Hertfordshire Student Paper	1 %
13	Jeffrey C Kimmell, Mahmoud Abdelsalam, Maanak Gupta. "Analyzing Machine Learning Approaches for Online Malware Detection in Cloud", 2021 IEEE International Conference on Smart Computing (SMARTCOMP), 2021 Publication	1 %
14	www.coursehero.com Internet Source	1 %

15	ijesi.org Internet Source	1 %
16	link.springer.com Internet Source	<1 %

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		

BKH1925010F

PAGE 1

Article Error You may need to use an article before this word. Consider using the article **the**.

PAGE 2

Article Error You may need to use an article before this word.

PAGE 3

P/V You have used the passive voice in this sentence. Depending upon what you wish to emphasize in the sentence, you may want to revise it using the active voice.

Article Error You may need to use an article before this word.

PAGE 4

Confused You have used **a** in this sentence. You may need to use **an** instead.

PAGE 5

Proper Noun If this word is a proper noun, you need to capitalize it.

Article Error You may need to remove this article.

Run-on This sentence may be a run-on sentence. Proofread it to see if it contains too many independent clauses or contains independent clauses that have been combined without conjunctions or punctuation. Look at the "Writer's Handbook" for advice about correcting run-on sentences.

Article Error You may need to remove this article.

Frag. This sentence may be a fragment or may have incorrect punctuation. Proofread the sentence to be sure that it has correct punctuation and that it has an independent clause with a complete subject and predicate.

PAGE 6

Article Error You may need to remove this article.

Confused You have used **to** in this sentence. You may need to use **two** instead.

Article Error You may need to use an article before this word. Consider using the article **the**.

P/V You have used the passive voice in this sentence. Depending upon what you wish to emphasize in the sentence, you may want to revise it using the active voice.

Frag. This sentence may be a fragment or may have incorrect punctuation. Proofread the sentence to be sure that it has correct punctuation and that it has an independent clause with a complete subject and predicate.

Article Error You may need to remove this article.