Towards the Development of a KNN-Based Intrusion Detection System for Mitigating Ransomware Attacks

+Oluwatobi, Adeyemo & *Olumide B. Longe (PhD)
+*Department of Computer Science
Caleb University
Lagos, Nigeria
*Systems & Multidisciplinary Research Group
Iinternational Centre for Information technology & Development
Southern University, Baton Rouge, LA, USA
E-mails: Mimiyoung2004@gmail.com; longeolumide@fulbrightmail.org
Phone: +2348064457448; +2348160900893

ABSTRACT

Ransomware infection is a challenging threat that encrypts a user's files while withholding the decipher key until a ransom is paid by the victim. This type of malware is a lucrative business for cybercriminals, generating millions of dollars annually. The spread of ransomware is increasing as traditional detection-based protection, such as anti-virus and anti-malware, has proven ineffective at preventing attacks. Encrypting ransomware targets small and large businesses as well as the regular home user. This thesis discusses ransomware evolution, using machine learning based algorithm to extract some signature features from a ransomware and use the extracted features to detect when next a new ransomware is at the verge of infecting the system.

Keywords: Sound, multi vibrator, microphone, relay, switch.

iSTEAMS Cross-Border Conference Proceedings Paper Citation Format
Oluwatobi, Adeyemo & *Olumide B. Longe (2018): A Clap Away from Light – Towards the Development of a KNN-Based Intrusion Detection System for Mitigating Ransomware Attacks. Proceedings of the 13th iSTEAMS Multidisciplinary Conference, University of Ghana, Legon, Accra, Ghana. Pp 211-216

1. INTRODUCTION

Ransomware is a type of malware that uses malicious codes to intrude the system before the end-users notices the intrusion, to encrypt some important files. It is the fastest growing threat that encrypts user's files and locks the computer and holds the key required to decrypt the files for ransom. Over the past few years, Impact of ransomware has increased exponentially.

There have been several reported high profile ransomware attacks, such as cryptolocker, cryptowall, wannacry, petya and bad rabbit which have collectively cost individuals and companies well over a billion dollars. In this project, a ransomware detection technique on windows operating system is proposed.

Statement of the Problem

If we look at the strategies that have been formed against ransomware today, they are summarized into four categories: prevention, detection, disruption and remediation.
In this thesis, the main focus is aimed at improving the detection strategy of ransomware using supervised learning approach of the machine learning method.

Significance/Justification of Study

This thesis will try to propose a reasonable and dependable solution to help detect ransomware in a system before it infects the system

Scope of The Thesis
This thesis will be focused on developing a desktop application that can detect the presence of ransomware before it infects the system. Ransomware infestation is a growing problem for end-users and because of its non-traceability it is very difficult to detect it before it takes over the system and locks the user out until the ransom is paid.

Aims & Objectives
The Aim Of This Research Is To Develop A Knn-Based Intrusion Detection System For Mitigating Ransomware Attacks.

To achieve the aim, the following objectives will be pursued:

i. Understudy the evolution of ransomware
ii. Extract similar features of ransomware through behavioural property extraction using the KNN machine learning algorithms
iii. Develop a software system that can detect ransomware before it locks computer systems
iv. Create a desktop application for ransomware detection and measure its performance

LITERATURE REVIEW

Through the literature analysis and analyzing the detection methods of current anti-ransomware product, various methods of detection, mitigation and indemnification were identified. This chapter presents other works and their findings divided into each of the methods.

Method of Detection	Publisher/year	Author	Title	Overview	Limitation
Monitoring File System And Performance	The University Of Texas At Alington /2017	Ashwini Balkrushna Kardile	Crypto Ransomware Analysis And Detection Using Process Monitor	Various Techniques Are Developed And Used To Monitor Malicious Sample File System Activity And I/O Access In Malware Analysis Environments.	As A Result, There Is A Risk That Ransomware May Run At The Kernel Level And Counter Some Hooks Which May Be Used By Process Monitor To Capture Registry Access Activities
	Hindawi Publishing Corporation /2015	Sanggeun Song, Bongjoon Kim, And Sangjun Lee	The Effective Ransomware Prevention Technique Using Process Monitoring On Android Platform	Processing Module Determines The Handling Of The Process Suspected As Ransomware By The Monitoring Module And Makes An Exception Or Isolation Of The Process.	This Process Is Only Suitable For Android Based Systems
	International Journal Of Engineering Development And Research/2017	S.Mahmudha Fasheem, 2p.Kanimozhi, B.Akoramurthy	Detection And Avoidance Of Ransomware	In Proposed System Two Things Are Absorbed One Is To Prevent Ransomware And Second Is The Detection Of The Ransomware And To Recover The Affected File Or Unblock The Access Control.	This Method Detects Ransomware After It Has Infected The System And The Files Infected With The Ransomware Will Be Removed.

	Ieee/2015	D. Bekerman, B. Shapira, L. Rokach, A. Bar And B. Sheva	Malware Detection Using Network Traffic Classification	This Method Analyzes Dns, Http, And Ssl Protocols, And Combines Different Network Classification Methods In Different Resolutions Of Network.	Future Work We Intend To Extend The Research On Transfer Learning Techniques To Improve Detection From Untrained Network
Honeypot Technique	Ieee/2016	Nolen Scaife, Henry Carter, Patrick Traynor,Kevin R.B.Butler	Detecting Ransomware With Honeypot Techniques	Using Ctb-Locker And Cryptolock They Can Detect Ransomware Based On Behavior Against User Data	When High Latency Creates In Operation That Time File Is Often Locked And Cannot Be Opened By Crypto Drop.
	Vectra Network/2016	Unknown	Surviving The Ransomware Pandemic	Another Low-Cost Technique To Foil Ransomware Attackers Involves Setting Up Canary Or Honeypot File Shares.	It Takes Time Also Depletes The Memory Space Allocation.
	Springer/2010	Mirosław Skrzewski	Monitoring Malware Activity On The Lan Network	Lan Network Traffic Monitoring Was Based On The Packet And They Assigned Ip Address.	Assigned Ip Addresses Do Not Generate Much Address Resolution Protocol(Arp) Traffic.
Machine Learning	Unknown /2015	Diane Duros Hosfelt	Automated Detection And Classification Of Cryptographic Algorithms In Binary Programs Through Machine Learning	This Detection Method Can Be Used To Detect When Crypto Ransomware Attacks The System And Starts Encrypting files.	It Is Only Able To Detect Programs Written In C And C++
	Xamk/2017	Kateryna Chumachenko	Machine Learning Methods For Malware Detection And Classification	The Goal Of The Project Lies In The Determination Of The Most Suitable Feature Representation And Extraction Methods, The Most Accurate Algorithm That Can Distinguish The Malware Families With The Lowest Error Rate	Currently, The Feature Extraction Is Performed After The Files Were Run In The Sandbox And The Reports Were Generated. This Approach Will Result In Delays In The File Analysis When Implemented.
Static-Based Analysis	Mwr Labs Whitepaper / 2018	Daniel Nieuwenhuizen	A Behavioural-Based Approach To Ransomware Detection	The Most Common Type Of Static Analysis, Which Is Commonly Used In Commercial Virus Scanners, Is Referred To As Signature Analysis.	The Fundamental Flaw Of Signature-Based Detection Is Its Inability To Detect Unknown Malware Which Has Yet To Be Turned Into A Signature.
Energy Consumption Monitoring	J Ambient Intell Human Computer	Azmoodeh, A, Dehghantanha, A, Conti, M And Raymond Choo, Kk	Detecting Crypto-ransomware In Iot Networks Based On Energy Consumption Footprint	To Develop A Fingerprint Of Ransomware's Energy Consumption, Initially, We Need To Record The Power Usage Of Targeted Applications.	It Involves A Lot Of Work To Be Monitoring The Power Consumption

Evolution-Based Classification	Elsevier/2 016	P. Zavarsky And D. Lindskog	Experimental Analysis Of Ransomware On Windows And Android Platforms	The Ransomware Families Is Analyzed On Their Evolution And Characterization.	To Prevent The Users' Data From Getting Into Unrecoverable State, A User Should Have Incremental Online And Offline Backups Of All The Important Data.

RESEARCH GAPS

After analyzing the various works and research on the diverse ways ransomware can be detected or mitigated I noticed some laxities which I would like to address in This thesis. They are:

i. Some Detection methods are limited to only one program type
ii. Some are not real-time
iii. Some methods are slow and manual.

METHODOLOGY

The first objective will be satisfied by an extensive literature research into understanding the evolution of ransomware. This will help provide a base line understanding into the functional properties, characteristics and features of ransomware.

The second objective will be satisfied by introducing various ransomware samples, that will be collected and used to train the system we are developing i.e the features of the training set will be extracted. All ransomwares have some similar basic features because they belong to the same family of malware. This trained dataset will be the database. A total of 100 Ransomware samples will be used to serve as our training and testing set.

The third objective will be achieved by using R and Python language to write the source code needed to build a detection software(back-end) which will be infused with the previously extracted features. The developed software is now subjected to an accuracy examination to see its efficacy in identifying the testing ransomware set as well other ransomware free files. The model will be able to detect the ransomware in the testing set.

The fourth objectives will be satisfied by using visual basic studios to build a front-end application that can interface between the back-end software and the end user.

WHY MACHINE LEARNING?

The rapid development of data mining techniques and methods resulted in Machine Learning forming a separate field of Computer Science. It can be viewed as a subclass of the Artificial Intelligence field, where the main idea is the ability of a system (computer program, algorithm, etc.) to learn from its own actions. It was firstly referred to as "field of study that gives computers the ability to learn without being explicitly programmed" by Arthur Samuel in 1959. A more formal definition is given by T. Mitchell: "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E." (Mitchell 1997).

The basic idea of any machine learning task is to train the model, based on some algorithm, to perform a certain task: classification, clusterization, regression, etc. Training is done based on the input dataset, and the model that is built is subsequently used to make predictions. The output of such model depends on the initial task and the implementation.

Machine learning approach

i. **Unsupervised learning**: This deals with learning useful structures without labeled classes, optimization criterion, feedback signal or any other information beyond the raw data input.
   Various methods are:
   a. Clustering(k-means)
   b. Taxonomy creation(hierarchical clustering)
   c. Trend detection(extrapolation)

ii. **Supervised learning**: this deals with the learning of useful structures with the help of a trained dataset gotten from the the raw data input. The data are labeled with pre-defined classes.
   Considering the nature of this work we will be using the supervised learning approach of machine learning.
   Various methods are:
   a. Regression
   b. Classification

We will be making use of the Classification method of the supervised learning approach.

C**lassification Methods**

From machine learning perspective, malware detection can be seen as a problem of classification or clusterization: unknown malware types should be clusterized into several clusters, based on certain properties, identified by the algorithm.

Various classification methods are:

a. Bayes tree/Bayesian network
b. Artificial Neural network
c. Random forest
d. Support vector machines (SVM)
e. K-nearest neighbor(K-NN)

We will be using the K-NN(K-nearest neighbor) algorithmic system for this thesis because it is one of the simplest, though, accurate machine learning algorithms. KNN is a non-parametric algorithm, meaning that it does not make any assumptions about the data structure. In real world problems, data rarely obeys the general theoretical assumptions, making non-parametric algorithms a good solution for such problems. KNN model representation is as simple as the dataset – there is no learning required, the entire training set is stored.

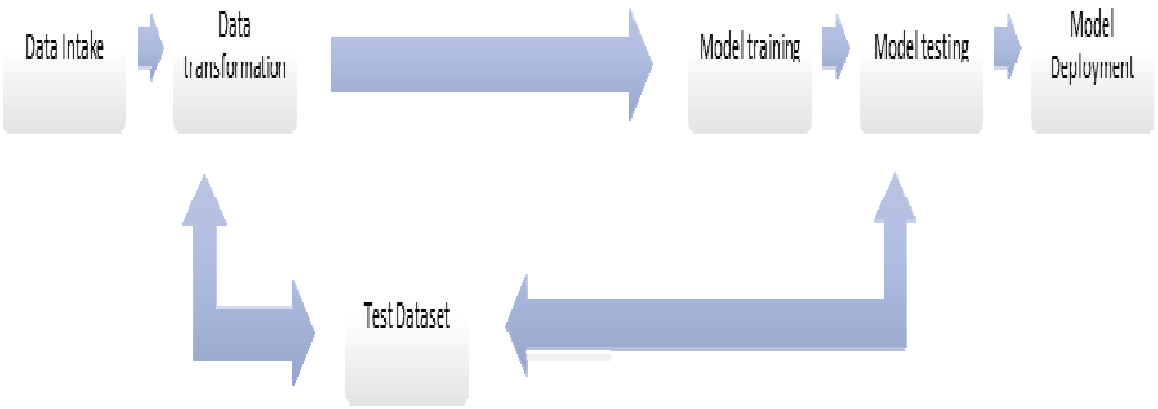

Fig 1: Workflow process
Source: Fieldwork

Feature Extraction

Such attributes are referred to as features, and the matrix is referred to as feature vector. The process of extracting data from the files is called feature extraction. The goal of feature extraction is to obtain a set of informative and non-redundant data. It is essential to understand that features should represent the important and relevant information about our dataset since without it we cannot make an accurate prediction.

WORKS REVIEWED/CONSULTED

1. ASHWINI, B.,(2017). *Crypto Ransomware Analysis And Detection Using Process Monitor.* The University of Texas.
2. Sanggeun, S., Bongjoon, K., (2015). *The Effective Ransomware Prevention Technique Using Process Monitoring On Android Platform. Hindawi Publishing Corporation.*
3. Fauzia, I., Muttu, K.,(2014). *Investigating the android Intens and permissions for malware detetection*," Xamk.
4. D. Bekerman, B. Shapira, L. Rokach, A. Bar and B. Sheva(2016), *Unknown Malware Detection Using Network Traffic Classification*," pp. 134–142.
5. Nolen S., Carter, H., Traynor, P., Kevin R.,(2016)"CryptoLock (and Drop It): Stopping Ransomware Attacks on User Data, "pp. 303
6. Nolen Scaife, et al., (2016). *Detecting Ransomware With Honeypot Techniques.* Institute of Electrical and Electronics Engineers.
7. Mirosław S.,(2010). *Monitoring Malware Activity on the LAN Network*. Springer.
8. Diane D. (2015). *Automated detection and classification of cryptographic algorithms in binary programs through machine learning.* In: arXiv: 1503.01186. url: http://arxiv.org/ abs/1503.01186.
9. P. Zavarsky and D. Lindskog,*(2016) Experimental analysis of ransomware on windows and android platforms : evolution and characterization.* vol. 94, pp. 465–472.
10. Nieuwenhuizen, D. (2018) *A behavioural-based approach to ransomware detection.* Mwr Labs Whitepaper.