Ransomware, Threat and Detection Techniques: A Review

SH Kok[†], Azween Abdullah[†], NZ Jhanjhi[†] and Mahadevan Supramaniam[††]

[†]School of Computer and Information Technology, Taylor's University, Malaysia
[††]Research & Innovation Management Centre, SEGI University, Malaysia

Summary
The popularity of ransomware has created a unique ecosystem of cybercriminals. Therefore, the objectives of this paper are to provide a thorough understanding of ransomware's threat and discuss recent detection techniques used. Successful ransomware attack has direct financial implication, which is fuelled by several mature enablers, such as encryption technology, cyber currency and accessibility. Encryption is effective and almost unbreakable. Anonymous cyber currency can avoid traceability. Easily obtainable ransomware code enables easy entry. A combination of these provides an attractive avenue for cybercriminals, producing specialist cybercriminals. In terms of detection techniques, it was found that machine learning (ML) via regression algorithms was the most technique adopted by researchers of ransomware. However, none of the researchers have produced any model to protect against ransomware attack. This research highlights the need of a solution using ML algorithm for the detection engine.

Key words:
Ransomware, Intrusion Detection (ID), Machine Learning (ML), Honeypot

1. Introduction

Ransomware has attracted great attention from cyber security experts in recent years because of the fast growth of its attacks and the creation of new variants capable of bypassing antiviruses and anti-malwares [1]. It is a relatively new malware but has generated much interest from cybercriminals because of its successful attack and direct financial interest. Ransomware objective is to block its victim from accessing their own resources by locking the OS or encrypting targeted files that seem valuable to the victim, such as images, spreadsheets and presentations. [2]. Basically, there are two types of ransomwares: locky and crypto. Locky ransomware locks the entire system from access by its user, but it is usually easy to resolve. However, crypto ransomware uses encryption technology to lock selected files from user access; this is much more difficult to resolve and the damage caused may be irreversible. Crypto ransomware is also the more popular type employed by cybercriminals. A third type of ransomware called scareware has been mentioned in the literature [3]. This ransomware does not actually damage the victim's computer but only scares the victim into paying the ransom. This type of ransomware is not discussed in this paper.

2. Methodology of Literature Review

In this work, we performed our research based on specific criteria related to ransomware. First, our research is based on recent studies, from year 2015 onwards. Thus, the information is the latest and is still relevant. Second, our sources comprise only scientific journals and conference papers; this is to ensure that the collected information is authentic. Third, we only focus on our topic of interest: the threats caused by ransomware and techniques for ransomware detection.

2.1 Contribution of Paper

This paper provides a detailed ransomware attack lifecycle and its characteristics, which can serve as a groundwork for future research on ransomware. In addition, existing techniques for ransomware detection, as well as the pros and cons of each technique, are reviewed. Based on our findings, we then propose a solution in terms of a model, which will be presented in our next paper.

3. Ransomware

Ransomware is a type of malware that prevents its victims from accessing their own data until they pay a ransom. This type of malware has direct financial implication, which has promoted an ecosystem of cybercriminals, who employ it as a business model. Ransomware as a service (RaaS) is a service that allows the easy acquisition of ransomware codes at a price. The price could be an outright purchase, or a profit sharing scheme could be used. This indicates that cooperation exists among criminals. One party is responsible for developing and creating the ransomware code, while another party is responsible for organizing the dissemination of the infection or an attack campaign, and both parties enjoy the profit from a successful attack. Ultimately, this will promote specialist criminals that authorities will find difficult to tackle [3].

Manuscript received February 5, 2019
Manuscript revised February 20, 2019

3.1 Enablers of Ransomware Attack

Ransomware attacks have expanded both in frequency and variant because of the facilitative actions of several enablers. These enablers arose mainly due to technology advancement and lifestyle change.

(i) Encryption Technology

Encryption is used for privacy purposes. In today's heavy dependence on the internet, large amounts of data are transmitted electronically. However, these data can easily be intercepted. Therefore, to ensure that the data are only read by the designated persons, encryption technology was invented.

This technology has proven to be a double-edged sword. Ransomware has exploited this technology to encrypt victim's files for extortion purposes. Ransomware mainly uses three types of encryption technology: symmetrical encryption, asymmetrical encryption and hybrid encryption.

Symmetrical encryption uses one key for both encryption and decryption process. Its advantage is the encryption process can be quickly completed. However, its downside is that it is less secure.

Asymmetrical encryption uses one key for encryption, called public key, and another key for decryption, called private key. The encryption process is slower but more secure.

Hybrid encryption combines both symmetrical encryption and asymmetrical encryption. Initially, the victim data are encrypted using symmetrical encryption, and then the key is encrypted using asymmetrical encryption. This enables a quick encryption process and high security.

(ii) Cyber Currency

Cyber currency is the main payment method for the ransom. This is mainly because such currency allows the recipient to remain anonymous to the authority. A cyber currency such as Bitcoin has received wide acceptance. This is true, especially with popularity of online stores that accept cyber currency.

Block chain technology is another form of encryption technology that uses a one-way hash function. This is the key technology employed by the cyber currency payment method to ensure legitimacy of currency [4].

(iii) Ransomware Accessibility

Ransomware codes can be easily obtained with the existence of RaaS. In addition, free development kits, such as Torlocker, TOX and Hidden Tear, are available for unskilled individuals. This greatly reduces the entry barrier of ransomware [3].

3.2 Ransomware Lifecycle

There are mainly seven steps in the lifecycle of ransomware, as shown in Fig. 1. The lifecycle shows the formation of a cybercriminal ecosystem, in which there is a close cooperation between 'creator' and 'campaigner'. The creator is the programmer that develops and produces the ransomware code, while the campaigner is the organizer of the attacking campaign. This kind of cooperation allows both parties to improve and sharpen their knowledge and skills in their areas of focus with each cycle. Ultimately, this produces specialist criminals.

(i) Creation

Creating ransomwares with programming codes is the main task of the creator. The creation stage also involves enhancing codes to increase the potency of the ransomware at the end of each cycle. Lessons learned from a cycle can be used for improvement in the next cycle.

(ii) Campaign

Distributing or disseminating the ransomware to the victim's system is the main task of the campaigner. Basically there are two types of target victims: individual and institutional victims.

For individual victims, the dissemination objective is usually to reach as many victims as possible. Dissemination to individual victims may be simpler, because not many victims are computer-savvy.

However, for institutional victims, the ransomware needs to be specifically targeted and highly sophisticated. This is because, usually, some form of security defence is already in place.

Some common infection vectors are email attachments, email links, compromised websites and social media. The campaign success depends on how effective the human psychology of fear and insatiability is exploited.

(iii) Infection

When the payload has reached the victim's system, the ransomware setup behaviour may begin. However, more sophisticated ransomwares may employ certain precautionary steps, which are detailed in Section 4.

(iv) Command and Control

Once the setup is complete, the ransomware may contact Command and Control centre for several reasons, the most common being to obtain the encryption key for the encryption process. Another possible reason is to download more files for more advanced infections.

(v) Search

After obtaining the encryption key, the ransomware can start searching for seeming valuable files such as text documents, spreadsheets, presentations, images and databases.

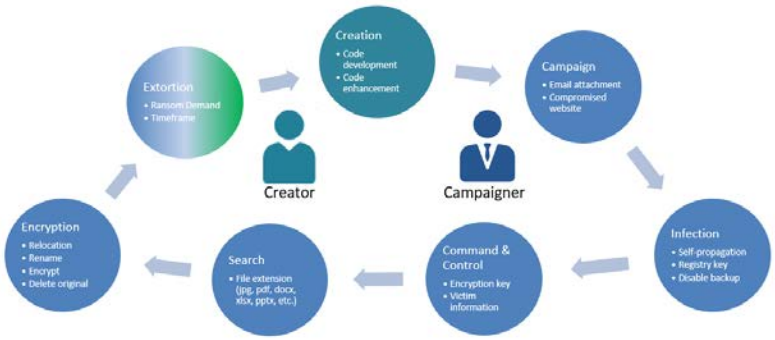

Fig. 1 Ransomware lifecycle

(vi) Encryption

Once all valuable file types have been identified, the ransomware starts the encryption process. Ransomwares normally use three types of encryption technology: symmetrical encryption, asymmetrical encryption and hybrid encryption. These are discussed in detail in the previous section.
Depending on the step(s) involved, the encryption process can be divided into three classes: class A, class B and class C, which are discussed in the next section.

(vii) Extortion

After the above steps have been completed, the final step in the cycle is to display the ransom demand. The demand note will inform the victim of the infection and specify the mode of payment. In addition, the note may also contain the ransom payment deadline, after which the ransomware will start deleting the valuable files.

3.3 Types of Ransomware Attacks

There are mainly two types of ransomware attacks: locky ransomware, which locks the system from being logged in, and crypto ransomware, which encrypts specific file types, making them inaccessible to the victim.

(i) Locky

Locky ransomware locks the system from being logged in by its victim. However, the system can be usually restored by rebooting or running in safe mode. Therefore, this type of ransomware is less detrimental and can be resolved quite easily.

(ii) Crypto

Crypto ransomware encrypts specific file types that are considered valuable to the victim such as documents, spreadsheets, pictures and databases. It can employ symmetrical, asymmetrical or hybrid encryption.
Depending on the step(s) involved, the encryption process can be classified into three: Class A, the file is encrypted but not renamed or relocated; class B, the file is encrypt and renamed, but not relocated; and class C, the file is encrypted, renamed and relocated, increasing the difficulty of tracking and restoring the file.

4. Ransomware Setup Behaviour

After the ransomware has been successfully uploaded into the victim's system, the setup step, as shown in Fig. 2, is crucial in ensuring a complete and successful infection. Ransomwares may employ one or more of the below precautionary actions [5].

Fig. 2 Ransomware setup behaviour

4.1 Payload Persistence

This action is to ensure that the attack can be persistent even after the system is rebooted. Common techniques used are placing an executable file in the startup directory, adding a new registry key and setting a scheduled task.

4.2 Restrict System Restore

This action is to prevent the victim from restoring the system to the pre-infection state. Common techniques used are to delete scheduled backup, backup system and backup files.

4.3 Stealth Mode

This action is to prevent the attack from being visible to the victim. Common techniques used are to execute from %AppData% directory as well as using the same name as the common system executable.

4.4 Environment Mapping

This action is to ensure that the infection is actually in the victim's system and not in a sandbox. A sandbox is the common setup for dynamic analysis of malwares. Common techniques used are to check the security setting and policies, geographical location, user language, file system architecture and network drives.

4.5 Communication Masking

This action is to ensure successful communication with Command and Control centre. A domain name can be randomly generated using an algorithm; this will complicate the tracking performed by the authority.

4.6 Privilege Elevation

This action is to enable the attacker perform actions as an administrator. Many system-related actions can only be performed by the administrator; therefore, elevating to administrator level will ensure all actions can be performed without restriction.

5. Types of Ransomware Analysis

The objective of ransomware analysis is to better understand how ransomware functions. Based on this understanding, defensive steps can be formulated to prevent future infections. Two types of analysis can be performed: static analysis and dynamic analysis, as shown in Fig. 3. Static analysis is based on the source code of the executable file. For dynamic analysis, the ransomware is executed in a controlled environment, and all its actions are recorded for analysis [6].

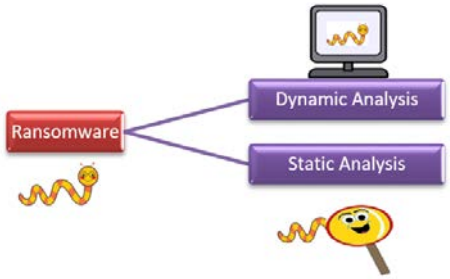

Fig. 3 Ransomware analysis

5.1 Static Analysis

Static analysis can be conducted quickly by examining the features of an executable piece of code and matching it to a previously observed malicious code.

(i) Pros

The malicious code is easily and quickly analysed. Successful detection here also means that the ransomware can be avoided without it having any chance to be executed.

(ii) Cons

It is susceptible to code obfuscation. Simple addition of normal operation codes can result in mismatch with previously identified malicious codes.
In addition, the analysis is also not effective when the code is encrypted. There is currently no efficient way to decrypt an encryption using brute force. It is simply time-consuming.
Static analysis is also not effective towards multi-phase attacks. The initial code could merely be a simple process to open a backdoor for additional codes to be downloaded and thus may not have a similar malicious action.

5.2 Dynamic Analysis

Dynamic analysis is also called behavioural-based analysis. Malicious code is executed in a controlled and monitored environment, usually a sandbox. All actions are captured for analysis.

(i) Pros

This type of analysis is less prone to obfuscation, and encrypted code can be analysed. Malicious action must be part of the process in order to achieve its objective. Encrypted code must be decrypted before the malware can perform its action.

(ii) Cons

Setup for this type of analysis is both costly and time-consuming. To accurately capture the ransomware behaviour, it is important that the environment setup closely imitates an actual environment.
As discussed previously, one of the setup behaviours of ransomware is environment mapping. If the analysis is performed using a virtual machine, which can cut cost and resources, the ransomware may discover this and prevent itself from exhibiting all its behaviours.

6. Ransomware Detection Techniques

This section discusses the various detection techniques used to discover and identify ransomware. The papers reviewed are summarized in Table 1, while the general techniques are discussed below.

6.1 Machine Learning

Machine learning (ML) involves learning the patterns in data to create a model. This model can then predict the outcome when fed with new data [7]. However, the difficulty in using ML is in finding the correct algorithm to match with the type of data and the needed outcome.

(i) Pros

The advantage of ML is that it can accurately predict the outcome with adequate training data. Training data should be varied with balanced distribution of outcomes to be predicted. Because ML involves learning the pattern in the data, it is less prone to obfuscation.

(ii) Cons

Finding the correct algorithm is often not straightforward and may require some runs of trial and error. Moreover, biasness and overfitting may occur if adequate caution is not taken.

6.2 Honeypot

Honeypot involves setting up decoy files for the ransomware to attack. Once these files are accessed, the ransomware can be identified.

(i) Pros

The traps or honeypot files can be set up, and then they simply wait to be attacked. Therefore, the technique does not require much maintenance or processing power from the system.

(ii) Cons

There is no guarantee that the honeypot files will be attacked by a ransomware. Therefore, it is important to know the characteristics of files that the ransomware will attack.

6.3 Statistic

Statistic can be used to analyse ransomwares to better understand their important characteristics. However, deploying this technique as a detection mechanism is tasking.

Table 1: Recent ransomware research from 2015 to 2018

Ref.	Purpose/ Motivation	Methodology	Result	Limitation/ Future Direction
[8]	Early detection can still be effective after the victim is infected.	Honeyfiles in Linux	Ransomware is immediately blocked, and user is notified for its removal	Not tested on other platforms such as Windows and Android Combine honeypot with tracking mechanism
[9]	Intrusion detection (ID) prevention system and antivirus as a single monitoring agent is complex and time-consuming and thus fails in ransomware detection.	Honeyfolder, a decoy folder modelled using social leopard algorithm (SoLA)	Better accuracy and precision, and recall software-defined networking improved network protection with simple rules.	Not tested on healthcare implants and other internet-connected gadget
[10]	Identify salient features of ransomware	Statistical comparison of API call between normal operation and ransomware	Eight APIs existed only in ransomware. Four APIs were found in ransomware to a statistically significant degree. Six API frequencies were more than three standard deviations away from the mean.	Cannot actually be used to detect ransomware.
[11]	Ransomware that can fingerprint environments can evade dynamic analysis.	Five ML algorithms were used for binary classification of ransomware using static analysis of opcodes transformed into n-gram using eight families of ransomware.	Both random forest and K-nearest neighbour produced the highest recall value of 99.8%.	Cannot adequately distinguish between CryptoWall, Locky and Reveton ransomwares according to accuracy metric for binary classification

[12]	Provide insights on how ransomware has evolved from its outset to March 2016	Analysis of ransomware families: 17 for Windows and 8 for Android.	In Windows, detection was possible by monitoring abnormal file system and registry activities. Android could be reduced with permissions request.	Other platforms such as Linux and Mac platform
[13]	Ransomware attack at fog layer nodes in cloud system has no direct contact with end users.	Three neural network deep learning algorithms were used for binary and multi-class classification of ransomware.	Long short term memory algorithm produced the best result for binary (F-measure of 0.996) and multi-class classification (true positive rate of 0.972 and false positive rate of 0.027).	Other deep learning algorithms such as sequential discriminative training of deep neural network and ensemble deep neural network.
[14]	Dynamic analysis can be used to detect ransomware, because it exhibits a set of characteristics at runtime that are common across families, which helps early detection of new variants.	EldeRan, mutual information (feature selection) and regularized logistic regression (classification)	EldeRan performed better than support vector machine (SVM) and naïve Bayes, but VirusTotal had the best performance.	Limitation: Ransomware that waits for user to perform a task. No other applications were running except those packed with fresh installation of Windows. Dataset consisted of a set of API that was not empty.
[3]	Explore research endeavours in ransomware Highlight issues and potential research directions.	Literature review	Ransomware taxonomy	Ransomware prediction Engineering new ransomware features Efficient detection technique Toolkit-based analysis and detection Ransomware classification Tracking ransomware payment Ransomware dataset Data recovery Ransomware advanced persistent threat attack detection
[15]	IoT involves resource-constrained devices.	Literature review	Taxonomy of IoT Security research	Application programs security Secure data perception Data transmission security Physical protection and availability
[2]	Evaluate the evolution of ransomware and its behaviour	Literature review	Ways to detect presence of ransomware Ways to prevent ransomware	Only two papers are specific to ransomware.
[1]	Ransomware malware is capable of avoiding antivirus and anti-malware applications; thus, this study investigates its prevention, detection and recovery.	Sixty-two papers were selected with specific criteria of filtering, and then cross referenced with 15 classes.	Most studies concentrate on behavioural-based detection. Detection in mobile devices has received much attention.	More research on detection and recovery of ransomware attack is needed.
[16]	Behavioural data are more difficult to obfuscate and take relatively longer time to capture, typically 5 min, meaning that malicious payload has likely been delivered by the time it is detected.	5s snapshot of behavioural data using recurrent neural network. Sample from VirusTotal executed from 0 to 20 s on Win 7.	Accuracy (96.01), time (19s), false positive (3.17), false negative (4.72)	Other platforms Snapshot and monitor per-process basis Automatic block further processing of detected malware Portability to other machine
[17]	Dynamic analysis monitors the behaviours of malwares, making them more difficult to be concealed, and majority relies on system calls. However, monitoring of system calls can be evaded; thus a new approach is required.	Bernoulli mixture model of probabilistic generative clustering to group unknown malware binaries through dynamic analysis of interaction with system resources in sandbox	Malware only: homogeneity (0.767), completeness (0.609), v-measure (0.679), avg # clusters (492); mixture of malware and legitimate operation: homogeneity (0.761), completeness (0.523), v-measure (0.620), avg # clusters (499)	Does not work for samples that do not interact with resources monitored by the sandbox
[18]	Dynamic malware analysis is time-consuming, especially for sophisticated timing-based evasion malwares.	Virtual time controller (VTC) sandbox, which is a Xen hypervisor modified with virtual clock to speed up system timer.	Ether ran for 120 s, precision (98.2); VTCSandbox@1x ran for 120 s, precision (98.2); VTCSandbox@8x ran for 102 s, precision (98.6).	May not work for a malware that is not time-based.
[19]	Classification of malware based on n-gram of API call sequence and an online update of regularization weight for each iteration.	Regularization weight (C) and n-gram (n). 5 ML, passive-aggressive I (PA I), passive-aggressive II (PA II), confidence weighted (CW) learning, adaptive regularization of weight, normal herd	Most accurate was CW with C = 4.0 and n = 6, train (0.940), test (0.918)	Continuous online update of classifier could be compromised by malware that slowly injects new accepted API.

[20]	ML classifier can be sabotaged by polluting training data by malware for mobile.	Adversaries detection (AD), camouflage detector using similarity analysis of negative outcome with very benign outcome and very malicious outcome to detect false negative.	Sophisticated attacker SVM w/o AD (63.30), w/ AD (89.30); random forest w/o AD (67.85), w/ AD (88.85); K-nearest neighbour w/o AD (72.00), w/ AD (90.45); KuafuDet (96.20)	Use reinforcement technology to prevent APK from reverse-engineering Use game theory to interpret attacker psychology
[21]	Low-profile attacker distributes concealed packets over a long period of time to mislead firewall and network ID system. Large number of features with relatively small number of samples leads to overfitting.	Feature selection using recursive feature addition (RFA) and bigram technique. Bigram technique to encode payload strings. SVM for classification. Combined metric that combines accuracy, detection rate and false alarm rate. ISCX 2012 dataset	Five hundred data, accuracy without bigram (82%), accuracy w/ RFA (92.9%), combined metric (87.8%)	GPU with many cores can parallel train multiple SVM at the same time Use RFA on web server to select most important features Use ensemble classifier. Use trigrams technique. Conduct test on Android botnet dataset.
[22]	Advanced persistent threats can obfuscate identifiable feature through encryption, custom code bases and in-memory execution. Machine activity metrics such as CPU, RAM, Swap and network traffic are inescapable footprints.	Self-organizing feature map (SOFM) creates unsupervised cluster of similar behaviour that can be used as features for classification. Creation of two maps: Good map from benign, Bad map from malicious	Malware operational plot review (MOPR) is end-to-end workflow from sample, sandbox, SOFM and logistic regression. % Accuracy: MOPR (93.76), random forest (86.52), BayesNet (77.70), multi-layer perception (MLP) (79.40), SVM (68.08)	Increase sample size. Increase granularity of data. Model phases of system behaviour to detect attack at early stage.
[23]	Dynamic analysis that model interaction with system resources and error messages	Vocabulary-based with multi-instance learning. Instance is set of pairs of names and types of resources. Vocabulary is clustering method dependent on type of resources and warning. classifier random forest, Linear SVM and MLP.	Accuracy random forest (0.943), SVM (0.944), MLP (0.938)	May not work with a malware that does not trigger any error messages.
[24]	Most other studies only use raw value of Portable Executable (PE) header field with other complimentary features such as Dynamic Link Library (DLL) and API.	Integrated feature (IntF) set using derived value of raw value to provide more meaningful feature	Logistic regression, raw (77.06), IntF (78.12); linear discriminant analysis, raw (91.71), IntF (92.45); random forest, raw (97.43), IntF (98.78); decision tree, raw (96.47), IntF (97.12); naïve Bayes, raw (56.04), IntF (50.09) K-nearest neighbour, raw (94.73), IntF (90.79)	Study the impact of application type such as multimedia, document processing and device drivers. Investigate the performance of a combination of the proposed IntF with another feature set.
[25]	Ransomware can be avoided through prevention techniques.	Literature review	Cyber security techniques involve four phases: predict, prevent, respond and detect.	Nil
[26]	Intensive review of ransomware and previous research	Literature review	Four types of ransomwares: misleading application, rogue antivirus, locker and crypto List of existing defence techniques employed	Nil
[27]	New patterns that can bypass existing vaccine systems are constantly created; therefore, a behavioural-based monitoring of processor, memory and storage is required.	Creation of priority protection area. Any new behaviour will be treated as suspicious, and user will be prompted for input. If the user approves as non-ransomware, it will be recorded in the database and automatically be allowed in the future. If user specifies it as ransomware its removal will be requested. In the future, user permission for removal will be triggered if similar behaviour is found.	Able to detect an author-created ransomware that can bypass two antiviruses	Nil

[28]	Many generic malware detection systems have been proposed, but none attempt to specifically address ransomware.	Automatic generation of artificial user environment to attract attack, then monitor filesystem I/O and detect screen locker	Able to detect known and unknown types of ransomware	Attacker may be able to detect artificial environment or run at a kernel level that can thwart monitoring of UNVEIL.
[29]	Prevent ransomware attack using detection and removal techniques	Literature review	Detection includes using honeypot, heldroid (Android) creating safe zone and v-detector. Prevention includes monitoring filesystem activities and test packet	Nil
[30]	New ransomware variants are able to avoid dynamic ML techniques.	NetConverse uses six ML algorithms, and then chooses the most effective.	J48 decision tree produces the best true positive rate of 97.1%	Develop real-time ransomware detection using cloud-based ML classifiers.
[31]	Attackers continue to create new variants of ransomware to bypass existing mitigation approach.	RanDroid extracts image and text to determine existing locking screen or threatening note. Tested on 100 ransomwares and 200 Goodware applications	91% accuracy	Obtain more data to improve accuracy, include detection of law enforcement logo, use natural language processing and spelling auto-correction.
[32]	Commercial solutions cannot detect latest variant of ransomware.	Talos uses concurrency workbench of new century, which compares app with temporal logic rules of ransomware	99% accuracy	Use Talos to detect other malware such as botnet and rootkit, also possible application in IoT.
[33]	Signature-based detection is insufficient and generic malware attack vectors are not robust enough to detect all ransomwares	Trap layer for early detection, ML layer to detect zero-day ransomware and file backup layer to preserve user data	Best result was from gradient tree boosting, with 98.25% true positive rate and 0.56% false positive rate	Evaluate in large-scale real setup.
[34]	Ransomware issue has increased in scale and affects cyber security.	Antibiotics enforce file access-control policy to prevent unauthenticated file modification	Total protection against contemporary ransomware	Allow granularity configuration for specific types of files and permit for specific processes.

Table 2: Analysis of ransomware detection techniques introduced from 2015 to 2018

Reference	Network	PC				Mobile			Data Source				Machine Learning										Outcome			Evaluation							
		Windows	Mac	Linux	Raspberry	Android	iOS	Cloud	Dynamic Analysis	Static Analysis	Honeypot	Statistic	Bayesian	Decision Tree	Dimension Reduction	Instance Based	Clustering	Deep Learning	Ensemble	Neural Network	Regularization	Rule System	Regression	Binary Classification	Multi-class Classification	Clustering	Accuracy	True Positive Rate	False Positive Rate	True Negative Rate	False Negative Rate	Precision	F-measure	Time
[8]				x							x													x									
[9]					x						x													x			x	x				x		
[10]		x							x			x																					
[11]		x								x			x	x		x								x			x	x				x	x	
[12]		x				x			x	x																							
[13]								x	x									x						x	x			x	x				x	
[14]		x							x				x		x						x		x					x	x					
[3]																																		
[15]																																		
[2]																																		
[1]																																		
[16]		x								x									x	x			x	x			x		x		x			x
[17]		x							x								x									x								
[18]		x							x															x				x				x		x
[19]		x							x												x			x			x			x		x		
[20]						x				x					x	x			x				x	x			x							
[21]	x														x								x	x			x					x		

[22]		x							x				x						x	x			x	x				x				x	x	
[23]		x							x										x	x			x	x			x	x	x					
[24]		x								x			x	x	x	x			x				x	x			x	x				x	x	
[27]						x			x			x													x									
[28]		x							x		x														x				x				x		
[30]	x	x											x	x		x			x	x					x				x	x			x	x	x
[31]						x			x	x															x			x							
[32]						x				x															x				x				x	x	x
[33]		x							x	x									x	x			x	x				x	x						
C	**2**	**14**	**0**	**1**	**1**	**5**	**0**	**1**	**13**	**8**	**3**	**2**	**5**	**3**	**4**	**4**	**1**	**1**	**7**	**5**	**2**	**0**	**8**	**18**	**1**	**1**	**9**	**12**	**6**	**1**	**1**	**9**	**7**	**4**

Note: C for count x represents item used by researcher

7. Result and Discussion

Table 1 shows recent studies performed on ransomware. The table summarizes the method used, result and limitation of the research, and it includes review and survey papers. Based on this list, we can conclude that many researchers are focusing on using ML techniques to detect ransomware. This is not surprising, because through ML, a model that recognizes ransomware's behavioural pattern, rather than a specific signature, can be built. This is important, because ransomware is constantly evolving and its signature can easily change; however, its attacking pattern is more difficult to change.

Table 2 shows a detailed breakdown of applied ransomware detection techniques; the table excludes review and survey papers. Based on this table, we can conclude that most studies were performed on Windows platform, because this is the most popular for personal computer.

In addition, researchers preferred to use dynamic analysis than static analysis. This is because behavioural data is more difficult to obfuscate, compared to source code.

As noted in Table 1, many researchers focused on ML techniques to detect ransomware. These techniques can be further sub-categorized based on the adopted algorithm, as suggested by [35]. The table shows that regression-type algorithms were the most preferred. This type of algorithm uses the relationship between predictors and outcome to build a model. This is an effective approach that can produce accurate binary classification.

Two important evaluation metrics used are accuracy and true positive rate. Both metrics indicate the extent to which the algorithm can provide accurate prediction.

8. Our Proposed Technique

To effectively detect ransomwares, a thorough understanding of ransomware infection and attacking pattern is required. Therefore, for our solution, we propose to create a ransomware detection model. To achieve this, we need to obtain data from both static and dynamic analyses. The pros and cons of both types of analysis are discussed in Section 5. We propose to utilize data from both types of analyses to ensure the most effective detection. However, for dynamic analysis, the ransomware samples need to be detected within the first 30 seconds of their execution, as suggested by [14]. This is to ensure detection with minimum damage, as damage incurred is irreversible.

In terms of the detection technique, we propose an ML technique with a possible hybrid of regression and rule-based algorithms. A regression algorithm builds its prediction model based on the relationship between predictors and outcome, while a rule-based algorithm generates set of rules for its prediction model. This combination can create a robust model that follows a set of rules from rule-based algorithm, and these rules are further strengthened with relationship formulation from the regression algorithm.

9. Conclusion

Our solution is twofold. First, we propose to build a ransomware attack model using data from both dynamic and static analyses, as this will ensure ransomware infection and attacking pattern are deeply understood. To the best of our knowledge, none of the previous researchers have adopted this approach.

In addition, we suggested using a hybrid algorithm instead of a single algorithm. This is because with the correct combination, the constituent algorithms of a hybrid algorithm can complement each other, and thus, the hybrid algorithm will outperform the individual algorithms.

References

[1] E. P. Torres P. and S. G. Yoo, "Detecting and neutralizing encrypting Ransomware attacks by using machine-learning techniques: A literature review," Int. J. Appl. Eng. Res., vol. 12, no. 18, pp. 7902–7911, 2017.

[2] J. P. Tailor and A. D. Patel, "A Comprehensive Survey: Ransomware Attacks Prevention, Monitoring and Damage Control," Int. J. Res. Sci. Innov., vol. IV, no. November, pp. 2321–2705, 2017.

[3] B. A. S. Al-rimy, M. A. Maarof, and S. Z. M. Shaid, "Ransomware threat success factors, taxonomy, and

countermeasures: A survey and research directions," Comput. Secur., vol. 74, pp. 144–166, 2018.
[4] M. Paquet-Clouston, B. Haslhofer, and B. Dupont, "Ransomware Payments in the Bitcoin Ecosystem," 2018.
[5] D. Nieuwenhuizen, "A behavioural-based approach to ransomware detection," 2017.
[6] D. Distler, "Malware Analysis : An Introduction." SANS Institute, USA.
[7] S. Kok, A. Abdullah, M. Supramaniam, T. R. Pillai, and I. A. T. Hashem, "A Comparison of Various Machine Learning Algorithms in a Distributed Denial of Service Intrusion," Int. J. Eng. Res. Technol., vol. 12, no. 1, pp. 1–7, 2019.
[8] J. A. Gómez-Hernández, L. Álvarez-González, and P. García-Teodoro, "R-Locker: Thwarting ransomware action through a honeyfile-based approach," Comput. Secur., vol. 73, pp. 389–398, 2018.
[9] C. D. D. Biomedico and C. Alberto, "SoLA: Social Leopard Algorithm for Intrusion Detection Honeypot to detect ransomware attacks," IEEE Trans. Cogn. Devlopment Syst., no. Submitted, pp. 16–23, 2018.
[10] N. Hampton, Z. Baig, and S. Zeadally, "Ransomware behavioural analysis on windows platforms," J. Inf. Secur. Appl., vol. 40, pp. 44–51, 2018.
[11] H. Zhang, X. Xiao, F. Mercaldo, S. Ni, F. Martinelli, and A. K. Sangaiah, "Classification of ransomware families with machine learning based on N-gram of opcodes," Futur. Gener. Comput. Syst., vol. 90, pp. 211–221, 2019.
[12] Monika, P. Zavarsky, and D. Lindskog, "Experimental Analysis of Ransomware on Windows and Android Platforms: Evolution and Characterization," Procedia Comput. Sci., vol. 94, pp. 465–472, 2016.
[13] S. Homayoun et al., "DRTHIS: Deep ransomware threat hunting and intelligence system at the fog layer," Futur. Gener. Comput. Syst., vol. 90, pp. 94–104, 2019.
[14] D. Sgandurra, L. Muñoz-González, R. Mohsen, and E. C. Lupu, "Automated Dynamic Analysis of Ransomware: Benefits, Limitations and use for Detection," 2016.
[15] I. Yaqoob et al., "The rise of ransomware and emerging security challenges in the Internet of Things," Comput. Networks, vol. 129, pp. 444–458, 2017.
[16] M. Rhode, P. Burnap, and K. Jones, "Early-stage malware prediction using recurrent neural networks," Comput. Secur., vol. 77, no. December 2017, pp. 578–594, 2018.
[17] J. Stiborek, T. Pevný, and M. Rehák, "Probabilistic analysis of dynamic malware traces," Comput. Secur., vol. 74, pp. 221–239, 2018.
[18] C. H. Lin, H. K. Pao, and J. W. Liao, "Efficient dynamic malware analysis using virtual time control mechanics," Comput. Secur., vol. 73, pp. 359–373, 2018.
[19] A. Pektaş and T. Acarman, "Classification of malware families based on runtime behaviors," J. Inf. Secur. Appl., vol. 37, pp. 91–100, 2017.
[20] S. Chen et al., "Automated poisoning attacks and defenses in malware detection systems: An adversarial machine learning approach," Comput. Secur., vol. 73, pp. 326–344, 2018.
[21] T. Hamed, R. Dara, and S. C. Kremer, "Network intrusion detection system based on recursive feature addition and bigram technique," Comput. Secur., vol. 73, pp. 137–155, 2018.
[22] P. Burnap, R. French, F. Turner, and K. Jones, "Malware classification using self organising feature maps and machine activity data," Comput. Secur., vol. 73, pp. 399–410, 2018.
[23] J. Stiborek, T. Pevný, and M. Rehák, "Multiple instance learning for malware classification," Expert Syst. Appl., vol. 93, pp. 346–357, 2018.
[24] A. Kumar, K. S. Kuppusamy, and G. Aghila, "A learning model to detect maliciousness of portable executable using integrated feature set," J. King Saud Univ. - Comput. Inf. Sci., 2016.
[25] S. B. Surati and G. I. Prajapati, "A Review on Ransomware Detection & Prevention," vol. IV, no. Ix, pp. 86–91, 2017.
[26] H. Shakir and A. N. Jaber, "A Short Review for Ransomware : Pros and Cons A Short Review for Ransomware : Pros and Cons," no. August, 2018.
[27] S. Song, B. Kim, and S. Lee, "The Effective Ransomware Prevention Technique Using Process Monitoring on Android Platform," vol. 2016, 2016.
[28] A. Kharaz, S. Arshad, C. Mulliner, W. Robertson, C. Mulliner, and W. Robertson, "UNVEIL : A Large-Scale , Automated Approach to Detecting Ransomware.," in USENIX Security Symposium, 2016, pp. 757–772.
[29] M. Poriye and V. Kumar, "Ransomware : Detection And Prevention," Int. J. Comput. Sci. Eng., vol. 6, no. 5, pp. 900–905, 2018.
[30] O. M. K. Alhawi, J. Baldwin, and A. Dehghantanha, "Leveraging Machine Learning Techniques for Windows Ransomware Network Traffic Detection," Adv. Inf. Secur., vol. 70, pp. 1–11, 2018.
[31] A. Alzahrani et al., "RanDroid : Structural Similarity Approach for Detecting Ransomware Applications in Android Platform," pp. 892–897, 2018.
[32] A. Cimitile, F. Mercaldo, V. Nardone, A. Santone, and C. A. Visaggio, "Talos: no more ransomware victims with formal methods," Int. J. Inf. Secur., vol. 17, no. 6, pp. 719–738, 2018.
[33] S. K. Shaukat and V. J. Ribeiro, "RansomWall : A Layered Defense System against Cryptographic Ransomware Attacks using Machine Learning," IEEE, 2018.
[34] O. Ami, Y. Elovici, and D. Hendler, "Ransomware Prevention using Application Authentication-Based File Access Control," 2018.
[35] S. H. Kok, A. Abdullah, N. Z. Jhanjhi, and M. Supramaniam, "A Review of Intrusion Detection System using Machine Learning Approach," Int. J. Eng. Res. Technol., vol. 12, no. 1, pp. 9–16, 2019.

SH Kok received the M.S. degrees in Information Technology from University Putra Malaysia in 2001. Currently, he is pursuing his Ph.D. degree in Computer Science at Taylor's University.

Azween Abdullah is a professional development alumni of Stanford University and MIT and his work experience includes thirty years as an academic in institutions of higher learning and as director of research and academic affairs at two institutions of higher learning, vice-president for educational consultancy services, 15 years in commercial companies as Software Engineer, Systems Analyst and as a computer software developer and IT/MIS consultancy and training.

Noor Zaman has completed his PhD. in IT from University Technology Petronas Malaysia. He has 19 years of teaching and administrative experience internationally. He has an intensive background of academic quality accreditation in higher education besides scientific research activities, he had worked for academic accreditation for more than a decade and earned ABET accreditation twice for three programs at College of computer sciences and IT, King Faisal University Saudi Arabia.

Mahadevan Supramaniam serves as Director, Research and Innovation Management Centre & Institute of Graduate Studies of SEGi University. Mahadevan's expertise lies in R&D development and policies, Enterprise Resource Planning, Computer Science & security system, Business Process Management and Integrated Technologies for industries. He has shared most of his experiences on his expertise area through public talks and has written many papers and books which has been published all over the world. Mahadevan holds a DBA from the Twintech International University College of Technology Malaysia and a Master's of Software Engineering degree from University Malaya.