Journal of Theoretical and Applied Information Technology
15th December 2019. Vol.97. No 23

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

AN INTELLIGENT AND REAL-TIME RANSOMWARE DETECTION TOOL USING MACHINE LEARNING ALGORITHM

[1]HIBA ZUHAIR, [2]ALI SELAMAT

[1]Dr, Department of Systems Engineering, College of Information Engineering, Al-Nahrain University, Baghdad, Iraq

[2]Prof. Dr, Malaysia Japan International Institute of Technology (MJIIT), Universiti Teknologi Malaysia (UTM), Kuala Lumur, Malaysia

E-mail: [1]hiba.zuhir@coie-nahrain.edu.iq or hiba.zuhair.pcs2013@gmail.com , [2]aselamat@utm.my

ABSTRACT

Abstract Zero-day ransomware still threaten users' and enterprises' survival in the cyber-space by disturbing electronic amenities, damaging information systems, and causing data and money losses. The publically used anti-ransomware software are trying to mitigate this security issue, however they are limited at identifying zero-day ransomware variants effectively in the real-time without performance overhead. Thus, this paper proposed intelligent, real-time, and three-tier model of ransomware detection tool to be performed well for protecting windows-based information systems. The proposed ransomware detection tool comprises a hybrid machine learning algorithm which hybridizes the decisive functions of two topmost machine learning algorithms (Naïve Bays and Decision Tree) to holistically characterize and accurately classify zero-day ransomware variants in real-time application. Empirical, comparative and realistic assessments demonstrate the adaptability and effectiveness of the proposed ransomware detection tool versus zero-day ransomwares. It achieves approximate accuracy rate of (96. 27%) and mistake rate of (1.32%) along with low misclassifications throughout real-time practice.

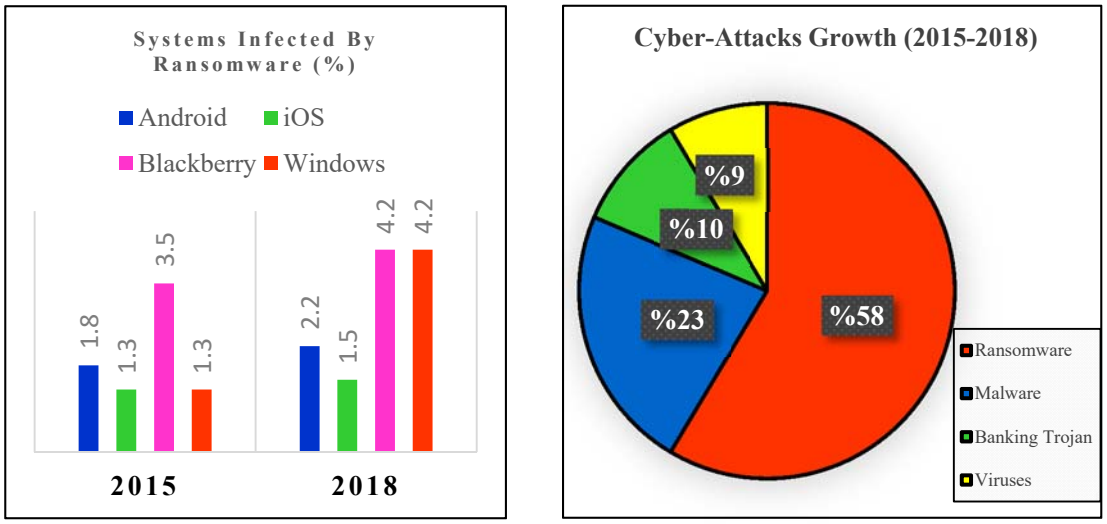
Keywords: *Zero-day ransomwares, Signature-based detection, Anomaly-based detection, Hybrid-based detection, Dynamic traits, Hybrid machine learning algorithms.*

1. INTRODUCTION

With the recent breakthrough of cyber-space services, almost world-wide enterprises involved in connected information systems that are integrated together at an exponential rate over the last few ears [1]. Incessantly, such information systems have been targeted by the cyber-criminals who evolved new families of their ransomware attacks that aiming at disrupting the electronic amenities, damaging the information systems, causing giant monetary losses to the related enterprises [1], [2]. Usually, ransomware attacks utilize systems' vulnerabilities and intrude them via add-ons, cookies, email attachments, faux and hoaxing links, and software downloads [1], [3]. They infect install ransomware file payload into the system, configure system registry and boot-up, seek out any backups and directories to remove them, disable windows backup, and launch their own servers during the internet connection. Then, they use their own keys to encrypt system files and pop-up messages on the screen asking the users to pay a ransom for file decryption and system unlock during a particular time [1]-[3]. As such cyber-criminals would gain giant and illegal profits, and they exploit ransomwares for terrorism activities in the existence of limited anti-ransomware techniques that affects cyber-security superior than other cyber-attacks. Furthermore, cyber-criminals target users of android, iOS, and Blackberry systems on various smart devices as shown in **Figure 1(a)** [1]-[3]. That, in turn, would cause wide spread, fast advancement, most targeting business and governmental industries, and great damages to both cyber-security and economy during 2015 and 2018 as shown in **Figure 1(b)** [1]-[3].

Journal of Theoretical and Applied Information Technology
15th December 2019. Vol.97. No 23

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195



(a) Outmost Infected Smart Devices And Computer Systems By Ransomware Attacks *(b) Supremacy Of Ransomware Attacks Among Cyber-Attacks*

Figure 1. Global Advancement Of Ransomware Attacks [1]-[3]

To protect users' privacy and survive systems' data in *windows* platforms, many anti-ransomware tools are developed in industry and academia. Among them are the signature-based anti-ransomware tools like *Kaspersky* and *Avast*. Others are the anomaly-based anti-ransomware tools like *McAfee* [4], [5]. However, these tools suffer from problems of frequent archive update, data loss and deterioration, obfuscation noise, false alarms, infirmity versus the scalable network traffic, and defeatism against zero-day ransomwares. Thus, they still need more improvements to work effectively against zero-day ransomwares and different ransomware families that have not been investigated before [4], [5].

To overcome the aforesaid problems, researchers develop hybrid-based anti-ransomware tools which are assisted by static and dynamic analyses of ransomware traits, and they are assembled by machine learning algorithms to detect zero-day ransomwares accurately with lower rate of false. For example, EldeRan, ShieldFS, and UNIVEIL etc. [4] - [20]. However, hybrid-based anti-ransomware tools are still obsolescent versus the adversary traits that the new ransomware usually evolve in real-time application. Deploying machine learning classifiers only while a ransomware's action runs encounters variable detection outcomes according to their ensemble design, different induction boundaries, and various sets of traits that they leverage.

To detect zero-day variants of ransomware families effectively in the real-time mode, the anti-ransomware tool requires an adaptive machine learning classifier that trains a big stream of data, adjusts its default functions, and regulates its induction parameters to make future predictions.

To do so, two supervised and unsupervised machine learning algorithms are hybridized to generate more decisive machine learning algorithm for ransomware detection, as it is presented in this paper. Furthermore, this hybrid machine learning algorithm is assisted by a set of the most distinctive ransomware traits for more holistic characterization of ransomware families. To be a well-suited ransomware detection tool for protecting windows-based information systems, the proposed hybrid machine learning algorithm is assembled into a three-tier ransomware detection model which involves analysis, learning, and testing tiers.

The analysis tier analyzes the dynamic actions and probable infections of different ransomware families in five-minute test routine which is implemented iteratively via a virtual testbed. Dynamic traits and infecting actions are characterized, synthesized, and archived for the next tier "learning tier". The learning tier; learns the extracted sets of traits by using the proposed hybrid machine learning algorithm to generate the ransomware detection model that must be adjusted whenever necessary during the real-time application mode. Whilst, the testing tier; applies the generated detection model to identify any unknown action of a

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

new ransomware variant "zero-day ransomware" in real-time application.

As such, the key contributions of this paper can be itemized as follows:

- A hybrid machine learning algorithm that analyzes the dynamic traits of zero-day ransomware variants,
- A robust schema of an anti-ransomware tool for protecting windows-based information system against zero-day ransomwares without tackling the entire system and machine.
- A Real-time model of anti-ransomware tool against daily evolving variants of different ransomware families.
- Realistic and comparative performance assessment of the proposed work.

For more description, the rest of this paper is organized as follows: a background of the ransomware and the related works will be presented in Section 2. Whereas, Section 3 will present the design and workflow of the proposed approach. Implementation and experimental results will be discussed in Section 4. At last, conclusions and future outlooks will be presented in Section 5.

2. RELATED WORK

Recently, various anti-ransomware tools have been developed by researchers in both academia and industry to detect ransomware families and their own variants online. These tools have been grouped under three categories: signature, anomaly, and hybrid approaches [4]-[20]. The signature-based approaches have identified ransomware variants whose traits have been previously archived in their related repositories [4], [5]. However, they have needed to update their archives frequently that might need human labor and took long time. Furthermore, some signature-based approaches like *BitDefender* have identified ransomware trait of particular ransomware families exclusively. Precisely, families of similar variants rather than families of different variants [4-20]. Whereas, anomaly-based approaches have recognized whether the system activities are benign or threat throughout data mining rules and/or machine learning classifiers [6]-[11]. However, they have encountered high fault alarms along with complex computations, time-consuming in implementation, and performance overhead on windows and other platforms [6]-[11]. To boost anti-ransomware, other achievements have adopted hybrid approaches for by assembling both signature and anomaly-based approaches simultaneously [12]-[20]. Even though, they have shown effective ransomware detection against zero-day variants; they have encountered performance overhead, maximal footprints, maximal misclassifications in online mode [12]-[20]. Inspiration of aforesaid categories have driven many related works published recently in the topic of anti-ransomware technology. **Table 1** synthesizes the related works in the terms of what baseline approaches they have adopted? , what ransomware families they have detected? , and what performance key factors they have leveraged but they still demanded?

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 1. *Synthesis of notable related works for anti-ransomware technology*

Baseline Approach	Brief Description	Key Factors	Drawbacks	Examples	Related Works
Signature	Relying on cryptographic primitive functions (static traits) and up-to-date databases to detect ransomwares	-Moderate accuracy -Lightweight	-Data loss -Obfuscation noise -Defeatist by zero-day attacks -False alarms -Frequently Updateable Database	BitDefender KasperSky McAfee Avast	[1]-[5]
Anomaly	Relying on runtime activities (dynamic traits) to detect ransomwares	-High accuracy -Lightweight	-Infirm vs. scalable network traffic -Defeatist by zero-day attacks -Impacts on files recovery -Infirm vs. arbitrarily large population of users	R-Locker RansomFlare	[6]-[11]
Hybrid	Relying on runtime activities (dynamic traits) and machine learning algorithms to detect ransomwares	-High accuracy -Low false alarms	-Evaded by adversary traits -Not real-time mode -Obsolescent against various ransomware families -Defeatist by zero-day ransomwares	EldeRan UNVEIL NetConverse ShieldFS 2entFOX Heldroid	[12]-[20]

3. METHODOLOGY

This section describes the proposed hybrid machine learning algorithms, the design of the anti-ransomware tiers, and the behavioral traits as well as the ransomware families that have to be detected by the proposed tool, as follows:

3.1 Ransomware Traits and Families

Cyber-criminals utilize different ransomware families to attack users and their computer-based systems as presented in **Table 2**. Accordingly, ransomware families varied in their activities, traits, and impacts. They exhibit traits that may infect the settings of operating systems, or users' data and files, or software applications, or control and command servers. To this end, 9 usually exploited traits by the topmost ransomware families are utilized to implement and testify the proposed tool. Both ransomware families and their corresponding traits are described in **Table 3.**

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 2. *The most risky ransomware families to windows-based information systems [1-5]*

Type	Place	Year	Method	Impacts
AiDS	USA	1989	It was delivered to computer-based information systems via floppy disks	Encrypting root directories Damage system files
GpCode	Russia	2005	It developed symmetric encryption algorithm to encrypt users' data files	Damage to management information systems of banks and real estate agencies
Archiveus	USA	2006	It applied RSA algorithm to encrypt system files	Defeat the original version of computer system Cause illegal and big money losses
WinLock	Russia	2008-2012	It locked computer system and demanding ransom via sending SMS to victim's phone number	Damage computer system Cause Data loss Cause Money loss
Reveton	Pakistan	2012	It impersonated law enforcement agencies to deceive users with rumor claims	Abuse the prepaid electronic payment platforms of e-business
Crypto-Locker	Europe	2013	It encrypts file's contents by RSA algorithm with private and public keys	Lock computer systems Target industrial organizations Cause financial loss
Crypto-Wall	USA	2014	It encrypted system files and injected malicious code which freezed the systems firewalls.	Halt the computer system Require ransoms in Bitcoins
Ransom as Service (RaaS)	USA and Europe	2015	Impersonated a malicious website in the dark web	Halt computer systems of victims who use are browsing the dark web

Table 3. *The Topmost ransomware families and their traits [11]-[20]*

Traits	Ransomware Families							
	CryptoWall	WinLock	Reventon	CryptoLocker	Archiveus,	GpCode	AiDS	RaaS
Windows API calls	✖	✓	✓	✖	✖	✖	✖	✖
Registry key actions	✓	✖	✓	✓	✖	✖	✖	✖
File system actions	✓	✖	✖	✓	✖	✖	✖	✓
Various File extensions	✓	✓	✓	✓	✓	✓	✖	✖
File names	✖	✖	✖	✖	✓	✖	✓	✖
Directory actions	✓	✓	✖	✓	✖	✖	✖	✖
Application folders	✓	✓	✖	✓	✖	✖	✖	✖
Control panel settings	✓	✖	✓	✓	✖	✖	✖	✖
Command and Control Server (C & C)	✓	✓	✓	✓	✖	✖	✖	✓

3.2 The Proposed Hybrid Machine Learning Algorithm

It can be observed from **Figure 2** that the proposed hybrid machine-based learning algorithm assembles the decision functions of Decision Tree (DT) and Naïve Bays (NB) synchronously to complement their pruning margins for more accurate categorization. Throughout tree building and pruning, DT generated its predictions of the traits within a tree structure: nodes, leafs, and branches. As such, DT's nodes denoted all traits in the input vector, the leaves referred to the predictions of the corresponding traits, and DT's branches related each

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

examined trait to its corresponding category [9]. Whilst, NB predicts the actual category of the overlooked traits in the indecipherable nodes of DT. So far, any unclassified trait is categorized by using the Bayes' probabilistic theorem with the assumption of that all traits are independent of each other [9].

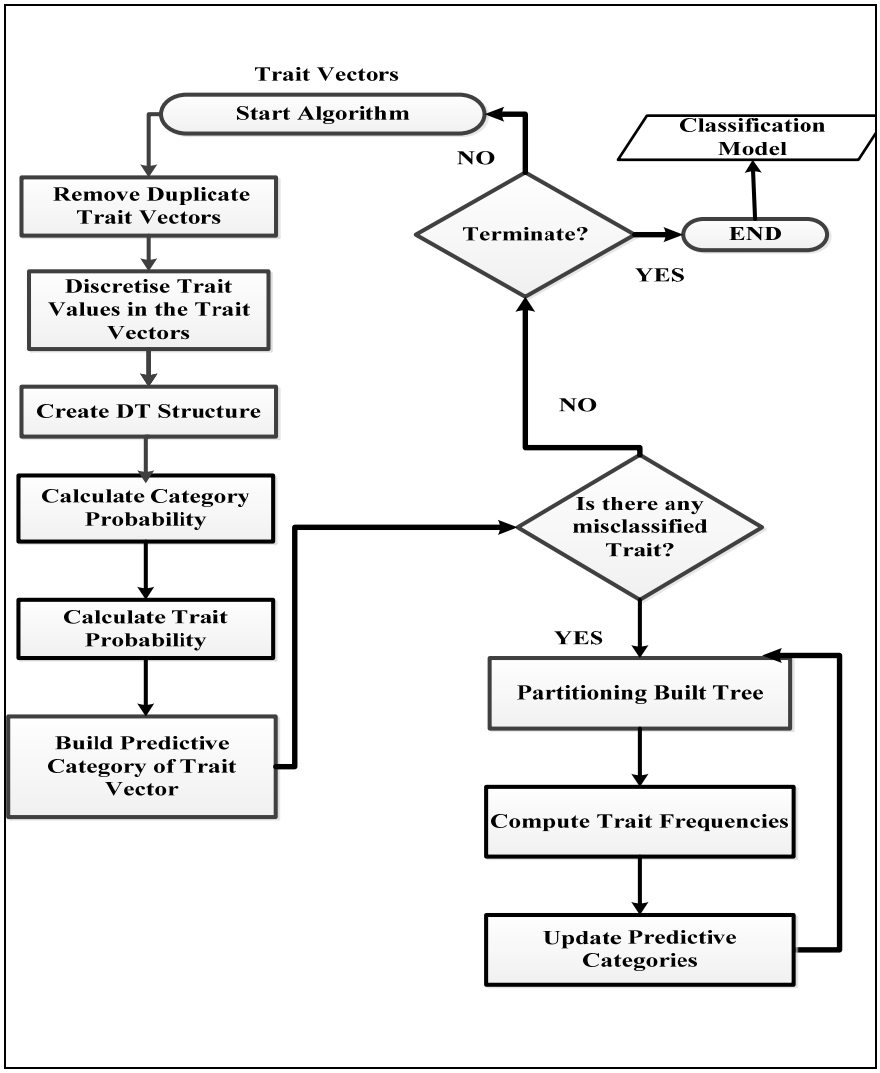


Figure 2: *Flowchart of the proposed hybrid machine learning algorithm*

To do so, the proposed algorithm learns the fetching trait vectors by leveraging the cutting back decision tree of *DT* to split the training trait vectors into sub-training vectors that would be pruned by *NB's* decision margins recursively. Thus, the training trait matrix ($T = \{T_1, \dots, T_m\}$) is given such that ($T_j = \{T_{j,i}\}_{j\in m, i\in|T_{i,j}|}$) with the prediction categories ($P_{category} = \{C_1, C_2\}: C_1 = 1, and\ C_2 = -1$). Each trait vector can be represented as $(T_j = \{C_k, T_{j,i}\}_{i\in|T_{i,j}|,k\in|C_k|}$. Then, the prior probability $P(C_k)$ is computed as per Equation (1) to predict how often each category occurs over (T) relatively to the trait vector (T_j). Whilst, the conditional probability of (T_j) is computed by Equation (2) to predict the relevance between the predicted category (C_k) and its corresponding trait ($T_{j,i}$) as it was indicated by ($P(T_{j,i}|C_k)$). Consequently, frequencies of each examined t_j in the current trait vector T_i would be checked-up across both the trait vectors and legitimate vectors to identify any redundant and/or misclassified trait such that $N_{t_i\to R}$ is the number frequencies of t_i in the trait vector, $N_{t_i\to B}$ is the number of frequencies for t_i in benign vector as per Equation (3)."

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$P(T_j|C_k) = P(C_k) \prod_{e=1 \to p} (T_{j,i}|C_k) \quad (1)$$

$$C_k = C_j \to P_{ms}(T_j, C_k) \quad (2)$$

$$P_r(P \mid t_i) = \frac{N_{t_i \to R}}{N_{t_i \to R} + N_{t_i \to B}} \quad (3)$$

3.3 Three-Tier Architecture Of Ransomware Detection Tool

Figure 3 illustrates the multi-tier architecture of the proposed anti-ransomware tool that consists of analysis tier, learning tier, and detection tier. In the analysis tier, the ransomware samples are analyzed during the five-minute routine on a virtual testbed to extract their distinctive traits. Correspondingly, the extracted traits are characterized into their related ransomware families. A virtual testbed is utilized to avoid the severe damage and malfunctions of ransomwares on the platform system. To do so, trap files (s) are created and disseminated into three different locations of hard disk partition (C: /). Once a ransomware's downloads on the targeting system, that ransomware runs and encrypts the trap file (s) before it halts the system. Then, all the traits that previously described in **Table 3**, are extracted the ransomware sample to be characterized and archived as a vector of traits belongs to the diagnosed sample. The learning tier retrieves all the archived trait vectors (both benign and ransomware trait vectors) from the data archive and learns them with the proposed hybrid machine learning algorithm to generate the classification model. Accordingly, the generated classification models will be used to check-up any suspicious sample in terms of its actions or traits in the detection tier to alert the system's user that a ransomware is going to infect the system probably.

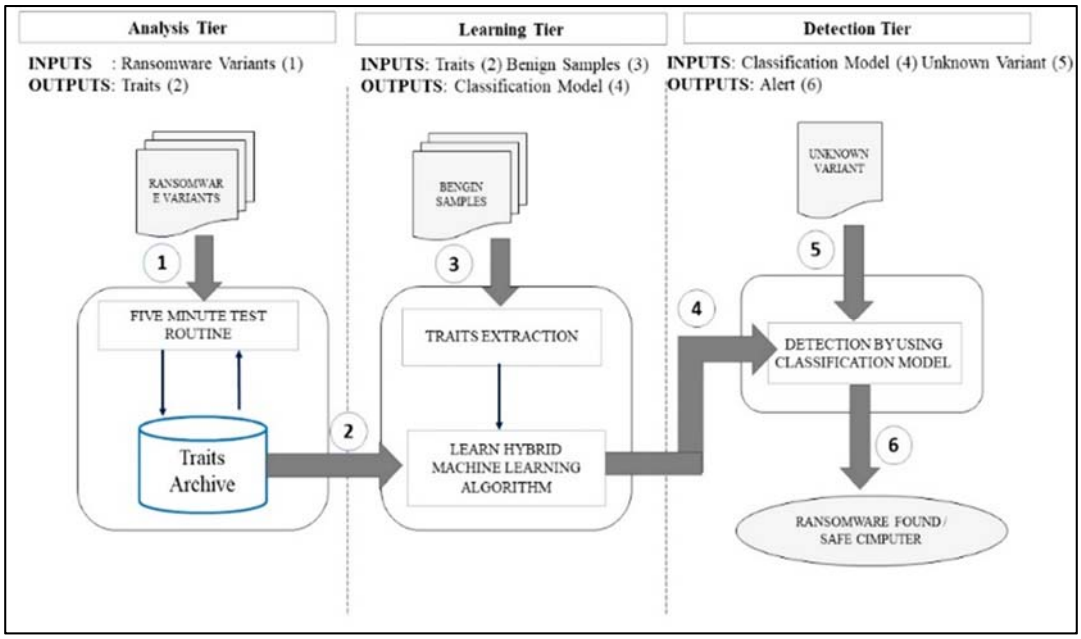


Figure 3. *The three-tier architecture of the proposed ransomware detection tool*

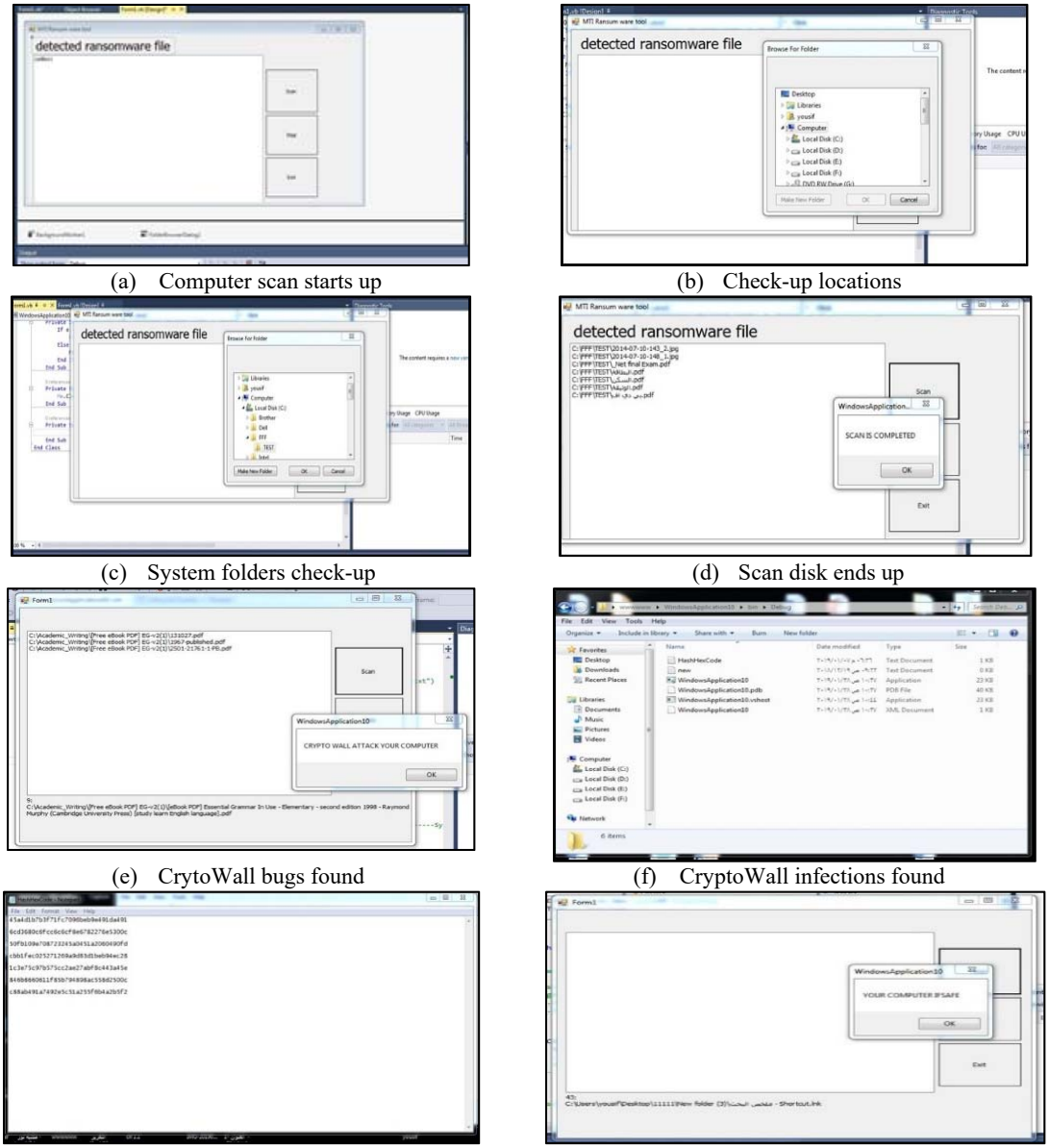
Journal of Theoretical and Applied Information Technology
15th December 2019. Vol.97. No 23

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

4. IMPLEMENTATION AND EXPERIMENTS

4.1 Implementation

To implement the analysis tier, a collected set of ransomware and benign variants are analysed in terms of their actions and infections to the trap file (s) on the virtual testbed during the five-minute test routine**.** Consequently, the learning tier is implemented such that the hybrid machine learning algorithm learns the aggregated set of samples to characterize the traits and to generate the classification models. Whilst, the detection tier pursues a computer scan to fix any new ransomware downloaded during the web browsing as it can be observed from **Figure 4**. If one or more of the downloaded file causing traits as those of the archive, then an alert will be popped up in user's screen acquiring him/her to stop the ransomware attack. Otherwise, the user is notified that his/her computer system is safe of ransomwares.



(a) Computer scan starts up (b) Check-up locations

(c) System folders check-up (d) Scan disk ends up

(e) CrytoWall bugs found (f) CryptoWall infections found

(g) CryptoWall infected files (h) User alarm declared

Figure 4: *Implementation of Learning and Detection Tiers*

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

4.2 Experimental Design

To implement and assess the developed multi-tier anti-ransomware tool, a collection of ransomware variants has been aggregated from notable data sources like *Virus Total* and *Malware Blacklist* during (1/9/2018 – 30/1/2019). Such data sources are usually adopted in the recently published studies for the same purpose [21]-[22]. Furthermore, the benign instances have been crawled from the web manually through the topmost software aggregators during the same period of time. By reducing the redundant variants, the malware variants besides excluding the suspicious samples that do not act as ransomwares, the benchmarking dataset contains 10000 valid and active ransomware variants of eight different ransomware families along with 500 benign instances as they are depicted in **Table 4**. Then, the benchmarking dataset has been formulated into a number of trait vectors that could be split up randomly into $\frac{1}{3}^{rd}$, $\frac{1}{3}^{rd}$ and $\frac{1}{3}^{rd}$ splits for analysis, learning and detection tiers respectively. To test and demonstrate whether the proposed tool could adapt the zero-day ransomware variants and their corresponding families, performance metrics including *Detection Accuracy Rate*, *Mistake Rate*, *Miss Rate*, and *Elapsed Time* along with plots of *ROC curve* have been utilized through experiments, see **Table 5**.

Table 4. *Characteristics of benchmarking dataset*

Characteristics		Benchmarking Dataset
Number of Ransomware Samples		10000
Number of Benign Samples		500
Data Archive		Malware Blacklist, Virus Total
Aggregation Time		1/9/2018-1/1/2019
Analysis Dataset		3500
Learning Dataset		3500
Detection Dataset		3500
Ransomware Families	AiDS	400
	GpCode	800
	Archiveus	1500
	WinLock	2620
	Reveton	400
	CryptoLocker	720
	CryptoWall	3250
	RaaS	310

Table 5*. Performance evaluation metrics as they adopted in [9]-[12]*

Measurement	**Description**	**Mathematical Modelling**
Accuracy Rate	It validates the effectiveness of classifier at detecting valid ransomwares (TP) and valid benign samples (TN) relatively to the whole dataset	$\text{Accuracy Rate} = \frac{TP+TN}{N_{total}}$ (4)
Mistake Rate	It validates the classifier ability to rationally detect valid ransomwares with least false detections	$\text{Mistake Rate} = \frac{FP}{N_B}$ (5)
Miss Rate	It validates the classifier's ability to rationally detect valid ransomwares with least misclassification cost	$\text{Miss Rate} = \frac{FN}{N_R}$ (6)
Elapsed Time	Calculating both execution and response time as they are spend by the developed tool to examine a batch of dataset with the nominal cost of computations	
Note:	N_{total} : N_R , and N_B denote the total number of all inspected samples included in the collected dataset, the number of correctly labelled ransomware variants, and the number of correctly labelled benign samples, respectively. FP, and FN : point out the number of benign samples that falsely detected as ransomwares, and the number of valid ransomwares that falsely identified as .benign samples	

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

4.3 Experimental Results and Discussion

Overall performance outcomes of the proposed tool across its three-tiers were cataloged in terms of *Accuracy Rate*, *Mistake* and *Miss Rates,* as shown in **Figure 5**. In addition, two experiments were conducted for *(i)* comparing the proposed hybrid algorithm versus the baseline machine learning algorithms like DT and NB via *ROC curve* plots, and *(ii)* appraising the proposed tool's outcomes versus those of its competitors like *McAfee*, *EldeRan*, *NetConverse*, and *R-Locker*, as shown in **Figures 6** and **7**. On the other hand, the plotted charts in **Figures 8** and **9** showed the progressive effectiveness of the proposed tool on the daily basis assessment during one-month test. Plots of *Accuracy Rate*, *Mistake Rate*, *Miss Rate* and *Elapsed Time* from the 1st day to the 30th day; inferred that the proposed tool could manifest its adaptive and effective classification against zero-day ransomwares. That was attributed to the proposed hybrid machine learning algorithm which hybridized and synchronized the classification margins of *NB* and *DT* into an iterative assembly to update the default margins on every batch of ransomware variants. Furthermore, the proposed tool deployed robust ransomware traits to characterize different ransomware families and identify their related variants. However, a radical escalating and/or de-escalating of performance trend line was reported at certain days. This was caused by the merits of ransomware families and their corresponding variants that might been varied in their traits, infections, crawling, analysis, and computations as well as elapsed time to thwart during the run of the proposed tool's multi-tiers. The aforesaid performance outcomes of the proposed tool as they were presented in Figures (5, 6, 7, 8 and 9) demonstrated the holistic analysis tier, the effective learning tier, and the progressive detection tier alongside the minimal performance overhead of the proposed tool during experimental, real-time and comparative assessments. Altogether, were caused by the following issues:

i. The comparable anti-ransomware tools fall short in characterizing all ransomware families and their exploitations;
ii. They varied in their detection approach (either signature, or anomaly, or hybrid based approaches) which assured different decisive rules at identifying ransomware variants;
iii. They varied in their adaptability to zero-day and/or new ransomware variants, therefore, they attained high to moderate *FNR*s;
iv. Also, they were limited in adjusting their own initial decisive margins by hybridizing different machine learning or data mining algorithms;
v. Existing anti-ransomware tools fall short in tackling ransomware variants in real-time mode.
vi. The proposed tool tackled ransomware variants during short elapsed time with a minimal computer system's footprints including: CPU utilization during processing time, detection time, and response time as well as memory usage. Thus, it outperformed its competitors with minimal performance overhead.

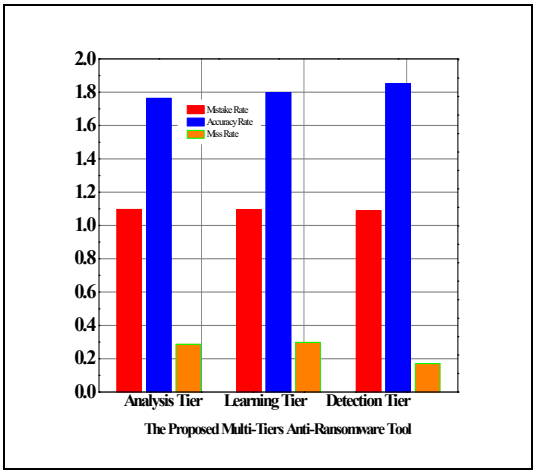


Figure 5: *Performance outcomes of the proposed tool on the benchmarking dataset*

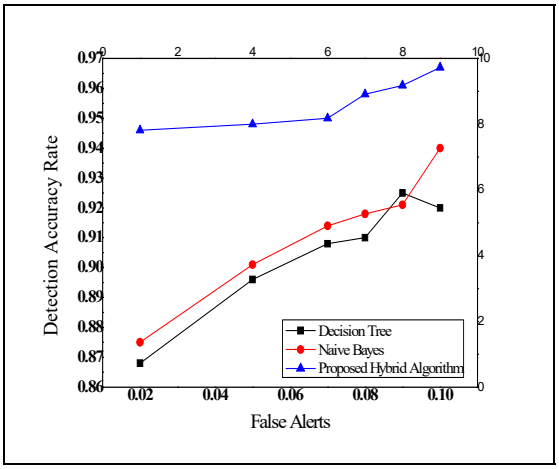
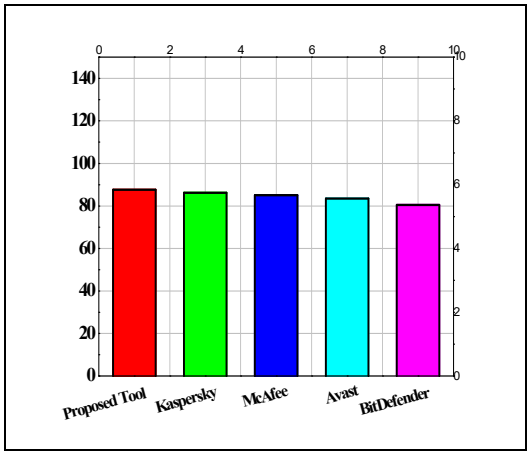
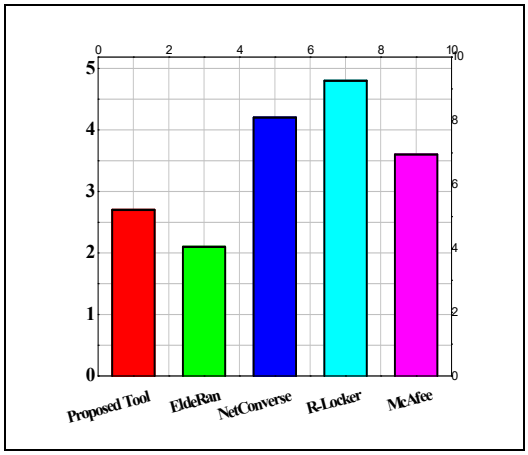


Figure 6: *ROC plots of the proposed hybrid machine learning algorithm vs. Decision Tree and Naïve Bayes*

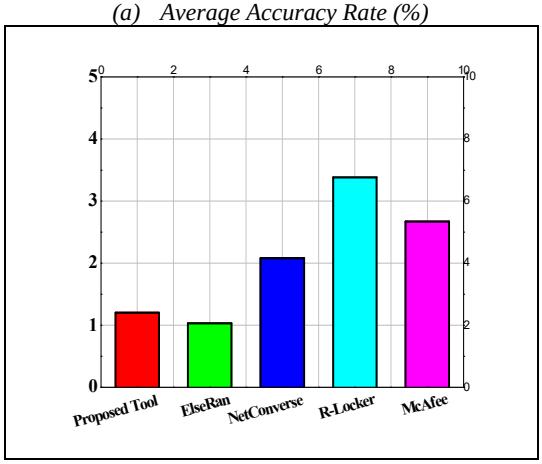
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195



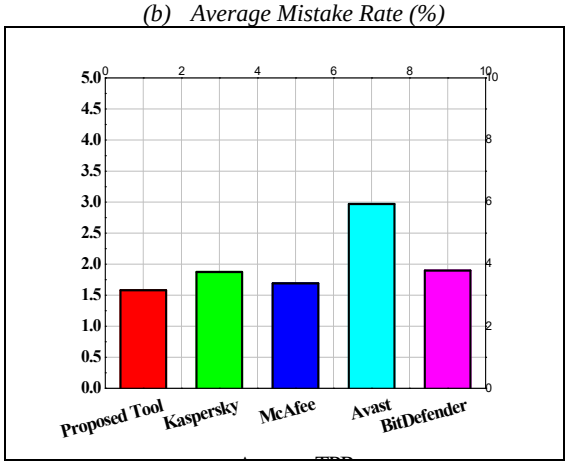
(a) Average Accuracy Rate (%)



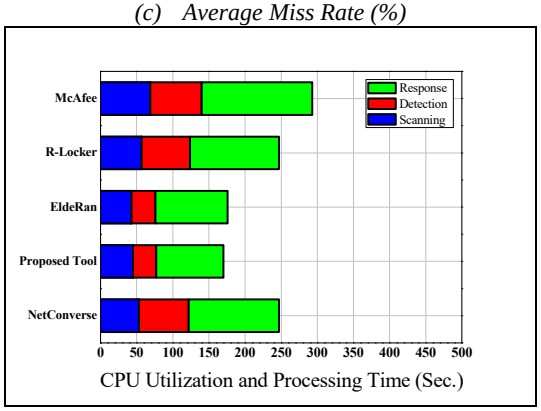
(b) Average Mistake Rate (%)



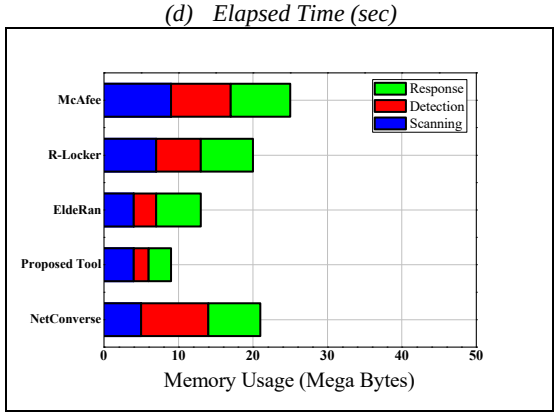
(c) Average Miss Rate (%)



(d) Elapsed Time (sec)



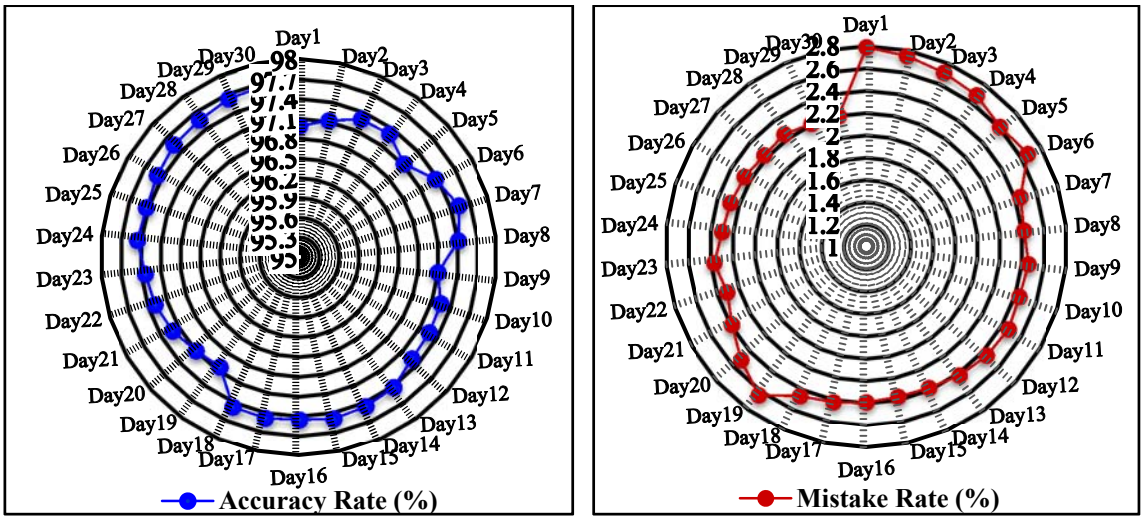
(e) CPU Utilization



(f) Memory Usage

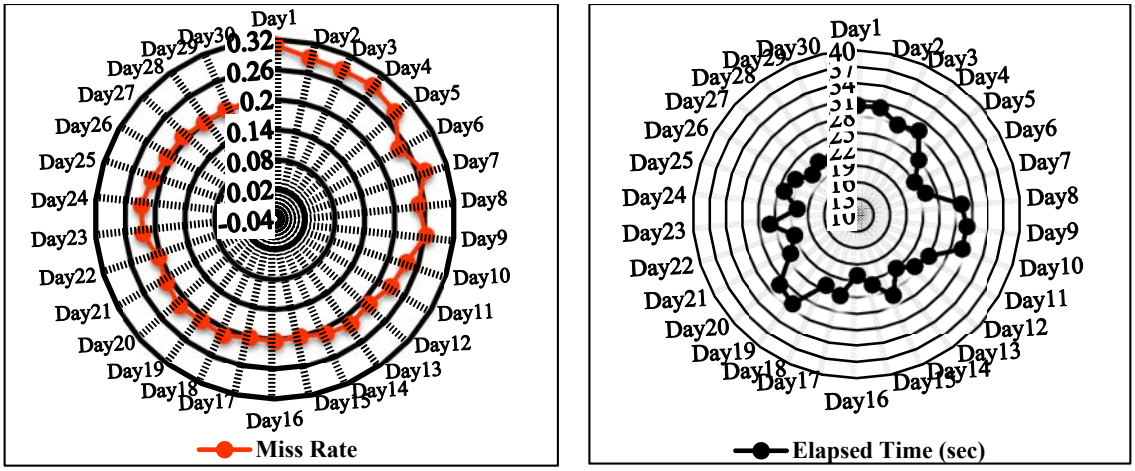
Figure 7. *Comparative analysis of the proposed tool versus anti-ransomwares tools*

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195



(a) Daily basis average Accuracy Rate *(b) Daily basis average Mistake Rates*

Figure 8: *Outcomes of real-time practice in terms of Accuracy and Mistake Rates*



(a) Daily basis average Miss Rates *(b) Daily basis average Elapsed Time*

Figure 9: *Outcomes of real-time practice in terms of Miss Rate and Elapsed Time*

5. CONCLUSIONS AND FUTURE WORK

This paper addressed the importance of hybrid machine learning algorithm that assisted by dynamic traits at characterizing ransomware families accurately, detecting the zero-day ransomwares effectively, and boosting the real-time defenselessness of hybrid-based ransomware detection tools proficiently through a three-tier paradigm. Conceptually, the proposed three-tier ransomware detection tool could synthesize and extract the dynamic traits of ransomware families during a five-minute test routine that implemented in virtual tested throughout the analysis tier. Whereas, the learning tier could learn the extracted by the proposed hybrid machine learning algorithm which hybridizes the induction functions of *Naïve Bayes* and *Decision Tree* into an iterative adjustment strategy to generate a decisive classification model. On the other hand, the detection tier could applied the generated decisive classification model adaptively at detecting zero-day ransomware variants on daily basis practice. Results of empirical, comparative, and realistic assessments showcased the effectiveness of the proposed hybrid machine learning algorithm individually and/or as it was designed in a three-tier ransomware detection tool. Precisely, both the proposed hybrid machine learning algorithm and the three-tier ransomware detection tool could achieve maximal detection

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

accuracy rates, minimal mistake rates, minimal misclassification rates, shorter elapsed time as well as minimal memory and CPU usage among their competitors. So far, the results approved that the designed ransomware detection tool can serve as real-time and publically used anti-ransomware software in the future versus the escalating number of ransomware families, their adversary traits, and their life-span variants on windows-based information systems. For future work, both static and dynamic traits of more ransomware families as well as the categorical selection of the best sets of those traits can be employed to learn the hybrid machine learning algorithm decisively.

REFERENCES:

[1] Bhardwaj, A., Avasthi, V., Sastry, H. and Subrahmanyam, G.V.B.: Ransomware digital extortion: a rising new age threat. Indian Journal of Science and Technology, 9(14), pp.1-5 (2016).

[2] Symantec. 2016 Internet Security Threat Report. https://www.symantec.com/content/dam/symantec/docs/reports/istr-21-2016-en.pdf, 2016.

[3] Tailor, J.P. and Patel, A.D.: A comprehensive survey: ransomware attacks prevention, monitoring and damage control. International Journal of Research, Science and Innovation, 4, pp.2321-2705, 2017.

[4] Richardson, R. and North, M.: Ransomware: Evolution, mitigation and prevention. International Management Review, 13(1), pp.10-21, 2017.

[5] Genç, Z.A., Lenzini, G. and Ryan, P.Y., Security analysis of key acquiring strategies used by cryptographic ransomware. In Proceedings of the Central European Cybersecurity Conference 2018 (p. 7), ACM, November, 2018.

[6] Morato, D., Berrueta, E., Magaña, E. and Izal, M., Ransomware early detection by the analysis of file sharing traffic, Journal of Network and Computer Applications, 124, pp.14-32, 2018.

[7] Gómez-Hernández, J. A., Álvarez-González, L., & García-Teodoro, P., R-Locker: Thwarting ransomware action through a honeyfile-based approach. *Computers & Security*, *73*, 389-398, 2018.

[8] Cabaj, K., Gregorczyk, M. and Mazurczyk, W., Software-defined networking-based crypto ransomware detection using HTTP traffic characteristics. Computers and Electrical Engineering, 66, pp.353-368, 2018.

[9] Hampton, N., Baig, Z. and Zeadally, S.: Ransomware behavioural analysis on windows platforms. Journal of Information Security and Applications, 40, pp.44-51, 2018.

[10] Sgandurra, D., Muñoz-González, L., Mohsen, R., and Lupu, E. C., Automated dynamic analysis of ransomware: benefits, limitations and use for detection, 2016.

[11] Zimba, A., Malware-free intrusion: a novel approach to Ransomware infection vectors, International Journal of Computer Science and Information Security, 15(2), p.317, 2017.

[12] Kharaz, A., Arshad, S., Mulliner, C., Roberson, W. K., and Krida, E., Unveil: a large scale, automated approach to detecting ransomware. In USINEX Security Symposium, pp. 757-772 (2016)

[13] Zavarsky, P. and Lindskog, D., Experimental analysis of ransomware on windows and android platforms: Evolution and characterization. Procedia Computer Science, 94, pp.465-472 (2016).

[14] Alhawi, O.M., Baldwin, J. and Dehghantanha, A., Leveraging machine learning techniques for windows ransomware network traffic detection, Cyber Threat Intelligence, pp.93-106, 2018.

[15] Nieuwenhuizen, D.: A behavioural-based approach to ransomware detection. Whitepaper. MWR Labs Whitepaper , 2017.

[16] Christensen, J.B. and Beuschau, N., Ransomware detection and mitigation tool. M.Sc. Thesis, Technical University of Denmark, 2017.

[17] Continella, A., Guagnelli, A., Zingaro, G., De Pasquale, G., Barenghi, A., Zanero, S. and Maggi, F., ShieldFS: a self-healing, ransomware-aware filesystem. In Proceedings of the 32nd Annual Conference on Computer Security Applications, pp. 336-347, ACM, December, 2016.

[18] Ahmadian, M.M. and Shahriari, H.R., 2entFOX: A framework for high survivable ransomwares detection. In 2016 13th International Iranian Society of Cryptology Conference on Information Security and Cryptology (ISCISC), pp. 79-84, IEEE, September, 2016.

[19] Andronio, N., Heldroid: fast and efficient linguistic-based ransomware detection, Doctoral Dissertation, University of Illinois, Chicago, USA, 2015.

[20] Feng, Y., Liu, C. and Liu, B., Poster: a new approach to detecting ransomware with

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

deception, In 38th IEEE Symposium on Security and Privacy, IEEE, 2017.
[21] MalwareBlackList - Online Repository of Malicious URLs. http://www.malwareblacklist.com.
[22] Virus Total-Intelligence search Engine, https://www.virustotal.com