POLİTEKNİK DERGİSİ

JOURNAL of POLYTECHNIC

Reliability of entropy-based malware detection as a single method in preventing ransomware attacks

Entropi temelli kötü amaçlı yazılım tespit yönteminin fidye yazılımı saldırılarını önlemede tek başına güvenilirliği

***Authors (Yazarlar)**: Abdulkerim Oğuzhan ALKAN[1], İbrahim Alper DOĞRU[2], İsmail ATACAK[3]*

ORCID[1]: 0000-0003-3505-196X

ORCID[2]: 0000-0001-9324-7157

ORCID[3]: 0000-0002-6357-0073

To cite to this article: Alkan A. O., Doğru İ. A. ve Atacak İ., "Effective Management of Rapid Intervention, Investigation, Analysis and Reporting Processes in Computer Crimes with New-Generation Digital Forensic Methods", ***Journal of Polytechnic***, *(*): *, (*).

Bu makaleye şu şekilde atıfta bulunabilirsiniz: Alkan A. O., Doğru İ. A. ve Atacak İ., "Entropi temelli kötü amaçlı yazılım tespit yönteminin fidye yazılımı saldırılarını önlemede tek başına güvenilirliği", ***Politeknik Dergisi***, *(*): *, (*).

To link to this article (Erişim linki): http://dergipark.org.tr/politeknik/archive

DOI: 10.2339/politeknik.1537076

Reliability of Entropy-based Malware Detection as a Single Method in Preventing Ransomware Attacks

Highlights

- *Malware detection methods to prevent ransomware attacks*
- *Reliability of entropy-based malware detection method*
- *Holistic approach to preventing ransomware attacks*

Graphical Abstract

Although entropy-based malware detection is a useful method to prevent ransomware attacks because it is easily applicable and provides fast results, the results of our study suggest that it is not sufficient on its own to prevent ransomware attacks.

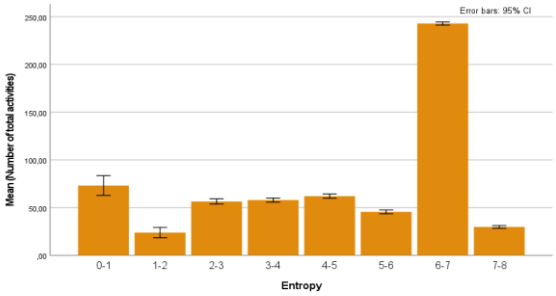

Figure: Graphical abstract

Aim

This study aims to evaluate the reliability of the entropy-based malware detection method.

Design & Methodology

As a result of the data obtained by using Binalyze AIR and Binalyze Tactical software on a computer known to be infected with malware, two groups were created by identifying software that was determined not to be malware and software that was determined to be malware. Comparisons were made between inter-group and intra-group entropy values.

Originality

By evaluating the entropy values of malware and non-malware software, the reliability of the entropy-based malware detection method was evaluated.

Findings

Evaluation was made on 44834 software that were determined not to be malware by Binalayze and 41447 software that were determined to be malware, a total of 86281 software. It was observed that the average entropy values of two groups were 6.3 ± 0.3 and 6.4 ± 0.9, respectively. No linear relationship was found between entropy, data size and total number of activities.

Conclusion

The entropy-based method alone is unreliable unless combined with hybrid models. More advanced and holistic approaches need to be adopted to provide effective cybersecurity defenses.

Declaration of Ethical Standards

The authors of this article declare that the materials and methods used in this study do not require ethical committee permission and/or legal-special permission.

Reliability of Entropy-based Malware Detection as a Single Method in Preventing Ransomware Attacks

Research Article

Abdulkerim Oğuzhan ALKAN1*, İbrahim Alper DOĞRU2, İsmail ATACAK2

1Department of Computer Forensics, Graduate School of Informatics, Gazi University, Ankara, Turkey
2Department of Computer Engineering, Faculty of Technology, Gazi University, Ankara, Turkey
(Received: 22.08.2024 ; Accepted: 22.09.2024 ; Early View: 03.10.2024)

ABSTRACT

As the complexity of ransomware attacks increases, traditional detection methodologies are often insufficient for detecting and preventing threats. Therefore, modern malware detection methods are used. These are the behavior-, system-, resource-, connection- and entropy-based ransomware detection methods. In this study, we evaluated the effectiveness of an entropy-based malware detection method in detecting ransomware attacks by evaluating the entropy values of malware detected using Binalayze AIR and Binalayze Tactical software. As revealed in the results of our comprehensive field study in which 41477 malware were evaluated, although the entropy-based malware detection method has advantages in that it is easily applicable, can be integrated with other methods, and provides fast results, it can give high rates of false-positive and false-negative results when used alone. The entropy-based method is unreliable unless it is used with hybrid models. More advanced and holistic approaches must be adopted for effective cybersecurity defense.

Keywords: Ransomware attacks, malware detection, entropy-based malware detection, Binalayze

Entropi Temelli Kötü Amaçlı Yazılım Tespit Yönteminin Fidye Yazılımı Saldırılarını Önlemede Tek Başına Güvenilirliği

ÖZ

Fidye yazılımı saldırılarının karmaşıklığı arttıkça, geleneksel yöntemlerin tehditleri tespit etme ve önleme konusunda yetersiz kalmaya başlaması sebebiyle modern kötü amaçlı yazılım tespit yöntemleri kullanılmaya başlanmıştır. Bunlar davranış temelli algılama, sistem temelli algılama, kaynak temelli algılama, bağlantı temelli algılama ve entropi temelli fidye yazılımı algılamadır. Bu çalışmada Binalayze AIR ve Binalayze Tactical yazılımları yardımıyla tespit edilen kötü amaçlı yazılımların entropi değerlerini değerlendirerek, entropi temelli kötü amaçlı yazılım tespit yönteminin fidye yazılımı saldırılarını tespit etme ve önlemedeki etkinliği değerlendirilmeye çalışılmıştır. 41477 kötü amaçlı yazılımın değerlendirildiği kapsamlı saha çalışmamızın sonuçlarında da ortaya konduğu üzere, entropi temelli kötü amaçlı yazılım tespit yönteminin kolay uygulanabilir olması, diğer yöntemlerle entegre olarak kullanılabilmesi ve hızlı sonuç vermesi gibi avantajları olmasına rağmen tek başına kullanıldığında yüksek oranda yanlış pozitif ve yanlış negatif sonuçlar verebilmektedir. Entropi temelli yöntem, hibrit modellerle birlikte kullanılmadığı sürece tek başına güvenilir değildir. Etkili siber güvenlik savunmaları sağlamak için daha gelişmiş ve bütünsel yaklaşımların benimsenmesi gerekmektedir.

Anahtar kelimeler: Fidye yazılımı atakları, kötücül yazılım tespiti, entropi temelli kötücül yazılım tespiti, Binalayze

1. INTRODUCTION

In a world where all infrastructure is digitalized, the importance of protecting data and digital infrastructure from ransomware attacks is increasing. As the complexity of ransomware attacks increases, traditional detection methodologies are often insufficient for detecting and preventing threats. This has led to the replacement of traditional digital forensic approaches with modern digital forensic techniques.

Modern forensic tools employ five ransomware-detection methods: behavior-based, system-based, resource-based, connection-based, and entropy-based detection methods [1-3].

Behavior-based detection targets signatures within system files that execute malicious actions. Despite its utility, this method has two significant drawbacks: failure in detection of new ransomware and high false-positive and false-negative rates [4].

*Corresponding Author
e-mail : aoguzhan.alkan@gazi.edu.tr

System-based behavior detection focuses on identifying malicious activities by employing integrity checks and behavior blocking. This technique validates the integrity of files and directories because ransomwares typically encrypt files, necessitating the monitoring of irregular file access behavior. Unfortunately, this method also results in high false-positive rates and requires extensive time for integrity verification [5].

Resource-based behavior detection aims to identify malicious activities that cause anomalies such as increased CPU and I/O usage. These abnormalities arise because file access requires I/O operations, and encryption requires CPU resources. However, this approach is time-consuming because of the need to gather resource information, which results in high false-positive rates [6].

Connection-based behavior detection monitors network activities, particularly the characteristics of ransomware connected to C & C servers, to obtain encryption keys. However, it cannot identify ransomware that does not establish network connections [7].

The entropy-based ransomware detection method, which is the focus of this study, detects files infected by ransomware by analyzing high-entropy levels typical of ciphertexts. Entropy, as defined by Shannon, measures the uncertainty or randomness in a dataset [8]. Encrypted malware typically exhibits higher entropy than benign files with more structured and predictable patterns [9].

Binalyze AIR and Binalyze Tactical (Binalyze, Ankara, Turkey) are modern forensic programs that help identify malware by accelerating digital evidence collection and automatic analysis processes. By leveraging machine learning algorithms and advanced analytics, Binalyze helps detect findings that may indicate malicious behavior, even without signs of known malware [10].

In this study, we evaluated the effectiveness of the entropy-based malware detection method in detecting ransomware attacks by evaluating the entropy values of malware detected using Binalyze AIR and Binalyze Tactical software. The rest of the study is outlined as follows. The methods section describes a method based on entropy-based statistical analyses of data collected from a malware-infected computer. The statistical analysis section gives the tests and analyses conducted using SPSS software. The results section presents the statistical findings derived from these analyses. The last section discusses and evaluates the feasibility of entropy-based approaches for attack detection.

2. METHODS

In our study, evidence was collected using Binalyze AIR and Binalyze Tactical software on a computer known to be infected with malware. An external storage disk containing the Binalyze Tactical software file, operable without installation, was connected to the supplied computer after turning it on. Then, to collect data, the file was run, and the data were collected. The data generated as a result of the collection process totaled 145 MB. Later, the same process was performed using Binalyze AIR software.

From the data obtained, two groups were created by defining the hashes that were determined not to be malware as "Malware 0" and the hashes that were detected as malware by the software as "Malware 1." Entropy values of the two groups were compared. In addition, the statistical relationship between the entropy values of malware hashes and "Size of data" and the "Total number of activities" and data were evaluated.

3. STATISTICAL ANAYSIS

Statistical analyses were performed using the Statistical Package for Social Sciences (SPSS) version 26.0 for Windows (SPSS Inc., Chicago, IL, USA). The mean, standard deviation, median, minimum and maximum value frequency, and percentage were used for descriptive statistics. The distribution of variables was checked using the Kolmogorov-Smirnov test. Kruskal-Wallis (Mann-Whitney U) tests were used to compare quantitative data. Spearman's correlation coefficient was used to examine the relationships among the measured variables. The statistical significance level was $P < 0.05$.

4. RESULTS

Within the scope of the study, an evaluation was made on 44834 software that were determined not to be malware by Binalyze and 41447 software that were determined to be malware, for a total of 86281 software. It was observed that the average entropy values between Malware 0 and Malware 1 groups were 6.3 ± 0.3 and 6.4 ± 0.9, respectively. Although the difference was statistically significant owing to the large number of observations, it has no practical meaning (Table 4.1).

Table 4.1. Comparison of entropy values between 2 groups

	Malware 0 (n = 44834)		Malware 1 (n = 41477)		p†
	Mean ± SD	Median (Min-Max)	Mean ± SD	Median (Min-Max)	
Entropy	6.3 ± 0.3	6.4(1.7-8)	6.4 ± 0,9	6.6 (0-8)	**< 0,001 ***

†Mann-Whitney U test, SD: Standard Deviation

As shown in Table 4.2 and Table 4.3, the relationship between the entropy values of 41447 hashes detected as Malware 1 and the "Size of data" and "Total number of activities" data was evaluated. Although a statistically significant result was obtained owing to the large number of observations, no linear relationship was observed between entropy and the other variables (Figure 4.1, 4.2).

Table 4.2. Correlation Analysis

	r	p	%95 CI	
			Lower	Upper
Entropy x Size of data	0,183	0,000	0,173	0,193
Entropy x Total number of activities	0,127	< 0,001	0,117	0,137

r: Spearman Correlation Coefficient, CI: Confident Interval

Table 4.3. Distribution of size of data and total number of activities according to entropy values

	Size of data		**Total number of activities**	
Entropy Category	Mean ± SD	Median (Min-Max)	Mean ± SD	Median (Min-Max)
0-1 (n=257)	291242 ± 1658690	0 (0-16855040)	73.21 ± 84.6	21 (0-404)
1-2 (n=71)	123508 ± 163667	98304 (8-601600)	23.89 ± 22.97	17 (0-68)
2-3 (n=311)	1573211 ± 1075506	2197504 (512-4329472)	56.5 ± 24.47	65 (1-244)
3-4 (n=704)	1249975 ± 911545	1503744 (1024-3067904)	57.98 ± 29.76	66 (0-287)
4-5 (n=1209)	413665 ± 661645	115712 (1024-79872000)	62.14 ± 39.34	75 (0-406)
5-6 (n=2047)	252513 ± 533194	63488 (512-7454720)	45.61 ± 46.32	37 (0-573)
6-7 (n=33658)	656173 ± 1085403	262144 (512-19537920)	242.94 ± 151.45	247 (0-692)
7-8 (n=3112)	594746 ± 791262	475392 (3072-22578176)	29.85 ± 42.05	3 (0-392)

SD: Standard deviation

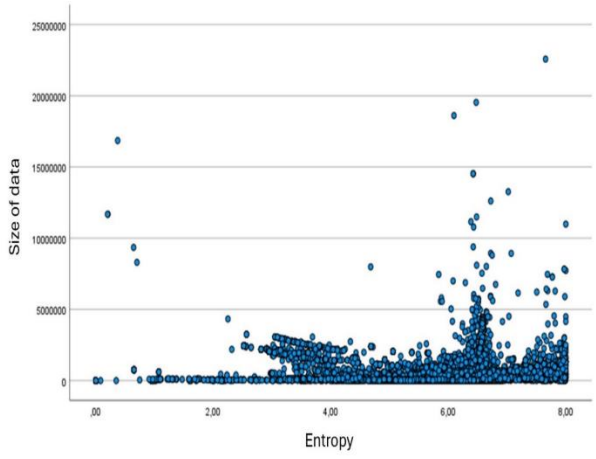

Figure 4.1. Scatter plot of the relationship between Entropy and Size of data

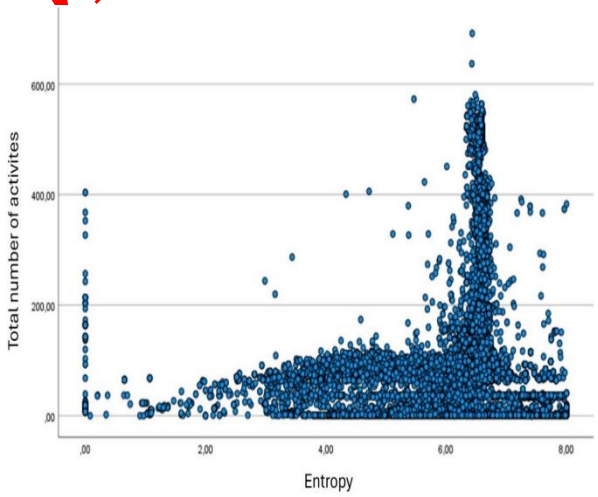

Figure 4.2. Scatter plot of the relationship between Entropy and Total number of activities

5. CONCLUSION and DISCUSSION

In the current study, the entropy values of malwares detected with the help of Binalayze Air and Binalayze Tactical software were examined and it was revealed within the scope of our laboratory study that the entropy-based malware detection method alone was insufficient in terms of preventing ransomware attacks.

Entropy-based malware detection methods measure the entropy of a file using the Shannon entropy formula. Malware often results in abnormal entropy levels compared to harmless files, owing to activities such as compression and encryption [11]. One of the most advantageous features of entropy-based malware-detection methods is their ease of application. Measuring entropy is easier than other complex methodologies, such as behavioral analysis or machine learning-based approaches. This brings this method to the forefront for cases in which rapid results are required. [12]. Because entropy measurements do not depend on content type, they can be used to analyze executable files, documents, or other file types that may potentially contain malicious code. This flexibility increases the applicability of the method in various environments ranging from personal computers to large-scale enterprise systems. [13]. Additionally, this method does not benefit from signature-based databases, which are important for traditional methods. Thus, entropy-based methods can be used to detect malware that escapes detection by being polymorphic or metamorphic with the help of code changes. [14, 15].

The applicability of entropy-based methods makes them easy to use in conjunction with other malware detection techniques, such as behavioral analysis and machine learning approaches. It is important to quickly filter suspicious files because this saves time and allows for a more detailed examination before further examination can be carried out with other techniques. [16].

Although the entropy-based malware detection method is widely used today owing to its easy applicability and integration into all small- and large-scale systems and other methods, it is not sufficient for malware detection alone, as revealed by the results of our field study. The most important reason for this is that this method does not allow behavioral analysis because it is static, and it may cause false-positive and false-negative results owing to different entropy values depending on the extension of the malware-infected file.

Document files (.docx,.pdf) exhibited a wide range of entropy values. Documents, particularly those containing text, exhibit low entropy levels. However, when these files are infected with malware, the entropy can increase significantly. The entropy level can fluctuate depending on how much of the document has been altered or obscured by malware [17, 18]. Script files (.js, .vbs) may naturally have higher entropy owing to their coding structure. Because these files contain various coding and data structures, they can also exhibit high entropy before being infected with malware. However, when script files are infected with malware, the addition of malicious codes can create higher levels of entropy, thereby increasing the variability and complexity of the content. [19, 20].

Multimedia files (.mp4, .jpg) generally exhibited high entropy. Because such files are created using compression algorithms that increase the randomness of the data and lead to high entropy values, they naturally exhibit complexity and high entropy. When malware infects these file types, the malicious content usually overlaps the original compression or encoding structure, leaving no significant change in the entropy value. However, in certain cases, malicious content can corrupt the file structure, resulting in unexpected changes in the entropy value. [9, 21]. Executable files (.exe, .dll) often exhibited higher entropy values when infected with malware.[22].

For these reasons, it may be difficult to determine the appropriate entropy thresholds to distinguish malware-infected files. Different software and file types yield a wide range of entropy values. Therefore, a uniform threshold value may lead to false-positive or false-negative results. As software and data diversity increase, entropy values may need to be constantly adjusted and manually calibrated [15, 23].

However, entropy-based detection techniques rely on the static analysis of file content and do not provide information regarding the behavior of a file when run. This lack of dynamic analysis fails to detect the possibility of malicious behavior even when the file is detected to be within normal limits in terms of entropy [19, 24-26].

The Binalayze software, which we use as proof of our results, detects malware using advanced techniques such as machine learning and behavior-based methods along with the entropy-based method. This is because even advanced methods can produce false positive and false negative results when used alone.

Many studies have reported that entropy-based methods are more successful when combined with other methods. Even the success of malware detection using machine-learning methods, whose use has increased recently, and successful results have been reported, is limited when used alone. The success of this method increases, even when combined with traditional methods [27]. Similarly, recent studies have shown that machine learning- and behavior-based methods are more successful when used together with other methods [18, 27-33]. In addition, when system-, resource-, and connection-based methods are used together with other entropy-based methods, a significant increase in success is achieved in preventing ransomware attacks [9, 16, 23, 28,

34-36]. All of these studies reported that the entropy-based method is successful at rates not exceeding 80% when used alone, whereas hybrid models can achieve a success rate of over 95%.

As a result, as revealed in the results of our extensive field study in which 41477 malware were evaluated, although the entropy-based malware detection method has advantages in that it is easily applicable, can be used in combination with other methods, and provides fast results, it can give high rates of false-positive and false-negative results when used alone. The entropy-based method is not reliable unless it is used with hybrid models and supported by other methods, such as machine learning and behavior-based methods. More advanced and holistic approaches must be adopted to provide an effective cybersecurity defiance. We believe that our study yielded important results that may serve as a basis for further research in this field.

DECLARATION OF ETHICAL STANDARDS

The author(s) of this article declare that the materials and methods used in this study do not require ethical committee permission and/or legal-special permission.

REFERENCES

[1] Cabaj K, Gregorczyk M, Mazurczyk W, "Software-defined networking-based crypto ransomware detection using HTTP traffic characteristics", ***Computers & Electrical Engineering,*** 66:353-68, (2018).

[2] Paik J-Y, Choi J-H, Jin R, Wang J, Cho E-S, "A storage-level detection mechanism against crypto-ransomware", ***Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security***, (2018).

[3] Al-Rimy BAS, Maarof MA, Shaid SZM, "Ransomware threat success factors, taxonomy, and countermeasures: A survey and research directions", ***Computers & Security***, 74:144-66, (2018).

[4] Kim D, Kim S, "Design of quantification model for ransom ware prevent", ***World Journal of Engineering and Technology,*** 3(03):203, (2015).

[5] Song S, Kim B, Lee S, "The effective ransomware prevention technique using process monitoring on android platform", ***Mobile Information Systems***, (2016-1):2946735, (2016).

[6] Nieuwenhuizen D., "A behavioural-based approach to ransomware detection", ***Whitepaper MWR Labs Whitepaper,*** 2017:20, (2017).

[7] Ahmadian MM, Shahriari HR, Ghaffarian SM, "Connection-monitor & connection-breaker: A novel approach for prevention and detection of high survivable ransomwares", ***2015 12th International Iranian society of cryptology conference on information security and cryptology (ISCISC),*** 2015: IEEE, (2015).

[8] Shannon CE, "A mathematical theory of communication", ***The Bell system technical journal***, 27(3):379-423, (1948).

[9] Davies SR, Macfarlane R, Buchanan WJ, "Differential area analysis for ransomware attack detection within mixed file datasets", ***Computers & Security***, 108:102377, (2021).

[10] Saxe J, Berlin K, "Deep neural network based malware detection using two dimensional binary program features", ***2015 10th international conference on malicious and unwanted software (MALWARE),*** 2015: IEEE, (2015).

[11] Lee K, Lee S-Y, Yim K, "Effective ransomware detection using entropy estimation of files for cloud services", ***International Symposium on Pervasive Systems, Algorithms and Networks***, 2019: Springer, (2019).

[12] Kornblum J, "Identifying almost identical files using context triggered piecewise hashing", **Digital investigation**, 3:91-7, (2006).

[13] Deng X, Jiang M, Cen M, "A Ransomware Classification Method Based on Entropy Map", ***2022 IEEE 21st International Conference on Ubiquitous Computing and Communications (IUCC/CIT/DSCI/SmartCNS),*** 2022: IEEE, (2022).

[14] You I, Yim K, "Malware obfuscation techniques: A brief survey", ***2010 International conference on broadband, wireless computing, communication and applications***, 2010: IEEE, (2010).

[15] Lyda R, Hamrock J, "Using entropy analysis to find encrypted and packed malware", ***IEEE Security & Privacy***, 5(2):40-5, (2007).

[16] Paik JY, Jin R, Cho ES, "Malware classification using a byte-granularity feature based on structural entropy", ***Computational Intelligence***, 38(4):1536-58, (2022).

[17] Shafiq MZ, Khayam SA, Farooq M, "Embedded malware detection using markov n-grams", ***International conference on detection of intrusions and malware, and vulnerability assessment***, 2008: Springer, (2008).

[18] Han J, Lin Z, Porter DE, "On the effectiveness of behavior-based ransomware detection", ***Security and Privacy in Communication Networks: 16th EAI International Conference, SecureComm 2020, Washington, DC, USA, October 21-23, 2020, Proceedings, Part II 16***, 2020: Springer, (2020).

[19] Davies SR, Macfarlane R, Buchanan WJ, "Comparison of entropy calculation methods for ransomware encrypted file identification", ***Entropy***, 24(10):1503, (2022).

[20] Hsu C-M, Yang C-C, Cheng H-H, Setiasabda PE, Leu J-S, "Enhancing file entropy analysis to improve machine learning detection rate of ransomware", ***IEEE Access***, 9:138345-51, (2021).

[21] Fridrich J, Goljan M, Du R, "Detecting LSB steganography in color, and gray-scale images", ***IEEE multimedia***, 8(4):22-8, (2001).

[22] Guo F, Ferrie P, Chiueh T-C, "A study of the packer problem and its solutions", ***International Workshop on Recent Advances in Intrusion Detection***, 2008: Springer, (2008).

[23] De Gaspari F, Hitaj D, Pagnotta G, De Carli L, Mancini LV, "Evading behavioral classifiers: a comprehensive analysis on evading ransomware detection techniques", ***Neural Computing and Applications***, 34(14):12077-96, (2022).

[24] Moser A, Kruegel C, Kirda E, "Exploring multiple execution paths for malware analysis", ***2007 IEEE Symposium on Security and Privacy (SP'07)***, 2007: IEEE, (2007).

[25] Christodorescu M, Jha S, "Static analysis of executables to detect malicious patterns", ***12th USENIX Security Symposium (USENIX Security 03)***, (2003).

[26] Jung S, Won Y, "Ransomware detection method based on context-aware entropy analysis", ***Soft Computing***, 22(20):6731-40, (2018).

[27] Gibert D, Mateu C, Planes J, "The rise of machine learning for detection and classification of malware: Research developments, trends and challenges", ***Journal of Network and Computer Applications***, 153:102526, (2020).

[28] Maniriho P, Mahmood AN, Chowdhury MJM, "A study on malicious software behaviour analysis and detection techniques: Taxonomy, current trends and challenges", ***Future Generation Computer Systems***, 130:1-18, (2022).

[29] Arabo A, Dijoux R, Poulain T, Chevalier G, "Detecting ransomware using process behavior analysis", ***Procedia Computer Science,*** 168:289-96, (2020).

[30] Hwang J, Kim J, Lee S, Kim K, "Two-stage ransomware detection using dynamic analysis and machine learning techniques", ***Wireless Personal Communications***, 112(4):2597-609, (2020).

[31] Chew CJ, Kumar V, ***Behaviour based ransomware detection***, (2019).

[32] Rosli NA, Yassin W, Faizal M, Selamat SR, "Clustering analysis for malware behavior detection using registry data", ***International Journal of Advanced Computer Science and Applications (IJACSA)***, 10:12, (2019).

[33] Urooj U, Al-rimy BAS, Zainal A, Ghaleb FA, Rassam MA, "Ransomware detection using the dynamic analysis and machine learning: A survey and research directions", ***Applied Sciences***, 12(1):172, (2021).

[34] Hurtuk J, Chovanec M, Kičina M, Biľlík R, "Case study of ransomware malware hiding using obfuscation methods", ***2018 16th International Conference on Emerging eLearning Technologies and Applications (ICETA)***, 2018: IEEE, (2018).

[35] Herrera Silva JA, "Ransomware detection by cognitive security, ***EPN***, (2023).

[36] Lee J, Lee K, "A method for neutralizing entropy measurement-based ransomware detection technologies using encoding algorithms", ***Entropy***, 24(2):239, (2022).