2024 International Conference on Electrical Electronics and Computing Technologies (ICEECT) | 979-8-3503-7809-2/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICEECT61758.2024.10738908

Ransomware Detection And Data Recovery

Amit Kumar Upadhyay
Department of computer Science and Engineering
Sharda University
Greater Noida, Uttar Pradesh, India
amitkumar.upadhyay@sharda.ac.in

Preeti Dubey
Department of computer Science and Engineering
Sharda University
Greater Noida, Uttar Pradesh, India
preetidubey.research@gmail.com

Sahil Gandhi
Department of computer Science and Engineering
Sharda University
Greater Noida, Uttar Pradesh, India
2020558875.sahil@ug.sharda.ac.in

Shreya Jain
Department of computer Science and Engineering
Sharda University
Greater Noida, Uttar Pradesh, India
2020555794.shreya@ug.sharda.ac.in

***Abstract*—Attacks using ransomware have become a major cybersecurity concern that is both widespread and constantly changing, causing misery on people, companies, and organizations all over the world. An in-depth analysis of the ransomware detection and data recovery environment is provided in this survey study. It looks at how ransomware has changed over time, examines the many types of ransomware detection techniques, and looks at the most recent developments in data recovery techniques. We examine the advantages and disadvantages of signature-based, behavioral analysis, machine learning, and other detection technologies, illuminating their potency against both well-known and newly discovered ransomware variants. Additionally, we discuss current developments in data recovery techniques, such as backup options, decryption tools, and incident response plans. Through this survey, we hope to give cybersecurity professionals, researchers, and decision-makers a comprehensive picture of the ransomware threat environment as well as the tools and approaches that may be used to effectively battle it.**

Keywords—ransomware, detection, data recovery, machine learning

I. INTRODUCTION

A. Background

In our increasingly digitalized world, the rise of ransomware and other cyberattacks poses serious risks to people, businesses, and governments. One of the most destructive and profit-driven cybercrimes nowadays is ransomware, which is notorious for encrypting files, jeopardizing data integrity, and extorting victims. Its quick evolution over the last ten years has presented difficulties for both researchers and cybersecurity specialists. Techniques for identifying and recouping from ransomware attacks need to evolve along with criminal techniques. This survey explores the intricacies of ransomware, looking at several ways to identify it, how to prevent it, and how this cyberthreat is always changing.

It also highlights new developments in data recovery methods that give hope for obtaining important data. The paper provides crucial insights for navigating this dangerous digital terrain by synthesizing research, technological advancements, and continuous efforts in the battle against ransomware.

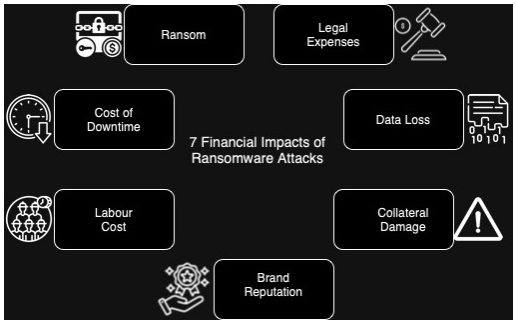

Figure 1. Financial Impact of Ransomware Attacks on Industries

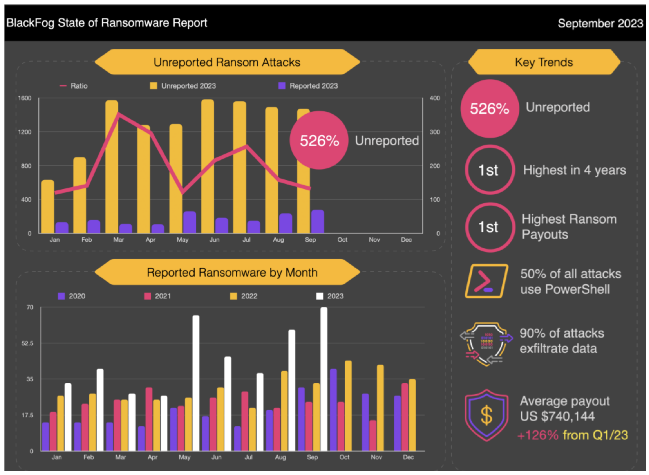

Figure 2. Ransomware Trends as in 2023

B. Evolution

Strong encryption and the emergence of anonymous payment systems like Bitcoin drove a considerable evolution in ransomware since the early 2000s. The decade of the 2010s saw an explosion of ransomware-as-a-service (RaaS) assaults, making it possible for individuals with little technical background to initiate attacks, resulting in massive data breaches and monetary losses. The seriousness of ransomware threats was highlighted by notable outbreaks such as NotPetya and WannaCry in 2017. Double extortion was used by cybercriminals during the late 2010s and early 2020s, when they threatened to release sensitive data unless ransoms were paid. The picture of ransomware became more complicated as laws and regulations tightened. Ransomware

979-8-3503-7809-2/24/$31.00 ©2024 IEEE

strategies changed in spite of these obstacles, focusing on vital infrastructure, supply lines, and sophisticated social engineering techniques, which made defenses constantly adjust.

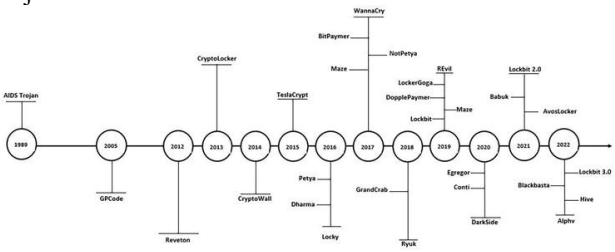

Figure 3. Timeline of Evolution of Ransomware

C. Imperative of Ransomware Detection and Recovery

Because of its adaptability and frequency, ransomware requires a strong protection plan that combines early detection, prevention, and efficient recovery techniques. The first line of security, detection, entails locating ransomware or its antecedents on the network or endpoints of an enterprise. This is a dynamic process that needs to be continuously improved upon to stay up with the changing threats. Many detection strategies are used, from conventional signature-based techniques to more sophisticated ones like machine learning, behavioral monitoring, heuristic analysis, and network traffic analysis. Even though detection is crucial, it is not infallible, therefore in order to recover encrypted or locked data and avoid giving in to ransom demands, additional efforts must be made in the data recovery domain. To reduce data loss, downtime, and financial damage, recovery techniques include backup systems, decryption tools, and incident response procedures. This survey report explores the many methods of detection and recovery, assessing the benefits and drawbacks of each approach while providing information on the most recent developments. The paper provides a thorough manual for researchers, practitioners, and policymakers, highlighting the importance of data recovery and protection in today's digital environment and illuminating the difficulties and state-of-the-art tactics in the continuous fight against ransomware.

D. Attack Mechanism

Cybercriminals are always improving their tactics to evade detection and boost extortion success, which is why ransomware attacks are always changing. Organizations need to take precautionary precautions against these threats, frequently backup their data, and warn consumers about the dangers of opening strange emails.

Usually, compromised downloads, malicious email attachments, or software flaws are how ransomware enters a system. Users are frequently duped into executing the harmful payload through the use of social engineering. After activation, the ransomware takes over, increases privileges, and turns off security measures in order to avoid discovery. After that, it encrypts the victim's data with powerful encryption methods, rendering it unreadable. In order to obtain the decryption key, a ransom message is shown, requesting money (often in cryptocurrency). Paying the ransom encourages more illegal activity, therefore it's risky because there's no assurance you'll get the key.

Victims can get their data back if they are given a decryption key. For future attacks to be avoided, it is imperative to remove the malware and bolster security protocols. It is possible to detect weaknesses and assist in efforts to apprehend the attackers by reporting the incident to law authorities and examining the ransomware strain.

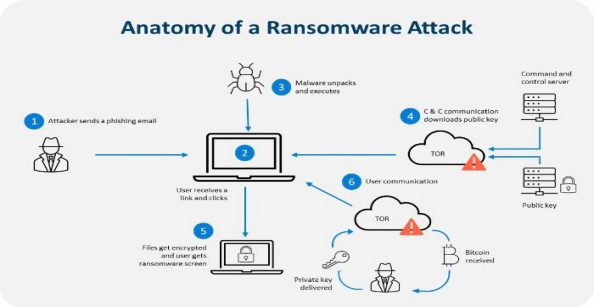

Figure 4. Anatomy of Ransomware Attacks

E. Targets

Ransomware mostly attacks people, companies, healthcare providers, academic institutions, and governmental organizations, concentrating on those with vital functions or important data. As vital infrastructure, such as transportation networks and electrical supplies, come under increasing threat, everyone who uses digital devices has to have strong cybersecurity protections in place.

II. METHODOLOGY

A. Techniques

1) Detection Techniques

a) Signature-based Detection

Signature-based detection uses distinct file characteristics or code snippets that are known ransomware signatures. To find and notify users of identified strains, the system searches documents or network traffic for these signatures.

Advantages:

- Effectiveness: highly effective for known strain identification
- False positives: since it depend upon known signatures, which are accurate, the false positive rate is low.

Limitations:

- Effectiveness: ineffective against unknown strands and needs frequent updates
- Polymorphism: polymorphic samples (that can change signature) can bypass this technique easily

b) Heuristic Detection

Heuristic detection identifies ransomware by analyzing unusual behavior and traits, rather than specific signatures.

Advantages:

- Effective against new and evolving strands
- Can detect zero-day ransomware

Limitation:

- Moderate false positive rate as trigger alerting behavior may be shown
- Source-intensive, impact the performance

c) Behavioural Analysis

Behavioral analysis keeps an eye on software behavior, focusing on system activity and creating a baseline to identify variations as possible hazards.

Advantages:

- Can identify ransomware at early stage
- Can easily detect abnormal or malicious behaviour a characteristic of ransomwares or unknown strands

Limitations:

- May alert false alarms in case of abnormal activity by legitimate software
- Resource intensive in large and complex environments

d) Machine Learning Algorithms

By examining data elements like file attributes and network activity, machine learning algorithms are able to identify ransomware. Models like neural networks—which are trained on past data—monitor in-the-moment activity, spotting possible dangers and notifying security professionals to stop them.

Advantages:

- Can easily adapt to evolving threats as they learn from new data in early stage, before extensive damage
- Decreased false positive rate

Limitations

- Can be resource intensive for deep learning approach
- May struggle with new variants due to no precedent in training

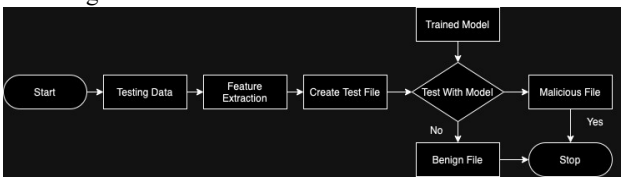

Figure 5. Workflow of the detection mechanism

e) Anomaly Detection

When anomalies like unexpected file encryption or dubious network traffic are detected, they can be used to identify potential security risks.

Advantages:

- Make adaptions to emerging threats, as it doesn't depend on preexisting signatures or patterns, or predefined rules
- Can detect insider threats and effective against new variant

Limitations:

- Need a baseline for comparison, sometimes, this can be difficult
- Sophisticated ransomware can bypass through evasion techniques

2) Prevention Techniques

- Regular Software Updates - Regular updates and patching reduce ransomware risk by closing vulnerabilities, but require ongoing monitoring and a structured patch management process.
- Email Security Measures - Employers should be trained to recognize phishing attempts and to use email security solutions to prevent harmful attachments and phishing emails.
- Employee Training and Awareness - Although employee education can help improve cybersecurity, human error is still a possibility. Train personnel to spot questionable emails, links, and attachments.
- Access Control and Least Privilege Principle - The impact of ransomware can be lessened by enforcing least privilege access, which calls for constant monitoring to prevent unauthorized access and lateral movement.
- Network Segmentation - Network segmentation limits ransomware's ability to propagate, limiting its damage. It makes setup and maintenance more difficult, though.
- Backup and Disaster Recovery - Backup important data to a secure, remote location. Regularly test backups for reliability to enable recovery without paying ransom.
- Endpoint Security Solutions - Sophisticated endpoint security tools use machine learning for proactive ransomware defense but require frequent updates and may produce false positives.
- Whitelisting Applications - Whitelisting blocks unauthorized software, reducing ransomware risk but can be impractical for large environments due to management complexity.
- Network and Intrusion Detection System - Network and intrusion monitoring can detect ransomware by spotting unusual traffic patterns early but may produce false positives and need fine-tuning.

3) Data Recovery Techniques

- Backup and Restore - Frequent offline or secure backups enable data recovery without having to pay a ransom. If backups are current, this strategy is dependable and efficient, although it can be resource-intensive and susceptible to attacks on backup systems.
- Decryption Tools - Security companies' decryption technologies make it possible to unlock files encrypted by particular ransomware strains without having to pay a ransom. They work well for strains that are known to exist, but they are restricted to readily available decryption keys and solutions.
- Incident Response Plans - A well-defined incident response strategy helps organizations manage ransomware attacks by outlining steps for alerting authorities, isolating systems, and containing damage. While it doesn't guarantee data recovery, it helps minimize the attack's impact.
- File Versioning and Shadow Copies - File versioning and shadow copies aid in restoring ransomware-encrypted data, but accessibility is limited and may not work against all strains.
- Ransom Payment - Negotiating with ransomware perpetrators might recover data if a working decryption key is provided, but it is discouraged due to promoting illegal activity, no guarantee of key functionality, and potential legal consequences.

B. Implementation

1) Random Forest

Utilizing randomized training data subsets, Random Forest is a resilient ensemble learning instrument that generates several decision trees and averages their classification or regression predictions. Effective against

intricate and dynamic ransomware patterns, this method reduces overfitting and increases robustness. Random Forest recognizes patterns associated with ransomware, having been trained on a variety of data sources including system logs and network traffic. It can change with the techniques of ransomware because of its versatility. It additionally facilitates the early identification of backup and recovery system breaches. Generally, Random Forest is a useful part of comprehensive ransomware prevention and recovery solutions because of its capacity to deal with high-dimensional data and generate trustworthy models.

2) Gradient Boost

Gradient Boosting, an ensemble learning technique that builds decision trees one after the other to improve prediction accuracy by fixing past mistakes, is used in the ransomware detection project. This method is excellent at finding intricate connections and patterns in datasets, which makes it a good fit for spotting changing ransomware behaviors. Gradient Boosting increases detection precision in dynamic threat situations because of its capacity to handle complex, nonlinear data. In order to ensure a strong protection against ransomware threats, the project focuses on meticulous hyperparameter tuning and striking a balance between model complexity and interpretability.

3) XGBOOST

The ensemble learning approach of XGBoost (eXtreme Gradient Boosting), which combines several weak models like decision trees for increased accuracy, makes it a potent machine learning tool useful in ransomware detection. By creating trees progressively and learning from past mistakes, it effectively manages unbalanced datasets. Because of its regularization, XGBoost is able to identify new ransomware variants and prevent overfitting. It works well with huge datasets because of its processing efficiency and capacity to handle missing data. To further aid in comprehending ransomware activity, XGBoost now offers comprehensible feature importance scores. For best performance, nevertheless, it requires a representative training dataset and careful parameter optimization to be successful.

4) Light GBM

Microsoft's LightGBM is a very effective gradient boosting system that is renowned for its speed, accuracy, and scalability. It selects useful samples using Gradient-based One-Side Sampling (GOSS) and discretizes continuous features using histogram-based discretization, which speeds up training and uses less memory. LightGBM's excellent performance and scalability make it ideal for handling big datasets with hundreds of thousands of samples and features. This allows for effective real-time deployment as well as quick model testing and improvement. LightGBM also excels at ransomware detection.

III. RESULT

To visualize the patterns and anomalies in data behaviour and identification of potential ransomware activity, this heatmap was produced. This heatmap indicates the distribution and intensity of ransomware related activities on the basis of extracted features.

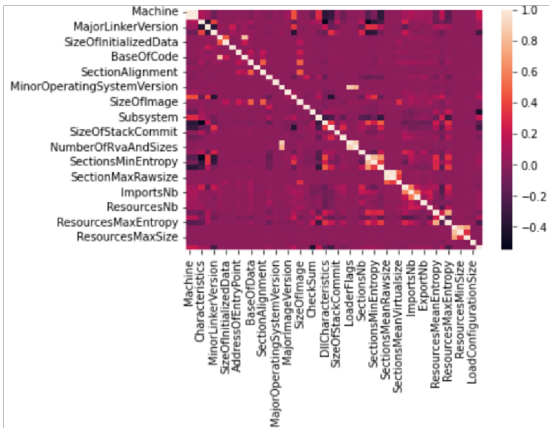

Figure 6. heatmap of Extracted Features

To evaluate the performance of model in classification of ransomware related activity, this confusion matrix is produced. By contrasting the expected labels with the actual labels of the data samples, it shows the categorization results.

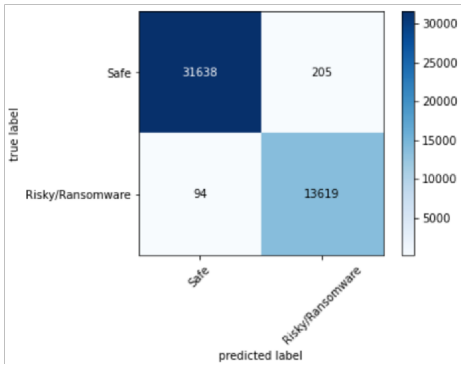

Figure 7. Confusion Matrix

To check the performance of various models on the given dataset, accuracy, Roc and F1 score was calculated for various machine learning models.

100%|

Model				
ExtraTreesClassifier	0.99	0.99	0.99	0.99
RandomForestClassifier	0.99	0.99	0.99	0.99
XGBClassifier	0.99	0.99	0.99	0.99
LGBMClassifier	0.99	0.99	0.99	0.99
BaggingClassifier	0.99	0.99	0.99	0.99
DecisionTreeClassifier	0.99	0.99	0.99	0.99
KNeighborsClassifier	0.99	0.99	0.99	0.99
ExtraTreeClassifier	0.99	0.99	0.99	0.99
AdaBoostClassifier	0.99	0.99	0.99	0.99
SVC	0.99	0.99	0.99	0.99
QuadraticDiscriminantAnalysis	0.98	0.98	0.98	0.98
BernoulliNB	0.97	0.97	0.97	0.97
NuSVC	0.98	0.97	0.97	0.98
LinearDiscriminantAnalysis	0.97	0.96	0.96	0.97
RidgeClassifier	0.97	0.96	0.96	0.97
RidgeClassifierCV	0.97	0.96	0.96	0.97
SGDClassifier	0.96	0.96	0.96	0.96
LogisticRegression	0.96	0.96	0.96	0.96
LinearSVC	0.96	0.96	0.96	0.96
CalibratedClassifierCV	0.96	0.96	0.96	0.96
NearestCentroid	0.94	0.94	0.94	0.94
PassiveAggressiveClassifier	0.92	0.93	0.93	0.92
Perceptron	0.89	0.91	0.91	0.89
GaussianNB	0.64	0.74	0.74	0.65
DummyClassifier	0.50	0.49	0.49	0.52

Figure 8. Accuracy of various models tested

Against the models performance, time taken for each model to run on the data set was noted.

Model	Time Taken
ExtraTreesClassifier	5.17
RandomForestClassifier	8.57
XGBClassifier	4.19
LGBMClassifier	0.62
BaggingClassifier	3.75
DecisionTreeClassifier	0.70
KNeighborsClassifier	18.53
ExtraTreeClassifier	0.23
AdaBoostClassifier	3.99
SVC	36.33
QuadraticDiscriminantAnalysis	0.31
BernoulliNB	0.18
NuSVC	959.23
LinearDiscriminantAnalysis	0.39
RidgeClassifier	0.31
RidgeClassifierCV	0.31
SGDClassifier	0.34
LogisticRegression	0.90
LinearSVC	17.85
CalibratedClassifierCV	56.62
NearestCentroid	0.18
PassiveAggressiveClassifier	0.25
Perceptron	0.28
GaussianNB	0.18
DummyClassifier	0.17

Figure 9. Time taken by various models to run

The trade-off between the true positive rate and false positive rate is evaluated by the ROC curve. It shows the model's performance and ability to discriminate across different threshold values.

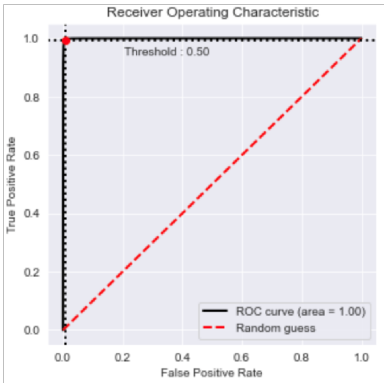

Figure 10. ROC Curve

As depicted by the following result, it can be observed that the true positive rate is high and false positive rate is low, indicating low error rate. This means that the model is quite accurate in detecting the ransomware and benign files.

Figure 11. Prediction Probability

IV. CONCLUSION

This study explores several techniques for detecting ransomware and recovering encrypted data, pointing out the advantages and disadvantages of each strategy. Although signature-based detection is low in false positives and effective against known threats, it necessitates constant updating. Although they can adapt to changing dangers, heuristic and behavioral assessments run the risk of raising false alarms and depleting resources. Although it requires consistent improvement and high-quality data, machine learning has promise. Regular backups are essential for data recovery, but they are also susceptible to attacks, and decryption software is limited to specific strains. Plans for handling events and minimizing damage are essential.

The study highlights the fact that no single approach can provide complete security; rather, a mix of strong recovery plans and detection measures improves security. To combat the ever-evolving ransomware menace, one must constantly innovate and adapt. The results emphasize how crucial proactive readiness and dynamic, multifaceted defenses are to cybersecurity.

REFERENCES

[1] S. Mohurle and M. Patil, "A brief study of wannacry threat: Ransomware attack 2017," International journal of advanced research in computer science, 2017.

[2] G. O'Gorman and G. McDonald, Ransomware: A growing menace, Symantec Corporation Arizona, AZ, USA, 2012.

[3] P. O'Kane, S. Sezer and D. Carlin, "Evolution of ransomware," Iet Network, 2018.

[4] R. Richardson and M. M. North, "Ransomware: Evolution, mitigation and prevention," International Management Review, 2017.

[5] O. Fedor, "93 Must-Know Ransomware Statistics [2023]," November 2022. [Online]. Available:

[6] https://www.antivirusguide.com/cybersecurity/ransomwarestatistics/?gclid=Cj0KCQjw0tKiBhC6ARIsAAOXutlxtBECRXkilm8YXWiEZ2VqqqPr77wXivK4o6NANBS5rNUWB4Z mRjIaAu7tEALw_wcB.

[7] M. Anghel and A. Racautanu, "A note on different types of ransomware attacks," Cryptology ePrint Archive, 2019.

[8] J. P. Tailor and A. D. Patel, "A comprehensive survey: ransomware attacks prevention, monitoring and damage control," Int. J. Res. Sci. Innov, 2017.

[9] H. Oz, A. Aris, A. Levi and A. S. Uluagac, "A survey on ransomware: Evolution, taxonomy, and defense solutions," ACM Computing Surveys (CSUR), 2022.

[10] I. Alexandru-Vasile, "Ransomware data recovery techniques," Cryptology ePrint Archive, 2023.

[11] S. Vasoya, K. Bhavsar and N. Patel, "Preventing ransomware attacks is crucial for individuals and organizations to safeguard their valuable data and protect against potential financial and reputational damage. Implementing effective preventive measures can significantly reduce the risk of fa," arXiv preprint arXiv:2212.04063, 2022.

[12] Z. Livingston, "Main Targets of Ransomware Attacks & What They Look For," December 2022. [Online]. Available: https://www.esecurityplanet.com/threats/what-ransomware-attackers-look-for/.

[13] A. Al Qartah, Evolving Ransomware Attacks on Healthcare Providers, 2020.

[14] U. Urooj, B. A. S. Al-rimy, A. Zainal, F. A. Ghaleb and M. A. Rassam, "Ransomware detection using the dynamic analysis and machine learning: A survey and research directions," Applied Sciences, 2021.

[15] B. M. Khammas, "Ransomware detection using random forest technique," ICT Express, vol. 6, 2020.

[16] S. Kok, A. Abdullah, N. Jhanjhi and M. Supramaniam, "Ransomware, threat and detection techniques: A review," Int. J. Comput. Sci. Netw. Secur, vol. 19, p. 136, 2019.

[17] B. Kenyon and J. McCafferty, "Ransomware recovery," Itnow, 2016.

[18] H. Alshaikh, N. Ramadan and H. A. Hefny, "Ransomware prevention and mitigation techniques," Int. J. Comput. Appl, 2020.

[19] A. H. Mohammad and others, "Ransomware evolution, growth and recommendation for detection," Modern applied science, 2020.

[20] S. Aurangzeb, M. Aleem, M. A. Iqbal, M. A. Islam and others, "Ransomware: a survey and trends," Journal of Information Assurance \& Security, 2017.

[21] C. Beaman, A. Barkworth, T. D. Akande, S. Hakak and M. K. Khan, "Ransomware: Recent advances, analysis, challenges and future research directions," Computers \& security, vol. 111, p. 102490, 2021.

[22] S. Young, "When ransomware strikes, what's your recovery plan?," Network Security, 2021.

[23] Ö. A. Aslan and R. Samet, "A Comprehensive Review on Malware Detection," IEEE Access, 2020.

[24] K. Shaukat, S. Luo and V. Varadharajan, "A novel deep learning-based approach for malware detection," Engineering Applications of Artificial Intelligence, 2023.

[25] E. Berrueta, D. Morato, E. Maga{\~n}a and M. Izal, "A survey on detection techniques for cryptographic ransomware," Berrueta, Eduardo and Morato, Daniel and Maga{\~n}a, Eduardo and Izal, Mikel, 2019.

[26] J. A. Herrera Silva, L. I. Barona LOpez, A. L. Valdivieso Caraguay and M. HernAndez-Alvarez, "A survey on situational awareness of ransomware attacks—detection and prevention parameters," Remote Sensing, 2019.

[27] M. S. Kumar, J. Ben-Othman and K. Srinivasagan, "An investigation on wannacry ransomware and its detection," in 2018 IEEE Symposium on Computers and Communications (ISCC), 2018.

[28] F. Noorbehbahani, F. Rasouli and M. Saberi, "Analysis of machine learning techniques for ransomware detection," in 2019 16th International ISC (Iranian Society of Cryptology) Conference on Information Security and Cryptology (ISCISC), 2019.

[29] N. Scaife, H. Carter, P. Traynor and K. R. Butler, "Cryptolock (and drop it): stopping ransomware attacks on user data," 2016.

[30] I. A. Chesti, M. Humayun, N. U. Sama and N. Jhanjhi, "Evolution, mitigation, and prevention of ransomware," in 2020 2nd International Conference on Computer and Information Sciences (ICCIS), 2020.

[31] M. Humayun, N. Jhanjhi, A. Alsayat and V. Ponnusamy, "Internet of things and ransomware: Evolution, mitigation and prevention," Egyptian Informatics Journal, 2021.

[32] F. A. Aboaoja, A. Zainal, F. A. Ghaleb, B. A. S. Al-rimy, T. A. E. Eisa and A. A. H. Elnour, "Malware detection issues, challenges, and future directions: A survey," Applied Sciences, 2022.

[33] P. Bajpai and R. Enbody, "Memory forensics against ransomware," in 2020 International Conference on Cyber Security and Protection of Digital Services (Cyber Security), 2020.

[34] D. Hitaj, G. Pagnotta, F. De Gaspari, L. De Carli and L. V. Mancini, "Minerva: A File-Based Ransomware Detector," arXiv preprint arXiv:2301.11050, 2023.

[35] A. Al-Sabaawi and T. Alrowidhan, "Ransomware Detection and Data Recovery (Case Study)," 2023.

[36] S. I. Bae, G. B. Lee and E. G. Im, "Ransomware detection using machine learning algorithms," Concurrency and Computation: Practice and Experience, vol. 32, p. e5422, 2020.

[37] A. Kapoor, A. Gupta, R. Gupta, S. Tanwar, G. Sharma and I. E. Davidson, "Ransomware detection, avoidance, and mitigation scheme: a review and future directions," Sustainability, 2021.

[38] A. Maurya, N. Kumar, A. Agrawal and R. A. Khan, "Ransomware: evolution, target and safety measures," International Journal of Computer Sciences and Engineering, 2018.

[39] S. Baek, Y. Jung, D. Mohaisen, S. Lee and D. Nyang, "SSD-Assisted Ransomware Detection and Data Recovery Techniques," 2021.

[40] I. Kara and M. Aydos, "The rise of ransomware: Forensic analysis for windows based ransomware attacks," Expert Systems with Applications, 2022.

[41] S. Baek, J. Jeon, B. Jeong and Y.-S. Jeong, "Two-stage hybrid malware detection using deep learning," Humancentric Computing and Information Sciences, 2021.