IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, July 2024

Cyber Threat Hunters : Machine Learning Survey for Security

Swarangi P. Saraikar, Shivani A. Dhomase, Dr. Sharmila S. More
MIT ACSC Alandi(D), Pune, Maharashtra
swarangiprashantsaraikar@mitacsc.edu.in, shivanianildhomase@mitacsc.edu.in, ssmore@mitacsc.ac.in

Abstract: *Machine learning (ML) has become available across various sectors, revolutionizing fields like healthcare, finance, and cybersecurity. It is not affected by its large potential, ML models are susceptible to diverse security threats. This survey delves into the intricate landscape of ML security, providing a comprehensive overview of the current state-of-the-art. We begin by outlining various attack vectors, including data poisoning, adversarial attacks, and model inversion, which can compromise the integrity and functionality of ML systems. Subsequently, we explore established and emerging defence mechanisms designed to mitigate these vulnerabilities. The survey further analyses the challenges and limitations are making more difficult for the development of strong and secure ML models. Finally, we discuss promising research directions and open problems that warrant further investigation to ensure the secure and trustworthy deployment of ML in security-critical applications.*

Keywords: Machine Learning, Security, Adversarial Attacks, Data Poisoning, Model Inversion, Defence Mechanisms, Security Challenges

I. INTRODUCTION

The remarkable growth of machine learning (ML) has transformed numerous domains, fostering advancements in areas like healthcare, finance, and, ironically, cybersecurity. ML algorithms excel at extracting patterns from vast datasets and making data-driven predictions, enabling them to tackle complex security challenges. However, the very foundation of ML, its reliance on data, introduces unforeseen vulnerabilities.

This survey paper delves into the crucial, yet often overlooked, aspect of security in machine learning. We navigate the intricate landscape of ML security, providing a comprehensive overview of potential threats, established defences, and remaining challenges. Machine learning approaches are also playing a vital role in improving the efficiency of detection and prevention techniques against threats to mobile devices [11].

The integrity and security of a computer system are compromised when an illegal penetration, unauthorized individual or program enters a computer or network intending to harm or disrupt the normal flow of activities [1]. To begin, we establish a foundational understanding of the various attack vectors that exploit inherent vulnerabilities within ML systems. These harmful tactics, including data poisoning, adversarial attacks, and model inversion, have the potential to manipulate, disrupt, and compromise the integrity of deployed ML models.

Following this exploration of threats, we shift focus towards the defence mechanisms employed to mitigate them. In this mechanism data should be confidential, integral, and available [2]. We discuss established and emerging techniques designed to safeguard the security of ML systems, fostering trust and reliability in their performance. However, ensuring robust security in ML is not without its challenges and limitations. We delve into these limitations, highlighting the complexities and open problems that hinder the development of foolproof ML security solutions.

Finally, we conclude the introductory section by presenting promising research directions that hold the potential to strengthen the security landscape of ML. We aim to ignite further exploration and collaborative efforts in this crucial field, paving the way for the secure and trustworthy deployment of ML, particularly in security-critical applications.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, July 2024

Objectives of Research:

Machine learning (ML) is a double-edged sword for security. While it offers powerful tools to detect threats and secure systems, ML models themselves are vulnerable to attacks. Cyber security is to protect the integrity of the data, networks, and programs from cyber threats to cyberspace [10].

Understanding Threats:

1) Classify different types of attacks on ML models, such as data poisoning, adversarial examples, and models and extractions.
2) Analyse the impact of these attacks on real-world security applications

Defensive Techniques:

- Evaluate existing methods for securing ML models, including data sanitization, adversarial training, and detection algorithms.
- Identify promising research directions for developing more robust defences

Standardization and Best Practices:

- Recommend best practices for secure development of ML-based security systems.
- Explore the need for standardization in areas like secure model sharing and evaluation.

Emerging Areas:

- Investigate the security implications of new ML techniques, such as federated learning and explainable AI
- Analyse the security challenges posed by the growing adoption of ML in Internet-of-Things (IoT) environments.

II. LITERATURE REVIEW:

Start by exploring how ML is currently used for cybersecurity tasks like intrusion detection, malware analysis, and anomaly detection. Reference relevant studies showcasing the effectiveness of different algorithms (e.g., Support Vector Machines, Random Forests) in specific security applications. Discuss the vulnerabilities of ML models to adversarial attacks Slike data poisoning and manipulation. Cite research papers that analyse the impact of these attacks on real-world security systems.

Review existing methods for securing ML models. This could include techniques like data sanitization, adversarial training, and detection algorithms used to identify and mitigate attacks. Analyse the limitations of current defence mechanisms. Highlight areas where further research is needed, such as developing more robust defences against evolving attack techniques. Discuss the importance of establishing best practices for secure development of ML-based security systems. Reference existing recommendations or highlight the need for new guidelines.

Explore the potential benefits of standardization in areas like secure model sharing and evaluation.

Additional Considerations:

- Include recent research on the security implications of emerging ML techniques, such as federated learning and explainable AI.
- Discuss the unique security challenges posed by the growing adoption of ML in Internet-of-Things (IoT) environments.

Machine Learning Techniques:

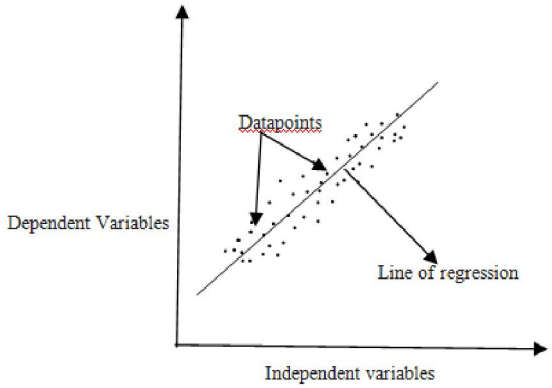
Machine learning (ML) is a field of computer science that gives computers the ability to learn without being explicitly programmed. This is achieved by training algorithms on data, which allows the algorithms to identify patterns and make predictions on new data. There are many different machine learning techniques, which can be broadly categorized into three main types: supervised learning, unsupervised learning, and reinforcement learning.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, July 2024

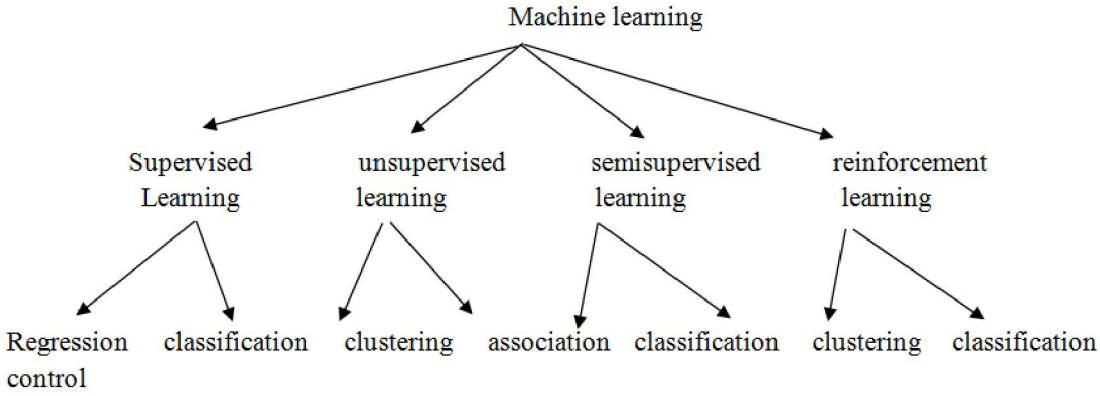
Supervised Learning

Supervised learning is a type of machine learning technique that involves training a model using labelled data. Labelled data consists of input data and its corresponding desired output. The model learns the relationship between the input and output data, and can then be used to make predictions on new, unseen data. Here are some common supervised learning techniques:

Regression: Regression is a technique for predicting continuous values, such as housing prices or stock prices.



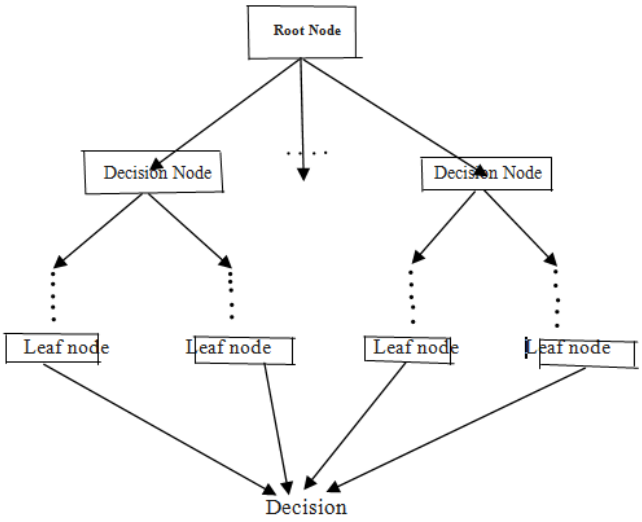
Classification: Classification is a technique for predicting discrete values, such as whether an email is spam or not.



Decision trees: Decision trees are a type of model that uses a tree-like structure to make decisions. They are often used for classification problems.

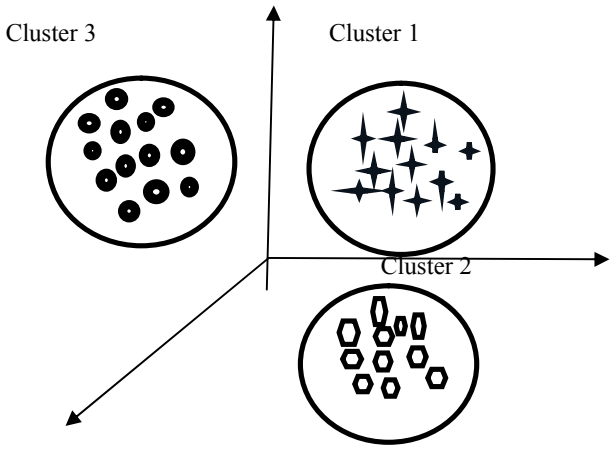
IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, July 2024

Unsupervised Learning



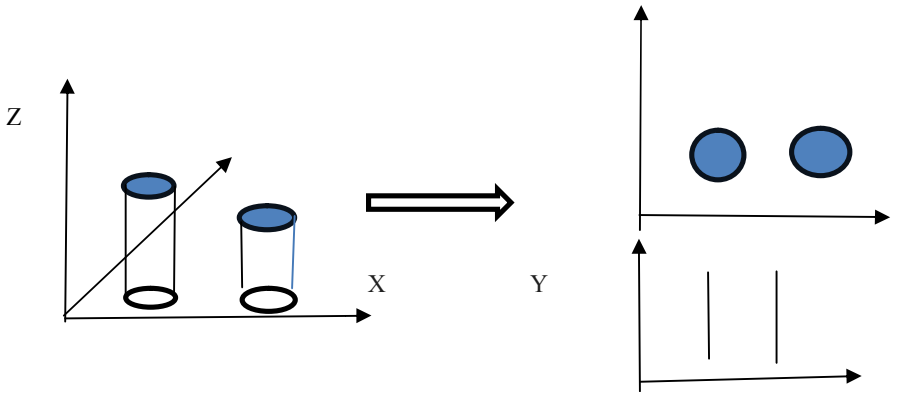
Unsupervised learning is a type of machine learning technique that involves training a model on unlabelled data. Unlabelled data does not have any corresponding labels or desired outputs. The model learns to identify patterns and relationships in the data on its own. Here are some common unsupervised learning techniques:

Clustering: Clustering is a technique for grouping data points together based on their similarities.



Dimensionality reduction: Dimensionality reduction is a technique for reducing the number of features in a dataset. This can be helpful for improving the performance of machine learning models.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, July 2024

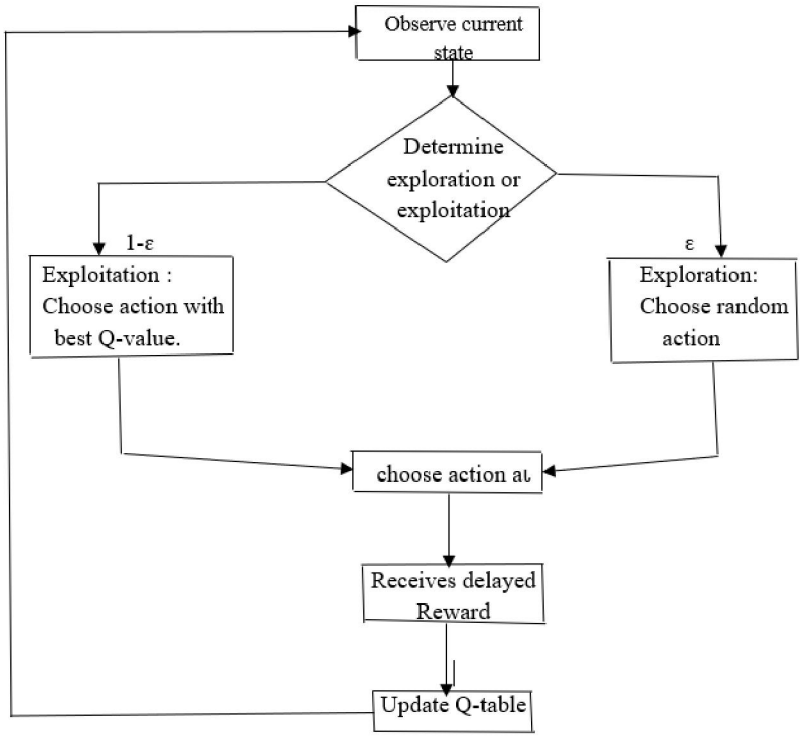


Principal component analysis (PCA): PCA is a common dimensionality reduction technique that is used to identify the most important features in a dataset.

Reinforcement learning

Reinforcement Learning is a type of machine learning technique that involves training a model through trial and error. The model interacts with an environment and receives rewards or penalties for its actions. The model learns to take actions that maximize its rewards. Reinforcement learning is often used for problems where it is difficult or expensive to obtain labelled data. Here are some common reinforcement learning techniques:

Q-learning: Q-learning is a reinforcement learning technique that is used to learn the value of taking different actions in different states.



IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, July 2024

Policy gradient methods

Policy gradient methods are a type of reinforcement learning technique that is used to directly learn a policy for taking actions.

These are just a few of the many different machine learning techniques that are available.

Application of machine learning in cyber security

Machine learning (ML) plays a crucial role in various aspects of cybersecurity, offering powerful tools to combat evolving threats and enhance security posture. Here are some key applications of ML in cybersecurity:

1. Threat Detection and Prevention:

Anomaly detection: ML algorithms can analyse network traffic, user behaviour, and system logs to identify unusual patterns that deviate from normal activity, potentially indicating malicious activity.

Malware classification: ML models can be trained to identify malicious software based on various features, such as code structure, network behaviour, and file properties. This helps in effectively blocking malware before it can infect systems[3]-[6].

Phishing detection: ML can analyse email content, sender information, and website characteristics to identify fraudulent emails attempting to steal sensitive information[7]-[9].

2. Security Automation and Efficiency:

Vulnerability scanning and prioritization: ML can automate vulnerability scanning processes and prioritize vulnerabilities based on their potential risk and exploitability, allowing security teams to focus on the most critical issues first.

Security incident and event management (SIEM): ML can be integrated with SIEM systems to analyse security events and logs, correlate them with threat intelligence, and prioritize incidents for faster response.

User and entity behaviour analytics (UEBA): ML can analyse user behaviour patterns to identify anomalous activities that might indicate compromised accounts or insider threats.

III. ADVANCED THREAT HUNTING AND INVESTIGATION

- **Proactive threat hunting:** ML can be used to analyse vast amounts of security data to uncover hidden threats and potential attack campaigns that traditional security tools might miss.
- **Threat intelligence gathering and analysis:** ML can automate the process of collecting and analysing threat intelligence from various sources, providing valuable insights into attacker tactics and motivations.
- **Incident investigation and forensics:** ML can assist in forensic investigations by analysing log data and identifying the root cause of security incidents, facilitating faster remediation.

IV. IDENTITY AND ACCESS MANAGEMENT (IAM)

- **Risk-based authentication:** ML can be used to assess the risk associated with a login attempt, enabling adaptive authentication measures based on the user's context and potential risk level.
- **Anomaly detection in user behaviour:** ML can analyse user access patterns and identify deviations from normal behaviour, potentially indicating compromised accounts or unauthorized access attempts.

Overall, ML offers a powerful toolbox for enhancing cybersecurity by automating tasks, improving threat detection and prevention, and enabling proactive security measures. However, it is crucial to remember that ML models are only as effective as the data they are trained on, and constant monitoring and improvement are essential to maintain their effectiveness against evolving threats.

Challenges and Limitations

Machine Learning can deliver a variety of benefits, including the ability to quickly find and respond to threats and better leverage insights from data analytics. However, it can also present challenges.

Data-related challenges:

Data quality: ML algorithms are heavilydepends on the quality and quantity of data they are trained on. In cybersecurity, data can be vast, diverse, and often riddled with inconsistencies or inaccuracies. This can lead to models making flawed decisions or failing to detect threats altogether.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, July 2024

Data privacy: Collecting and using vast amounts of data for security purposes raises concerns about user privacy. Balancing effective security measures with ethical data collection practices is crucial.

Model-related challenges:

Interpretability and explainability: Complex ML models can be like black boxes, making it difficult to understand how they arrive at their decisions. This lack of transparency can hinder trust in their effectiveness and make it challenging to identify and rectify errors.

Adversarial attacks: Malicious actors can exploit vulnerabilities in ML models to manipulate their behaviour. This could involve feeding the model poisoned data to trigger false positives or negatives, compromising the system's integrity.

Other significant limitations:

Evolving threats: Cybercriminals are constantly devising new attack methods. While ML models can adapt to some extent, staying ahead of the curve remains a challenge.

Resource requirements: Training and running complex ML models often demands significant computational resources, which can be a barrier for smaller organizations.

Despite these challenges, research in improving the robustness, transparency, and efficiency of ML models for cybersecurity is ongoing. By acknowledging these limitations and working towards solutions, organizations can leverage the power of ML to enhance their security posture while mitigating associated risks.

V. FUTURE DIRECTIONS AND RESEARCH OPPORTUNITIES

Evolving Attack Landscape:

Analyse how attackers are adapting their techniques to exploit vulnerabilities in new ML architectures, like transformers and generative models.

Discuss the need for continuous monitoring and adaptation of defence mechanisms to stay ahead of the adversarial curve.

Explainable AI and Security:

Explore how advancements in explainable AI can be leveraged to understand and mitigate security risks in ML models.

Investigate techniques for building inherently secure ML models by design, incorporating security considerations into the training process.

Privacy-Preserving Machine Learning:

Discuss the challenges of securing ML models while preserving the privacy of sensitive data used for training.

Explore the potential of federated learning and homomorphic encryption for secure training on decentralized data sources.

Human-in-the-Loop Security Systems:

Analyse how human expertise can be integrated with ML-based security systems for improved decision-making and attack response.

Investigate methods for building trust and transparency in ML-powered security solutions for human users.

Security for Resource-Constrained Environments:

Explore the development of lightweight and efficient defence mechanisms for securing ML models on devices with limited computational power, particularly relevant for IoT security.

Discuss trade-offs between security and performance when deploying ML models on resource-constrained devices.

Standardization and Certification:

Analyse the ongoing efforts to establish standards and certifications for secure development and deployment of ML models.

Discuss the challenges and opportunities for creating a robust and effective regulatory framework for ML security.

Interdisciplinary Research:

Explore the potential for collaboration between machine learning, security, and other disciplines like psychology and economics to address the complex challenges of ML security.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, July 2024

Investigate the human factors involved in ML security breaches and design training programs to raise awareness and improve security practices.

Federated learning:

Securely training ML models on distributed datasets without compromising privacy is a critical challenge with vast potential for security applications.

Adversarial machine learning:

Research on both developing more robust models and creating more sophisticated attacks will continue to be an arms race.

These are just a few examples, and the specific research opportunities you choose will depend on your specific areas of interest within the broader field of machine learning security.

VI. CONCLUSION

In conclusion, machine learning presents a double-edged sword for security. While it offers powerful tools for threat detection and prevention, it also introduces new vulnerabilities. By acknowledging these challenges and actively researching advancements in secure ML practices, we can harness the full potential of machine learning to create a safer digital landscape.

REFERENCES

[1]. V. Ambalavanan, "Cyber threats detection and mitigation using machine learning" in Handbook of Research on Machine and Deep Learning Applications for Cyber Security, Hershey, PA, USA:IGI Global, pp. 132-149, 2020.

[2]. T. Thomas, A. P. Vijayaraghavan and S. Emmanuel, "Machine learning and cybersecurity" in Machine Learning Approaches in Cyber Security Analytics, Singapore: Springer, pp. 37-47, 2020, [online] Available: https://link.springer.com/chapter/10.1007/978-981-15-1706-8_3.

[3]. Z. Ma, H. Ge, Y. Liu, M. Zhao and J. Ma, "A combination method for Android malware detection based on control flow graphs and machine learning algorithms", *IEEE Access*, vol. 7, pp. 21235-21245, 2019.

[4]. S. Saad, W. Briguglio and H. Elmiligi, "The curious case of machine learning in malware detection", *arXiv:1905.07573*, 2019, [online] Available: http://arxiv.org/abs/1905.07573.

[5]. P. Jain, "Machine learning versus deep learning for malware detection", pp. 704, 2019, [online] Available: https://scholarworks.sjsu.edu/etd_projects/704/.

[6]. D. Sahoo, C. Liu and S. C. H. Hoi, "Malicious URL detection using machine learning: A survey", *arXiv:1701.07179*, 2017, [online] Available: http://arxiv.org/abs/1701.07179.

[7]. R. S. Rao and A. R. Pais, "Detection of phishing Websites using an efficient feature-based machine learning framework", *Neural Comput. Appl.*, vol. 31, no. 8, pp. 3851-3873, Aug. 2019.

[8]. O. K. Sahingoz, E. Buber, O. Demir and B. Diri, "Machine learning based phishing detection from URLs", *Expert Syst. Appl.*, vol. 117, pp. 345-357, Mar. 2019.

[9]. M. Alauthman, A. Almomani, M. Alweshah, W. Omoushd and K. Alieyane, "Machine learning for phishing detection and mitigation" in Machine Learning for Computer and Cyber Security: Principle Algorithms and Practices, New York, NY, USA:CRC Press, pp. 26, 2019.

[10]. D. Craigen, N. Diakun-Thibault and R. Purse, "Defining cybersecurity", *Technol. Innov. Manage. Rev.*, vol. 4, no. 10, pp. 13-21, Oct. 2014.

[11]. B. Arslan, S. Gunduz, and S. Sagiroglu, ‘‘A review on mobile threats and machine learning based detection approaches,’’ in Proc. 4th Int. Symp. Digit. Forensic Secur. (ISDFS), Apr. 2016, pp. 7–13.