DIGITAL FORENSICS AND ARTIFICIAL INTELLIGENCE FOR CYBER SECURITY

R. Sri Devi [1], M.Mohankumar[2]

ABSTRACT

We have the benefit of technological advancements and digitalization and also security threats that come along with them. Therefore, security of information in the network is a vital element today. Many researchers suggest different security mechanisms such as models, steganography and cryptography algorithms for a secure way of access information in cyberspace. But,there are unethical and malicious activities in the network infrastructure that humans cannot identify. The Digital Forensics and Artificial Intelligence (AI) are gifts to cyber security offering asignificant solution to hackers' malicious activities, and AI also makes stronger information security by security tools based ondeep learning and machine learning. The aim of this article is to survey various algorithms for deep learning and machine learning approaches.

Keyword: Artificial Intelligence, Cyber Security, Deep Learning, Digital Forensics,Machine Learning, Semi supervised, Supervised, Unsupervised

I. INTRODUCTION

In the aspect of cyber security, there are constant threats,and securing the information is a very difficult task. The main aim of Infosec is to assure the business against malware and offer securityby providing availability, integrity, confidentiality and non-repudiation. As per Internet Security Threat Report, 2017, MongoDBis an open source database programshowing increased severity of the ransomware attacks [18].

Application is a machine-learning-algorithm that allows machines to learn automatically and surpass from experience. To identify dangerous risks in network AI gives accurate result in predicting malware [17]. Structured data can be processed by machine-learning-algorithms,whichare used for malware, detection Intrusion Detection System/ Intrusion Prevention System (IDS/IPS), spam detection, phishing, hardware security, cloud security and IoT security.Deep learning depends on multilayers of Artificial Neural Networks (ANN). Deep neural network is used to solve complex data with mathematical model, which imitates a network form consisting of hidden layers with a large number of neuron layers [7].

l Digital Forensic Investigations (DFIs) are the foremost step in the determination of security threats. In 1980, law enforcement departments began to setup expert groups to handle cyber crimes cases [19]. Digital forensics solution is provided to civil,criminal, corporate andmilitary investigations [14].

Machine learning and deep learning algorithms areexplained in section III, digital forensics steps andtools in section IV andconclusion given in the last section.

II. MACHINE LEARNING AND DEEP LEARNING ALGORITHMS

Machine learning can be categorized as semi-supervised, unsupervised or descriptive, supervised or predictive, reinforcement learning, transduction and learning to learn.

[1]*Department of Computer Science, Karpagam Academy of Higher Education,* Coimbatore-21, Tamilnadu, India

[2]*Department of CS,CA& IT, Karpagam Academy of Higher Education,* Coimbatore-21, Tamilnadu, India

Semi-supervi sed learning is a combination of supervised and unsupervised learning for processing unlabeled and labeled data, for example Constrained clustering. Unsupervised learning has the facts without the preferred output, for examples Clustering and dimensionality reduction. Supervised learning is used to identify and response to cause reasonable predictions to set of classes [2]. Examples: Regression and classification. Reinforcement learning has vital role in environment that gives feedback for further learning process [8] [16].

Transduction predicts new input or output according to new facts. Learning to learn method learns from own inductive prejudice on older experience [8].Further, it can be classified from the above types of learning.

Classification is a task of separating things into different groups and using themto identify different classes of network attacks such as scanning and IP spoofing [4].Clustering is similar to classification, but grouping things by similarity groups, it is used for forensic analysis [2]. Association Rule Learning (ARL) is an unsupervised data mining approach andobserves a correlation among variables in a range of data. Regression is a task of predicting the next value based on the previous values and is used to predict the network-packet parameters and compare them with the normal ones.Generative model is a task of creating something based on previous knowledge of the distribution.Dimension reduction or Dimensionality reduction is ageneralization process of searching common variables and lower the value of random variables [2].

TABLE1

Machine learning and Deep Learning Algorithms for Supervised, Unsupervised and Semi supervised Approach

Machine Learning				Deep Learning				
Regression	Classification	Clustering	Dimensionality Reduction	Regression	Classification	Clustering	Generative Models	Association Rule Learning
Linear Regression	Logistic Regression	K- means	Principal Component Analysis (PCA)	Artificial Neural Network (ANN)	Artificial Neural Network (ANN)	Self Organized Maps (SOM) or Kohonen Networks	Markov Chains	Deep Restricted Boltzmann Machine (RBM)
Polynomial Regression	K- Nearest Machine (K-NM)	Mixture model (LDA)	Singular Value Decomposition (SVD)	Recurrent Neural Network (RNN) [1]	Convolution Neural Network (CNN)	-	Genetic algorithm	Deep Belief Network (DBN)
[2] Decision trees- ID3, ID4,ID5, ID5R, C4.5 algorithm [2]	Support Vector Machine (SVM) [2]	Bayesian	Linear Discriminate Analysis (LDA)	Long Short Term Memory (LSTM) [1]	-	-	-	Stacked Auto Encoder
SVR (Support Vector Regression)	Kernel SVM	Gaussian Mixture Model	Latent Semantic Analysis (LSA)	Generative Adaptive Neural Network (GANN)	-	-	-	-
Random Forest	NaviceBayes [2]	Agglomerative	Factor Analysis (FA)	Echo State Networks (ESN) [1]	-	-	-	-
Linear regression	Decision Tree Classification	Mean - shift	Independent Component Analysis (ICA)	-	-	-	-	-
Polynomial regression	Random Forest Classification	-	Non – negative Matrix Factorization (NMF)	-	-	-	-	-

Learning algorithms can be used for security task using static and dynamic analysis.The machine learning approaches are practiced for network traffic scanning, process, application, user-intrusion-detection and endpoint. The following outlinesthe basic learning algorithms used in this survey. It is clear that supervised learning approaches are commonly used algorithms for identifying threats. Semi-supervised learning methodsare inexpensive and less time-consuming [3].

III. DIGITAL FORENSICSTEPS AND TOOLS

Digital forensics is the process of finding and processing electronic data. The goal of the study is to preserve evidence by collecton, identification, interpretation, validation, documentation, preservation and presentation.The following steps are used in digital forensics for implementing a structured investigation by seizure, acquisition, analysis and reporting for the purpose ofrestoring past events.Seizure.

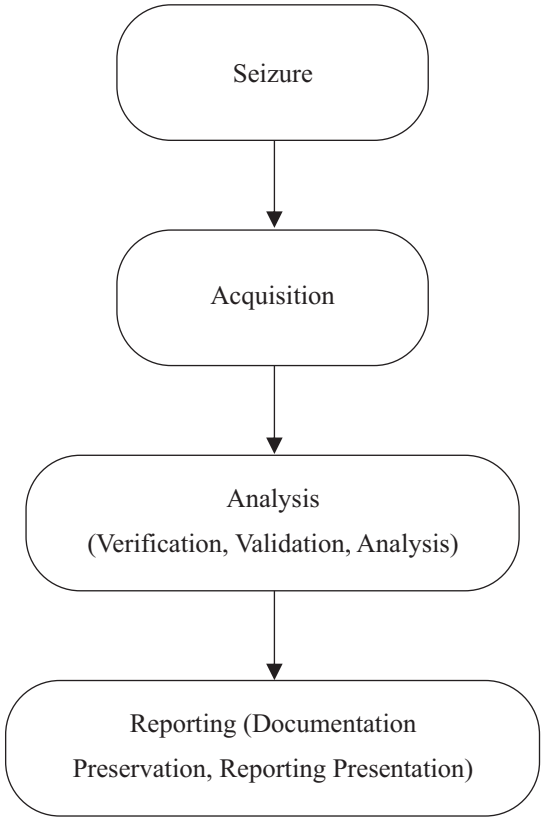

Fig 1. Steps in Digital Forensics

A. Tools for Digital Forensics

The need of digital forensic investigation tools is to correctly present all facts that accumulate during the computer crime activities; it is proof of evidence by the investigator.Translation and presentation are the tools used for analysis. Intelligent analysis method is applied for investigating offline malicious network occurrence and intrusion data information.

TABLE 2

Tools for Digital Forensics Investigation

Tools	Descriptions
Autopsy	worked in The Sleuth Kit, GUI, to flag relevant sections of data.
Volatility	Analysis RAM
Oxygen forensics	Digital forensics application, access insights faster and critical data
Bulk extractor tool	scans a file, disk image, or a directory and extracts vital information
Redline	Memory forensic tool
Computer Aided Investigative Environment (CAINE)	Software tools,GUI modules
Xplico	Cloud analysis tool
Parrot	GUI, Java to investigate CAN traffic, recording, GPS tracking
Digital Evidence & Forensics Toolkit (DEFT)/ Deft OS	Mobile forensics tool

IV. CONCLUSION

There are various cyber security threats and data breaches happening, and digital information has more risks in today's digital world. Digital forensics and learning algorithm consist of valuable tools for cyber security for detecting malicious software. From the survey, we understand that Information security can be strengthening by digital forensics using AI applications. This is the primary stage for doing research to a great extent.

REFERENCES

[1] M. Sewak, S. K. Sahay and H. Rathore, "Comparison of Deep Learning and the Classical Machine Learning Algorithm for the Malware Detection," *2018 19th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, Busan, 2018, pp. 293-296.doi: 10.1109/SNPD.2018.8441123

[2] Kwon, D., Kim, H., Kim, J., Suh, S. C., Kim, I., & Kim, K. J. (2017). A survey of deep learning-based network anomaly detection. *Cluster Computing*, 1-13.

[3] G. Chandra and S. K. Dwivedi, "A Literature Survey on Various Approaches of Word Sense Disambiguation," *2014 2nd International Symposium on Computational and Business Intelligence*, New Delhi, 2014, pp. 106-109. doi: 10.1109/ISCBI.2014.30

[4] G. Dayanandam, E. S. Reddy and D. B. Babu, "Regression algorithms for efficient detection and prediction of DDoS attacks," *2017 3rd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT)*, Tumkur, 2017, pp. 215-219. doi: 10.1109/ICATCCT. 2017.8389136

[5] Yaniv Harel, Irad Ben Gal, and Yuval Elovici. 2017. Cyber Security and the Role of Intelligent Systems in Addressing its Challenges. ACM Trans. Intell. Syst. Technol. 8, 4, Article 49 (May 2017), 12 pages. DOI: https://doi.org/10.1145/3057729

[6] ChafikaBenzaïd, AbderrahmanBoulgheraif, Fatma Zohra Dahmane, Ameer Al-Nemrat, and Khaled Zeraoulia. 2016. Intelligent detection of MAC spoofing attack in 802.11 network. In Proceedings of the 17th International Conference on Distributed Computing and Networking (ICDCN '16). ACM, New York, NY, USA, Article 47, 5 pages. DOI: https://doi.org/10.1145/2833312.2850446

[7] Tobiyama, S., Yamaguchi, Y., Shimada, H., Ikuse, T., & Yagi, T. (2016, June). Malware detection with deep neural network using process behavior. In *2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC)* (Vol. 2, pp. 577-582). IEEE.

[8] Ayodele, T. O. (2010). Types of machine learning algorithms. In *New advances in machine learning*. IntechOpen.

[9] Baig, Z. A., Szewczyk, P., Valli, C., Rabadia, P., Hannay, P., Chernyshev, M., ... & Syed, N. (2017). Future challenges for smart cities: Cyber-security and digital forensics. *Digital Investigation*, *22*, 3-13.

[10] Sarno, R., Dewandono, R. D., Ahmad, T., Naufal, M. F., &Sinaga, F. (2015). Hybrid Association Rule Learning and Process Mining for Fraud Detection. *IAENG International Journal of Computer Science*, *42*(2). [11]Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks*, *61*, 85-117.

[12] R. Montasari and R. Hill, "Next-Generation Digital Forensics: Challenges and Future Paradigms," *2019 IEEE 12th International Conference on Global Security, Safety and Sustainability (ICGS3)*, London, United Kingdom, 2019, pp. 205-212.

[13] Carrier, B. (2002, August). Defining digital forensic examination and analysis tools. In *Digital Research Workshop II*.

[14] Beebe, N. (2009, January). Digital forensic research: The good, the bad and the unaddressed. In *IFIP International Conference on Digital Forensics* (pp. 17-36). Springer, Berlin, Heidelberg.

[15] Casey, E. (2011). *Digital evidence and computer crime: Forensic science, computers, and the internet*. Academic press.

[16] Xiaojin Zhu; Andrew Goldberg, "Introduction to Semi-Supervised Learning," , Morgan & Claypool, 2009

[17] AnamitraDeshmukh , " Artificial Intelligence," in Technical Publications,2016

[18] https://www.symantec.com/content/dam/symantec/docs/reports/istr-24-2019-en.pdf

[19] https://en.wikipedia.org/wiki/Digital_forensics