A Light-weight Watermarking-Based Framework on Dataset Using Deep Learning Algorithms

Muhammad Tayyab
School of Computer Science and Engineering (SCE)
Taylor's university Lake-side Campus, 47500
Subang Jaya, Malaysia
muhammadtayyab@sd.taylors.edu.my

Mohsen Marjani
School of Computer Science and Engineering (SCE)
Taylor's university Lake-side Campus, 47500
Subang Jaya, Malaysia
mohsen.marjani@taylors.edu.my

N. Z. Jhanjhi
School of Computer Science and Engineering (SCE)
Taylor's university Lake-side Campus, 47500
Subang Jaya, Malaysia
noorzaman.jhanjhi@taylors.edu.my

Ibrahim Abakr Targio Hashem
College of Computing and Informatics, Department of Computer Science
University of Sharjah, 27272
Sharjah, UAE
ihashem@sharjah.ac.ae

2021 National Computing Colleges Conference (NCCC) | 978-1-7281-6719-0/20/$31.00 ©2021 IEEE | DOI: 10.1109/NCCC49330.2021.9428845

***Abstract*—In most decision-based security applications Deep Learning (DL) algorithms have been widely using for improvement. For better performance, a large amount of dataset has been used for training the DL algorithms. As DL has been remained a key element in the performance of the application, hence, several privacy and security issues have reported, which have affected the performance. Such security attacks have also affected the performance by taking the advantage of the huge dataset, because it is easy for an attacker to add executable noise into the dataset to get the information of the dataset and the model used. Most common security attacks like poisoning and evasion attacks have been considered challenging attacks that have caused misclassification and wrong prediction. Hence, a secure metric is needed to mitigate the effects of such attacks from the dataset. Therefore, in this paper, a light-weight watermarking framework has been proposed that provides security to the dataset before training the DL algorithms. We have implemented our proposed framework using the most common Convolutional Neural Network (CNN) and Artificial Neural Network (ANN) against security attacks. The proposed framework has been evaluated based on accuracy, precision, and computational cost, and has maintained the accuracy up to 98.89% and a precision of 0.96, which has maintained the level as in recent literature. We have also reduced the computational cost for the proposed framework. We believed that the proposed framework can be used to mitigate the security issues in DL algorithms and enhanced toward other security applications.**

Keywords— Deep Learning (DL) Convolutional Neural Network (CNN), Artificial Neural Network (ANN), Poisoning Attacks, Evasion Attacks

I. INTRODUCTION

Most of the modern applications which rely on numerous datasets for a decision have been using Deep Learning (DL) algorithms for improvement. It is well known that DL has shown remarkable performance results while solving decision-based problems in real-world scenarios. The most popular areas of DL applications areas: Internet of Things (IoT) [1], smart cities [2, 3], Modern education systems [4], surveillance models [5], vulnerability and malware detection [6], drones jets [7], robotics and voice-controlled devices [8]. Such application areas have a large number of datasets that are used to automate the process by applying the DL models for better performance and high accuracy. More importantly, DL has also gained confidence in solving issues in security applications. In modern security-sensitive applications and health, models have been modified and improved by DL algorithms[9]. Detection of spam and malicious emails [10], fraud detection [11], malicious intrusion detection [12] are the sensitive models, that have been enhanced and improved by DL algorithms. Moreover, DL has introduced many techniques to not only improve the performance of the security-sensitive application but also has reduced human involvement in both supervised and non-supervised learning like Convolutional Neural Network (CNN), linear regression [13] decision trees [14], and Naive Bayes [15].

With all these innovations and features introduced by the DL algorithms, many security issues have been addressed, that have affected the performance. Poisoning and evasion attacks are considered the most exciting security attacks that are caused by adding noise into the dataset. An attacker can apply simple noise into the dataset that is used to train the model and can get valuable information. These attacks are considered active attacks in DL algorithms which can mislead the classification of the model towards attacker motive [16]. While in exploratory attacks commonly known as evasion attacks, some "Good words" are added into the spam emails, to avoid the detection system and labeled as desirable email. This can create a loop whole in the email box and an attacker can get valuable information from the user's account. Similarly, DL algorithms have played a very important role in classification and predicting models in various environments [17]. Datasets used for DL algorithms can be collected from several untrustworthy sources. Although, it is considered that DL can work honestly in any decision-based environment, however, and attackers have the primary motive to get useful information from the dataset and model by adding some malicious executable noise. This is referred to as poisoning of the dataset which causes misclassification and wrong prediction [18]. Hence, it is needed a secure framework for the dataset that can not only provide security but also maintain the integrity of the user's data.

Although in previous studies, there had been proposed security techniques based on Homomorphic Encryption (HE) schemes, which has provided security to the model and dataset, the computational cost was very high like Faster CryptoNets, CryptoNets [19], CryptoDL-1 [20]. etc., Moreover, some additional information was also added into the dataset, while