**NTCC REPORT-2020**

AMITY SCHOOL OF ENGINEERING AND TECHNOLOGY



|  |  |
| --- | --- |
| **SUBMITTED BY** **:** | **ARYAN RAWAT** |
| **UNDER THE GUIDANCE OF :** | **Dr. AKASH PUNHANI** |
| **ENROLLMENT NO :** | **A2305219448** |
| **SEMESTER** **:** | **2** |
| **BATCH** **:** | **2019-2023** |
| **TOPIC** **:** | **DEEP LEARNING IN CYBERCRIMES** |
| **FIELD :** | **MACHINE LEARNING** |

**CERTIFICATE**

This is to certify that the project report entitled, “Deep Learning in Cybercrimes” submitted to **Amity School of engineering and technology, Uttar Pradesh, India** in fulfillment of the requirement for the award of the degree of B.Tech. in computer science, embodies original work carried out by **ARYAN RAWAT**. This work is original and has not been submitted so far in part or full for any other degree or diploma of any other university.

Director

**Dr. S.K. Shukla**

Professor & Director,

AIFS, Noida

Signature of the mentor

**Dr. Akash Punhani**

ASET, Noida

**ACKNOWLEDGEMENT**

I would like to express my special thanks of gratitude to Dr. Akash Punhani, my mentor for helping me in choosing this topic. It was under his guidance and support that I was able to patch this project and make it full proof success. His suggestions and his instructions have served as the major contributors to achieve completion in this project.

Then I would like to thank my parents and friends who have helped me with their valuable suggestions and guidance that has been helpful in various phases of the completion of this project.

Last but not the least I would like to thank my classmates who have helped me a lot.

**INDEX**

1. **TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **S.NO.** | **TOPIC** | **PAGE NO.** |
|  | CERTIFICATE | 2 |
|  | ACKNOWLEDGEMENT | 3 |
|  | ABSTRACT | 5 |
|  | Introduction | 6 |
|  | Types of cybercrimes in brief | 8 |
|  | Deep learning related solutions to the cybercrimes | 10 |
|  | **3.1** Android malware detection guided by deep learning | 10 |
|  | **3.2** deep learning method for prediction of DDoS attacks | 11 |
|  | **3.3** deep learning in detecting phishing emails | 12 |
|  | **3.4** spam filtering with deep learning | 13 |
| **4.** | conclusion | 15 |
| **5.** | REFERENCES | 16 |

**ABSTRACT**

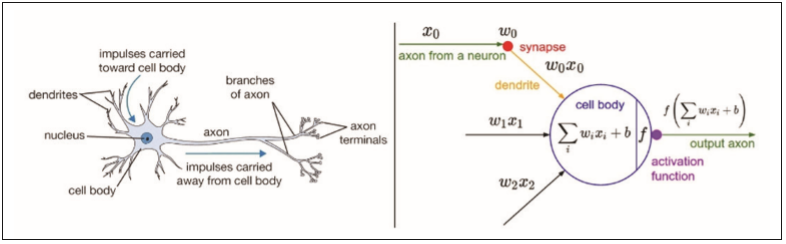
Technology in cybercrimes there is also development in technology countering these crimes in the recent year technology has developed several folds. This advancement has led to a considerable increase in cyber related crimes. Humans has developed a more effective and accurate way to solve these problems known as deep learning which is basically solving problems precisely with machine brain. In this paper we discuss majorly four types of cyber-attacks that are android malware detection. Various DDOS attacks in which the server is flooded by bad traffic and it make it impossible for real clients to approach the server, which leads to large amount of money loss for example DDoS attacks on credit card companies, email phishing, spams messages on social media. Here we will discuss some cyber-attacks and their counter which has used deep learning.

1. **INTRODUCTION**

Due to the digitization whole data was forced to be turned into digital form which increased the data breach and stealing of data. According to a census there are 3.5 billion smartphone users in the world which is almost half of the total population and these smartphones are most vulnerable to be attacked by hackers. CISCO blocked 7 trillion cyberattacks and potential threats that's 20 billion attacks each day. Including invasion of privacy, wire threats, money laundering, false identity enrollments of domain names, espionage, theft of highly classified treaties, and other crimes — which put a huge burden of money loss on governments and individuals scaling to around tens of billions dollars.

There is a large scope for attacks as everything around us is approachable through internet and the point that users download applications and download files from non-official stores and sites. Windows in laptops, androids in mobile phones which also support smart TVs and iOS in apple. They can easily be inflicted and corrupted and can be used to get personal info and credit card information. It is impossible to tackle each one of these problems manually so this need is completed by artificial intelligence [1].

Deep learning is the architecture which is based on links of complex artificial neural networks. The machine brain helps us to maneuver over these problems collectively and precisely. An artificial neuron is made in an easy way, from a mathematical equation, it resembles the neuron from biology which is the elementary unit of human brain and this artificial neuron is the smallest unit of an artificial neural network [1] [5]. See fig.1.



**Figure 1.** Artificial neuron [5]

In the previous time, deep learning has been applied effectively in many fields. For example, deep learning can be used in conversion machines to mend dependability, in reference systems to propose what clients are attracted in or want to get later, and in photo detection systems to distinguish objects [2].

1. **Types of cyber-crimes in brief**

* Phishing-this is defined as the process of obtaining card information or login information using false ID of a genuine source. Phishing attack includes hackers sending malicious emails and URL to victims to access their data. The hackers are becoming more cautious and they furbish and mend their messages in such a way that they are not listed as spams. Victims are forced to either change their information or their important data is stolen. Phishing is done by numerous other types such as vishing which is voice phishing that includes calling to a victim while giving him false identity. These fake individuals maybe representing a bank or organization and will force you to tell them your card details or personal information. Once you follow their commands your security is compromised. So as a result we should answer every call with care and presence of mind and should not ever give personal address or data. Various organizations have guidelines and lists of precautions regarding these fake calls [14].
* Denial of service (DoS) attack it’s an attempt by the attackers to block services provided by the server. It involves overflowing servers with artificial requests generated by malicious computers so that the input of requests becomes more then the server can accept. This leads to overload of server and disruptions in working of server. As a result of this attack the online services provided by the server becomes scarce and the traffic assortment fails. Huge corrupted machineries known as the botnets are made to efficiently perform this attack. The hackers at the point when server is down try to break in through malicious malwares. There is also a sub type of DoS attacks i.e. DDoS or distributed denial of service attack where more than one source are involved in flooding the server. To stop the working of server hackers choose targets that contain large database on servers and have huge user base, prominent sites such as bank and charge cards are primary targets. Major organizations such as flip cart, Facebook, twitter are major targets for these kinds of attacks because they have a large data base.
* Malware-they are malicious software that have the tendency to attach themselves to files and can easily transfer from one computer system to another through network services. These malware disrupts PC procedures and have an influence on the info by modifying it or by deleting it completely without permission. They can easily uninstall an important software and applications from the system which is highly undesirable. Malwares inflicts spyware and adware which can be prevented by using appropriate anti- viruses. Ones sub type of malware is Trojan horses which has quite a different type of propagation. They are generally listed or attached on genuine records such as emails which are shown authentic and from legit sources. The user can also download Trojans by mistake from drive downloads or URLs on malicious sites without even realizing it. A Trojan can disrupt working of a system by many means such as stealing, deleting or tampering the data.
* Spams- spams are messages intended as promotions, business prices, lucky draws or large amount of cash. They are basically tricks made by the influencers which forces the victim to give away personal information or when the URLs are used they steal data. These types of spamming attacks are hard to stop because botnets which spread these kinds of mails are made especially to dodge the filter by spam segregating channels. Spamming is message bombarding. Large amounts of messages are sent to countless people without any pattern. Opening these links may lead to a malicious site or a promotion page or advertisement which basically wastes times and can be irritating. Spammers gather email information from records of service providing companies, news sites, databases of online blogs, location books and these information are sold to other spammers. These spams are often directed from illegal email addresses [14].

1. **DEEP LEARNING RELATED SOLUTIONS TO THE CYBER CRIMES** 
   1. **ANDROID MALWARE DETECTION GUIDED BY DEEP LEARNING**

Malware-it’s a malicious software that’s solemnly designed to inflict the damage to server, computer or a client. There are a variety of malicious malwares such as virus, Trojans horses, worms, adware etc.

Android operating system attracts the malware manufacturers, as every normal personal keeps their most important and private information in mobile phones, such as their contacts, social accounts, emails, and bank accounts details. Existing malware analysis can be divided in two kinds namely dynamic and static. In static analysis binaries and main source code are checked in order to detect unrecognized patterns or malicious patterns whereas dynamic analysis is behavioral kind of analysis in this malicious software is separated in separate environment while examining its change in behavior. The static method used was time consuming and takes too much and on the same side dynamic method for malware detection was centered to identify malware based on uneven battery consumption [1][2].

So a deep learning method was created to scan malwares on the basis of network load, battery usage, memory and computer chip consumption which is several folds effective. The basic approach is via two deep learning architectures namely encoder decoder and LSTM [1][2].

1. The encoder decoder architecture is based on a neural network which has an encoder (which gives input or state and encodes in simpler form) and a decoder (which decodes the input or state and give output).see fig. 2

Figure 2. The encoder- decoder model

1. The next method is LSTM which stands for long-short term memory. They are basically an artificial system designed to identify designs in arrangements of data, such as mathematical time’s series data, stock markets and government activities (but also counting writing, genomes, writing and the vocal term)

Both of these architectures are very effective deep learning approach and perform well on platform [1].

* 1. **DEEP LEARNING METHOD FOR PREDICTION OF DDoS ATTACKS**

AI has proved to be very useful for solving and detecting problems. Health illnesses and credit card theft are two examples of usage for this technology. One of the toughest complications to solve, in cyber-security, is battling distributed denial of service attack (DDoS).

DDOS stands for distributed denial of service. The DDOS attacks are generally flooding a server with countless of signals from a fake address or network, it generally looks like hundreds of infected computers shooting signals. So after this the victim server faces difficulties coping with so much traffic which is generally fake traffic which means that they are not real requests by clients. Server then becomes unable to sort the good or real requests and bad traffic or fake requests by various computers. The task to write codes to distinguish bad and good traffic is really hard as the fake traffic is made in such a way that it almost fully resembles real traffic. Deep learning helps to detect the DDoS attacks early so that some part of traffic can be channeled away and the main server can be saved [3].

An example of how harmful a DDoS attack can be for a business is best defined in the situation that a large bank gets confronted. During the attack, the bank will shut off their servers and prevent this malicious circulation from incoming their private data centers. During this down-time, no customer of the bank will be able to make dealings or use any of the set’s facilities. In over-all, banks can lose around $5 million per 60 minutes that their facilities are unavailable (counting their reputational harm, which based on the extent of the bank can be a lot bad) [3].

Various algorithms are made using deep learning solemnly to detect these DDOS attacks earlier by the cyber security teams of major companies. Machine intelligence in general is very efficient at finding different patterns within data patches compared as when picked manually. In neural systems, we start off by making neurons and transfer random masses to the nets (they aren’t frequently mentioned to as synapses in AI). By serving in labeled data, the neural network learns what the ideal masses are as long as they are given the correct labels. As the archetypal reads in more and more labeled data, it learns the accurate masses better and better to the point where ultimately, individual can provide in information and the AI will be able to select correct label [3][4].

* 1. **DEEP LEARNING IN DETECTING PHISHING EMAILS**

Email communication has become a necessary part of our life. They have become very important part of financial communication in business all over the world with this there is a significant rise in phishing emails which are emails specifically designed to steal personal information and financial records without our consent. The scary part of phishing emails is that they can successfully steal our information without us even noticing [6][7][9].

The most common way to get trapped by these phishing emails by entering any fraud site. Recently there are three ways to counter the phishing emails which are as follows-

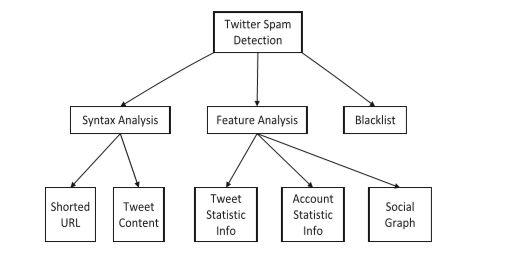
1. The most common way is to spread public awareness and not entering personal details everywhere.
2. The second most effective and commonly accepted is by blacklisting all fraud websites spreading phishing emails but the major drawback is that the site will generally be detected way after spreading phishing emails.
3. The third and most effective of all ways is terminating a detecting phishing emails early using deep learning algorithms. This works easily as the machine brain predicts if a page or website is fraud based on data from already busted real phishing emails from sites [6][7]. The problems with deep learning approach is that-

* It need a skillful individual trained in deep learning specifically.
* The time taken to train machine brain to detect site and writing specific points to detect phishing sites is considerably long.
* It is hard to set up the criteria or pin point characteristics of fraud phishing emails.

Presently phishing is divided in two parts the email phishing and website phishing. Most phishing emails contain a URL (uniform resource locator). When the user click on the URL lead straight to the malicious site [11][8]. To write the criteria to distinguish the phishing website from others it’s important to include some major points and write its code such as-

* The site should have relevant material.
* Exclusion and inclusion of certain specific characteristics.
* Search relevant documents.
  1. **SPAM FILTERING WITH DEEP LEARNING**

Spam email or message is generally commercial in nature it focuses on trapping a large amount of people rather than targeting it to an individual like in phishing emails. Spam tends to be less private whereas phishing email may already have a person’s name in them. Spams maybe legit sites or companies trying to sell their products or increase their reviews. Spam messages usually don’t try to be fake and do not force you to click on malicious links [12].

Recently there is tremendous rise in social media spam messages which are generally on twitter. The researchers and cyber security teams has till now used many machine learning methods such **Figure 3.** Spam detection categories [12]

as blacklisting spam activities on social media. According to investigation these techniques are about 87% accurate because of difficulties such as spam drift and information fabrication.

As the blacklist method cannot be continued because of variations or mutations in spamming activities and inspecting each link by hand is very time consuming so deep learning plays a major role here too. The composition of every tweet is enrolled and trained by deep learning. A binary classifier is paired with deep learning architecture so that spams can be distinguished from regular tweets. This is called syntax analysis. Another technique to tackle this problem is to use feature analysis which takes take accounts features such as age of user account as fake accounts are usually new, the number of following, number of followers. To efficiently use those features some modern mining methods were established. The two main problems with this method is that spam drift affects the results of a trained classifier. The second problem is that data features are available on twitter and can be obtained by statistical methods [12][13]. The data fabrication is also one of the major problems. See fig. 3.

1. **CONCLUSION**

Deep learning is necessary for us as it saves us from major online cyber related crimes all over the world. It runs the data through many stages and each stage passes the information to next in a simplified form than previous one. Deep learning algorithms are also self-dependent that means they learn to do a task and they can filter or select right or wrong on their own. It can scan through a large amount of dataset easily which is as accurate as handpicking while using a very less amount of time so it’s an ideal replacement of manual labor. Deep learning neural network can be efficiently applied to large dataset for knowledge, predictions and applications. Basically deep learning is itself a sub part of machine learning and the time taken to train deep learning algorithms maybe much longer even days and months but once completed it works fluently and efficiently. Currently most of the businesses depends on deep learning like media, journalism, movie industry, social media platforms and manufacturing industries. Overall deep learning in machine learning provides us artificial neurons which can help us make artificial or machine brains and these brains are capable of doing large tasks for us in every field and it mostly surpasses human precision in many fields. Deep learning is the bright future in AI for upcoming generations.

1. **REFERENCES**
2. Milosevic, N., & Huang, J. (2019). Deep learning guided Android malware and anomaly detection. *arXiv preprint arXiv:1910.10660*.
3. Nguyen, K. K., Hoang, D. T., Niyato, D., Wang, P., Nguyen, D., & Dutkiewicz, E. (2018, April). Cyberattack detection in mobile cloud computing: A deep learning approach. In *2018 IEEE Wireless Communications and Networking Conference (WCNC)* (pp. 1-6). IEEE.
4. Alguliyev, R. M., Aliguliyev, R. M., & Abdullayeva, F. J. (2019). Deep Learning Method for Prediction of DDoS Attacks on Social Media. *Advances in Data Science and Adaptive Analysis*, *11*(01n02), 1950002.
5. Mittal, S., Das, P. K., Mulwad, V., Joshi, A. and Finin, T. (2016). CyberTwitter: Using Twitter to generate alerts for cybersecurity threats and vulnerabilities. Proc. IEEE/ACM Int. Conf. Advances in Social Networks Analysis and Mining. IEEE Press, Piscataway, pp. 860–867
6. VREJOIU, M. H. (2019). Neural Networks and Deep Learning in Cyber Security. *Romanian Cyber Security Journal*, *1*(1), 69-86.
7. Hiransha, M., Unnithan, N. A., Vinayakumar, R., Soman, K., & Verma, A. D. R. (2018, March). Deep learning based phishing e-mail detection. In *Proc. 1st AntiPhishing Shared Pilot 4th ACM Int. Workshop Secur. Privacy Anal.(IWSPA)*. Tempe, AZ, USA.
8. Benavides, E., Fuertes, W., Sanchez, S., & Sanchez, M. (2020). Classification of Phishing Attack Solutions by Employing Deep Learning Techniques: A Systematic Literature Review. In *Developments and Advances in Defense and Security* (pp. 51-64). Springer, Singapore.
9. Hajgude, J., & Ragha, L. (2012, October). Phish mail guard: Phishing mail detection technique by using textual and URL analysis. In *2012 World congress on information and communication technologies* (pp. 297-302). IEEE.
10. Marchal, S., Armano, G., Gröndahl, T., Saari, K., Singh, N., & Asokan, N. (2017). Off-the-hook: An efficient and usable client-side phishing prevention application. *IEEE Transactions on Computers*, *66*(10), 1717-1733.
11. Selvaganapathy, S., Nivaashini, M., & Natarajan, H. (2018). Deep belief network based detection and categorization of malicious URLs. *Information Security Journal: A Global Perspective*, *27*(3), 145-161.
12. Deng, L. (2014). A tutorial survey of architectures, algorithms, and applications for deep learning. *APSIPA Transactions on Signal and Information Processing*, *3*.
13. Wu, T., Wen, S., Liu, S., Zhang, J., Xiang, Y., Alrubaian, M., & Hassan, M. M. (2017). Detecting spamming activities in twitter based on deep‐learning technique. *Concurrency and Computation: Practice and Experience*, *29*(19), e4209.

[13] Benevenuto, F., Magno, G., Rodrigues, T., Almeida, V., & Detecting Spammers on Twitter," Collaboration. (2010). electronic messaging, anti-abuse and spam conference (CEAS).

[14] Buch, R., Ganda, D., Kalola, P., & Borad, N. (2017). World of Cyber Security and Cybercrime.

# Aryan Rawat

## ORIGINALITY REPORT

# 2% 1% 1% 1%

SIMILARITY INDEX INTERNET SOURCES PUBLICATIONS STUDENT PAPERS

|  |  |
| --- | --- |
| PRIMARY SOURCES |  |
| research.ijcaonline.org 1  Internet Source | <1% |
| Submitted to University of Westminster  2  Student Paper | <1% |
| Rasim M. Alguliyev, Ramiz M. Aliguliyev,  3  Fargana J. Abdullayeva. "Deep Learning  Method for Prediction of DDoS Attacks on  Social Media", Advances in Data Science and  Adaptive Analysis, 2019  Publication | <1% |
| Submitted to Study Group Australia  4  Student Paper | <1% |
| onlinelibrary.wiley.com 5  Internet Source | <1% |
| techie-buzz.com 6  Internet Source | <1% |

Exclude quotes On Exclude matches < 8 words

Exclude bibliography On