International Journal of Advanced Research in Computer and Information Security
Vol.3, No.1 (2019), pp.1-8
http://dx.doi.org/10.21742/ijacis.2019.3.1.01

Applications of Machine Learning in Cyber Security

N. Thirupathi Rao[1] and Debnath Bhattacharyya[2]

Dept. of Computer Science and EngineeringVignan's Institute of Information Technology (A)Visakhapatnam 530049, AP, India
[1]nakkathiru@gmail.com, [2]debnathb@gmail.com

Abstract

Machine learning is one of the latest trends models and methods that can be implemented and the users can get good results. The growth of this area was developing day to day by almost all areas of applications and research. The other important area to be considered was the cyber security. Most of the common people also being trapped by these technologies and the people are losing their valuable data in some cases and in some other cases the people or losing their valuable money also. In some other cases, the people are losing their lives due to the sharing of their personal data to the public domain. Hence, people need to think of these problems in a different manner and also to solve these technical problems by using these advanced technologies. In the current article, some of the cyber security issues and threats that were being occurring in these days were highlighted and how the utilization of these machine learning techniques will be used to identify such threats and can be avoided or can be protected from these sorts of attacks.

Keywords: *Machine learning, Algorithms, Attacks, Cyber security, Shells, Networks, Components, Hubs, Switches, Routers, Servers*

1. Introduction

Artificial Intelligence and machine learning is the two thrust areas that were being developed in a fast and very innovative way in these days. Machine learning is one of the application areas of artificial intelligence and it is the branch of the field of computer science. This branch of machine learning deals with the working of the robots, machines and automation of all types of applications based on some data that was being provided by the developers [1][2]. The machines will react and proceed on working for the above said data being supplied by the users and also the machines will consider such data and they will get training from that previous data and act on various situations based on the previous actions that were performed for solving such similar type of problems. The main goal in the working of this machine learning and its related applications are the development of the algorithms such that these algorithms can able to make the decisions based on the supplied previous data. The data might be used for various applications. The utilization of human intervention or human presence can be reduced day to day by the proper development and proper utilization of the resources. Machine learning is mainly focused on the usage of data analytics and data analysis.

By the utilization of these techniques and methods, the human intervention and human involvement is limited very much. The techniques and the methods that we are using today in

Article history:
Received (July 20, 2019), Review Result (August 27, 2019), Accepted (October 6, 2019)

Print ISSN: 2208-1658, eISSN: 2208-1666 IJACIS

machine learning models are not the same those methods were developed and proposed earlier days of the machine learning. Basically these methods were developed for the pattern recognition and for other models. But today it had developed further to the extent that the machines are learning based on the data and acting based on the suggestions or the decisions that were already taken in the previous models. The methods and the techniques that we are using today in machine learning are some of the techniques that already existed in the modeling based on predictions and the data mining [3]. The two methods will follow the similar type of methods that can be used for the method to find the data or the methods to find the required data. Most of the famous applications we are using today in our lives are the shopping interests of the people based on the searches that they are being performed while browsing for some items in the market or in the social marketing websites [4]. Some of the other important tasks that can be performed by using these sorts of applications are like the detection of fake money, current notes identification, personalized marketing, fraud detection, spam filters, security to the networks, detection of threats and updating the news feeds etc.

Figure 1. Machine learning example [2]

The working of the machine learning and its related applications can be processed in various forms and various methods. The algorithms that were in the group of machine learning algorithms are classified into two types like supervised algorithms and unsupervised algorithms. In the case of supervised learning algorithms, the scientists or the experts based on the applications are present in the process of implementation of the applications and time to time submission of input data corrections to the existing supplied data are performed such that to make the implementation process more simple and easy. Whereas in the other case, the unsupervised learning includes the methods of learning in which the intervention of the presence of human are not required for the submission of data or to correct the data that was already being started submission to the system. In the case of these unsupervised learning models, the data training or the data submission was not much required due to the reason that they use another method called the deep learning with which the data and application processing can be done and completed [5][6][7].

Some of the application areas of machine learning are the business intelligence, business intelligence models, customer relationship models, human resource systems, human resource management, virtual assistant systems, self-driving cars, virtual assistants, cyber security etc. Nowadays, the utilization of machine learning techniques in cyber security applications is also increased a lot. In order to analyze the models, several types of factors and applications are to be monitored and read under several situations. In the following sections, a detailed analysis on the factors and applications to be considered for the machine learning applications and how they will help in the successful completion of the functioning of such algorithms with such methods are explained in detail as follows.

2. Cyber security threats that machine learning can protect

In the current section, a detailed analysis about the various threats and problems to be discussed in detail and type of threats that were being come to the users and by using the machine learning techniques how the users can avoid or can protect their systems and machines from being corrupted or being hacked or being disturbed the smooth functioning of the systems.

They are,

- Watering hole
- Web shell
- Spear phishing
- Remote Exploitation
- Ransom ware

3. Watering hole

This watering hole attack is one of the important most considered attacks where the attackers will target a group of users from an organization or group of users working in a company or an organization. The basic thing that attackers will implement in the current methods are that they will try to observe the list or group of people who are visiting a particular website and based on the list of customers, the attackers will target them and the data of those users can be stolen and used for some other purposes [6]. The attackers may not attack all the group of people in the list at a time, the attackers will attack them selectively based on some search criteria. The users or the attackers will check the security protection that was provided to that website will be checked and then the attack will be planned. The current concept was implemented based on the old methods or the statements that the animals whoever want to kill or take their prey, they will create a small pond of water or a situation and they will wait at various small locations which cannot be identified like an ambush and whenever the animals will come for water, they will attack and kill those animals and will take them as prey for them.

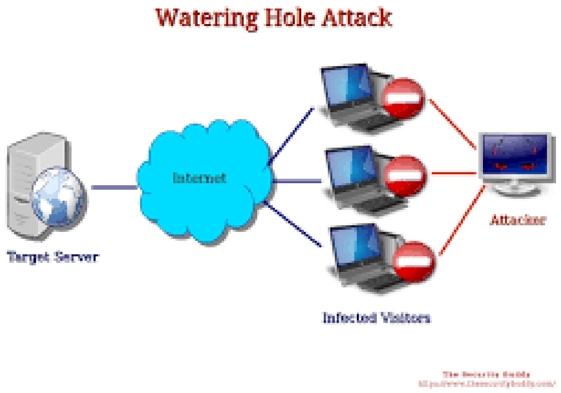


Figure 2. Watering hole attack model [8]

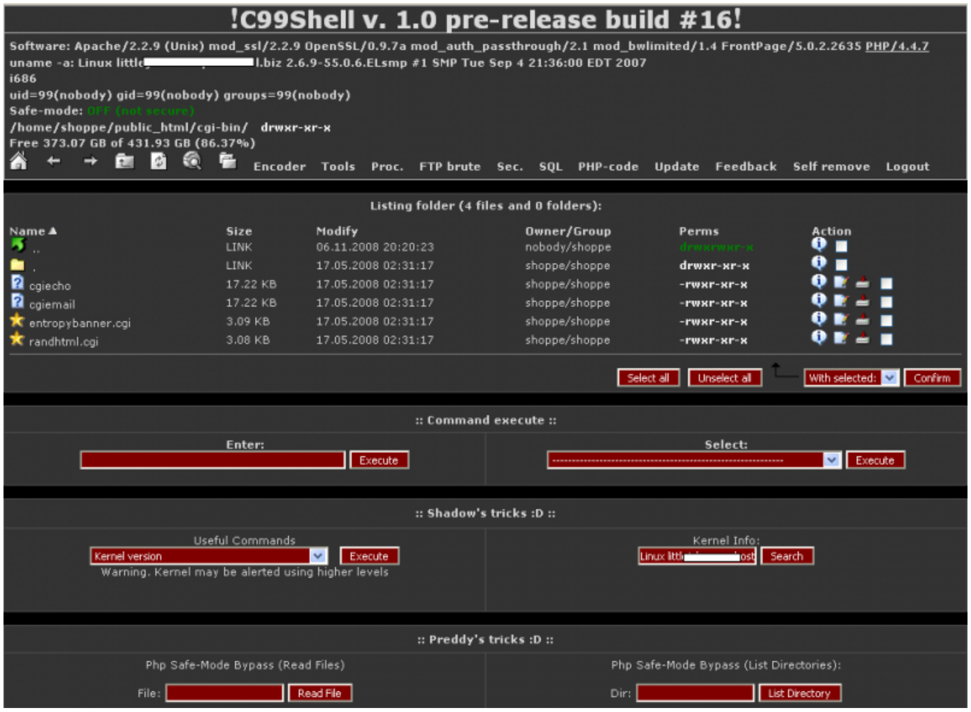
In the working of these attack models, several machine learning methods, algorithms and techniques are sued for the better working and better utilization of these methods. The attackers target or the main thought of breaking the network was the private networks, office networks and government office networks where the work can be done so easily. The best example can be easily understood is that think of there exists a coffee shop near to an industrial area. The employees of the companies surrounded will visit this shop daily once and have a cup of coffee and tea and will get to again their offices. The people whoever visiting the coffee shops will use the internet facility located at that particular shop for a while. Now, the attackers will not enter into the employee's company's websites and tries to hack the data of these employees and will do the bad things. Instead the hackers will hack the network of this coffee shop and will try to access the entire data from the website of all the employees of several companies and tries to blackmail them and do all necessary bad things.

By using the machine learning algorithms, the security to such systems and websites located at various locations of the companies and other sources of networks. The machine learning algorithms use various benchmark standards and tradeoffs and other security provided networks and websites such that by analyzing the traversals of the paths of the networks and websites. The software's or the algorithm models will detect the path traversals and the list of users whoever is working and whoever is using the facilities of the websites and also tries to identify the malicious software's and users through the websites. Several machine learning traversal of path detection algorithms are present in the source such that those can be sued for the detection on intruders or the people who are entering without any proper login detail or credentials. This algorithm also finds the source of the path from where the data or the data packets are being transferred to identify the actual sender and receiver of the data.

4. Webshell

The hackers or the people who want to store the data and try to commit some cyber crimes or cyber attacks will have several forms of types of crime. The webshell programming method is one among them. It is famous and very easy to implement also. Basically this webshell is a small piece of program or small lines of code that can be loaded into the web server or the program to where the attackers want to enter and make some fraud. This line of code was entered into the server where the root files of the hosting website or the targeted website were located and these files can be used for further more implementations. Once these lines of code was entered into the server files, the web interface of the website or the server files or the log files can be infected and also the system run commands also be infected. The codes or the lines of code entered will execute as backdoor programs from the web browsers that were being used. Here the major concern for the implementation of this code shell is that the webshell code you want to implement must be in the same lines of code for the original files language. If the file was in java, the webshell program also should in java or if the website was developed under php, then the webshell programming also must be in the php program only.

The main problem of these webshell programs or the malicious software's was these lines of code will enter into the root cause products of the system and the server files will be changed. The directory files of the servers also being changed or hacked as a result the functioning of the server files also will be under those people who had planted the webshell programs in the servers. The full access to the databases and their files will be under the trap of these people who are tracking these files. The major concern or the major problems being faced by the people are the ecommerce websites and the people who are purchasing the products through these websites by using their credit cards and other cards. The attackers will simply access these personal data of the customers and identifies the customer's credit card details and the data will be given stolen and misused by these attackers.



```
!C99Shell v. 1.0 pre-release build #16!
Software: Apache/2.2.9 (Unix) mod_ssl/2.2.9 OpenSSL/0.9.7a mod_auth_passthrough/2.1 mod_bwlimited/1.4 FrontPage/5.0.2.2635 PHP/4.4.7
uname -a: Linux littl[redacted]l.biz 2.6.9-55.0.6.ELsmp #1 SMP Tue Sep 4 21:36:00 EDT 2007
i686
uid=99(nobody) gid=99(nobody) groups=99(nobody)
Safe-mode: OFF (not secure)
/home/shoppe/public_html/cgi-bin/   drwxr-xr-x
Free 373.07 GB of 431.93 GB (86.37%)
```

Listing folder (4 files and 0 folders):

Name ▲	Size	Modify	Owner/Group	Perms	Action
..	LINK	06.11.2008 20:20:23	nobody/shoppe	drwxrwxr-x	□
.	LINK	17.05.2008 02:31:17	shoppe/shoppe	drwxr-xr-x	□
cgiecho	17.22 KB	17.05.2008 02:31:17	shoppe/shoppe	-rwxr-xr-x	□
cgiemail	17.22 KB	17.05.2008 02:31:17	shoppe/shoppe	-rwxr-xr-x	□
entropybanner.cgi	3.09 KB	17.05.2008 02:31:17	shoppe/shoppe	-rwxr-xr-x	□
randhtml.cgi	3.08 KB	17.05.2008 02:31:17	shoppe/shoppe	-rwxr-xr-x	□

:: Command execute ::

Enter:

Select: --

:: Shadow's tricks :D ::

Useful Commands: Kernel version

Warning. Kernel may be alerted using higher levels

Kernel Info: Linux littl[redacted]ost

:: Preddy's tricks :D ::

Php Safe-Mode Bypass (Read Files)

File:

Php Safe-Mode Bypass (List Directories):

Dir:

Figure 3. Webshell example [9]

These sorts of attacks can be reduced by using machine learning techniques in general. The models or the machine learning techniques can be used to identify these attackers and can warn or can remove the access to such customers from using these websites. The machine learning models will be used to observe the behaviour of the people whoever is using these websites and the operations or the tasks that those users were performing also observed thoroughly. If any user's behaviors were observed suspicious, the user's data can be blocked access. The machine learning models can be sued to identify such users by training the techniques with similar types of various models and their behaviour and the systems will always be ready to identify such users and will be caught. Several users and their data models will be given to the machine learning systems such that the machines will have a look always on the behaviors of the users and block the access to the users if any user behaves differently than the other user's behaviour with some malicious interests.

5. Spear phishing

As the growth of technology is increasing, the ways of cheating the public and the people who don't have full knowledge on these topics will be cheating more in these days. There is no way of escaping from this sorts of people, only the way can be used was to follow the preventive steps being suggested by the experts on that particular area of work in the technology. Spear phishing is one of the latest methods of techniques that were used to cheat the public. Basically, this method was sued to spoof the people's emails and tries to access the details of the mails and tries to identify the sensitive information of the public. The mails were related to an organization or an employee in the organization with some good and confidential data. The data might be regarding the financial details, financial transactions and quotes for various biddings etc. Most of the times, these sort of attacks are being done not only from the other types of attackers, these works can also be done by both the competitors and the enemies in the business lines or in their sector of competitions.

Figure 4. Spear fishing example [10]

The people being cheated by using these sorts of emails are more compared to other cases of cheating. The mails that come in these means of models are the trusted party mails such that the people whoever will observe those emails, they may think that the mails are being generated from a trusted source and can be useful. Hence, they will open the mails and their data and every content in the mails will be attacked or hacked by the users. Machine learning is one of the important choices for overcoming from these sorts of attacks. The old and normal phishing techniques are very old and can be very slow to identify such customers or users and within such slow time the hackers or the attackers will enter and will do the necessary damage. But, in the current model of operations, this damage can be reduced to some extent by using these methods.

6. Remote exploitation

The other types of network or data attacks that can be observed are the remotely exploiting the functioning of the servers or the machines. In this remote exploitations models, the attackers may sit a different or remote places which cannot be identified easily the operations will be processed form such locations. This sorts of attacks or occur on the components of the networks like routers, switches, network components and tries to access the servers or the systems data and malicious software's will be loaded and the data will be changed or hacked. The main goal of these sorts of these attacks was the stealing of the sensitive data from the servers. After stealing the data, misusing the data for their personal benefits. Machine learning techniques or the algorithms can be used to access such threats and can be eradicated. When the network algorithms are trained such that, these algorithms will always keep a close watch on these network applications and if any change in the normal behavior of the components will be observed, immediately the network can be accessed or can be treated as the network was been hacked or misused. Algorithms can be used to train for various applications and various scenarios such that to analyze the network behaviour or the network components behaviour.

7. Ransom ware

The other important and the last method attack was the ransom ware attack. It is a malware that can be incorporated to an application in the personal computer or for an office system. In this method of attack, the intruders will hack some applications that were being running in your personal or office computers and demand for money to release access to their applications and files. In most of the cases, the major motive for this sort of attacks are the money related topics. They will demand some money to release the applications and can be threaten to share the data to other users too. The major source for the spreading of this sorts of application based malwares are the emails, email attachments, software apps, applications, storage devices like pen drives, hard drives etc. These attacks also can be implemented remotely by sitting at very remote locations and the operations can be concluded or can be implemented.

Basically, this malware was is of two types. One is the file coder which will encrypt the data or it will lock the data and the second part is the lock screen which will help in locking the screen for further applications. Neural networks and deep learning techniques are the best methods to analyze this malware presence. In most of the applications, these two algorithms can be sued for the detection of data analysis. These algorithms can detect the unknown elements in a dataset. Hence, by using these neural network models the ransom malware or its related issues can be avoided.

8. Conclusion

In the current article, the day to day rise of the threats and problems from various attackers to the networks and network components, servers and systems being used in the offices, personal purposes and public places etc. Cyber security was one of the major concern areas in these days for the better utilization and better usage of the personal data and for personal financial sources. Hence, in the current article an attempt has been made to provide the details of several attacks that may occur in normal cases and the usage of machine learning algorithms how those attacks can be prevailed or controlled or identified in the earlier phase and can be protected themselves being attacked by these people. Most of the applications are given in detail and can be useful for reference and can be useful for protecting themselves from these sorts of attacks.

References

[1] RAJASIMHA KOPPULA, “Applications of machine Learning in cyber security you need to know about,” apiumhub, in technology industry trends, **(2018)**

[2] Expert System, “What is machine learning? A definition,” https://www.expertsystem.com/machine-learning-definition, **(2017)**

[3] Das, Rishabh, Morris, and Thomas, “Machine learning and cyber security,” International Conference on Computer, Electrical & Communication Engineering (ICCECE), USA, pp.1-7, **(2017)** DOI: 10.1109/ICCECE.2017.8526232

[4] Apruzzese Giovanni, Colajanni Michele, Ferretti Luca, Guido Alessandro, and Marchetti Mirco, “On the effectiveness of machine and deep learning for cyber security,” 10th International Conference on Cyber Conflict (CyCon), pp.45-56, **(2018)** DOI: 10.23919/CYCON.2018.8405026

[5] James B. Fraley and James Cannady, “The promise of machine learning in cyber security,” Southeast Con 2017, Charlotte, NC, USA, pp.55-62, **(2017)** DOI: 10.1109/SECON.2017.7925283

[6] Philip K. Chan and Richard P. Lippmann, “Machine learning for computer security”, Journal of Machine Learning Research, vol.7, pp.2669-2672, **(2006)**

[7] Chandra, Yogesh., “Application of machine learning in cyber security”, National Conference on Digitization at Delhi, **(2016)**

[8] Amrita Mitra, “What is watering hole attack?” The Security Buddy, March 9, **(2017)**

[9] Silver Moon, “What are web shells - Tutorial,” BinaryTides, Coding, Software, Tech and Reviews, **(2019)**

[10] Anirudh Bharadwaj, “How machine learning can help counter spear phishing attacks,” Dles Technologies, August, **(2018)**