SURVEY ON DATA ANALYTIC TECHNIQUES USED FOR DETECTING CYBER THREATS

V. Rama Krishna¹, Dr. R.Subhashini
¹Research Scholar, School of Computing, Sathyabama University, India
²Professor, School of Computing, Sathyabama University, India

ABSTRACT
Cyber security becoming key concern in cyber world. Attackers taking different dimensions to fly under the radar. Legacy methods like firewalls, IDS not sufficient as per the latest demands. The growth in the network speed and massive storage is making traditional systems tough to meet the latest requirements. Machine learning is key in the new approaches it was combined with clustering and Artificial intelligence is getting used widely. This paper we are exploring different cyber network threats and the data analysis methods used for detecting them.

Keywords: Cyber Security; Machine Learning; Data analytics

I. INTRODUCTION
In current world large amount of date is flowing through the network. Companies like Google, Amazon, Facebook, Yahoo using the user data to enhance their user experience. This making the cyber attackers to use that for malicious attacks and miss use for political and financial gains. Cyber security focuses on securing the devises like servers, client system (mobiles, laptops, desktops etc.) and data used in them from the attacks. Cybersecurity plays key role in security world it was set of tools, guidelines used to protect the networks [1]. Cyber security becoming critical with two reasons one the society is having high demand of digitization it usage growing to different sectors like ecommerce, health care, social network, second was attackers becoming more sophisticated using different approaches to break the systems. Bellow are the challenges faced by cyber security experts 1) collecting large amount of data 2) Managing and analyzing the unstructured data 3) identifying and predicting the threats from that data. Cyber security analysts using Machine learning, Artificial intelligence and big data for detecting the threats.

II. DIFFERENT THREATS IN CYBER SECURITY:
Phishing: By using phishing techniques cyber attackers will retrieve the confidential information from the customers

Pretending as other person: In social websites it was difficult to identify whether he is the actual person (or) not. They may keep fake profiles, photos, mail id and phone numbers.

Theft of confidential passwords: By using some malicious programs attacker can identify the passwords by using keyboard ticks. Some sites will ask users to create accounts user tend to give similar passwords that can make them enter to the other confidential sites (or) financial sites to theft the information, money.

URL fluxing: this used to shorten the URL used for the websites. Attackers can use the loop holes in existing domain search algorithms and can redirect to malicious sites.

Theft of information: By using hardware devise (or) malicious programs data can be theft from the user workstation such mailers list, passwords and confidential information
Promotion of anti-social activities, terrorism and drugs distribution

Threats coming from IoT: Wide spread of devices are connected to network starting from CCTV cameras, door regulators, refrigerators, TV, Vehicles. These electronic items share large amount of data between them. If any of these devices are compromised with security then it will be entry door for the attackers.

DOS and DDOS attack: This attacks will make the server to go down and it cant be responded to the genuine users. This attack will eat up the resources. It will be targeted to bring down the essential services. TCP flood and smurf attack are famous DDOS attacks.

![DDOS attack](image1)

**Fig 1.** DDOS attack

Eaves dropping: Intercepting the confidential messages transmitted in the network. Attacker will just listen the message he will not alter the message just will copy the messages.

![Eaves dropping](image2)

**Fig 2.** Eaves dropping

Message alteration: Attacker intercept the confidential messages and they will alter the message.

![Message alteration](image3)

**Fig 3.** Message alteration

Impersonation: In this attacker will pretend as some one else and will try retrieve the confidential information.

![Impersonation](image4)

**Fig 4.** Impersonation

Malicious program (or) software: It was program enters in to system to perform desired malicious activity.
Fig 5. Types of malicious programs

**Trapdoor:**
Initial days they used to debug the program, it trap door entry to the program attacker who is aware of this can enter to the privilege levels with out doing authentication

**Logic bombs:**
Malicious program will be hiding in the legitimate program, will start executing with special trigger such as sequence of commands, particular time based on clock clicks

**Trojan horse:** Malicious program hidden in the genuine program. can achieve certain actions that unauthorized user permitted to perform

**Viruses:** Malicious software that can infect other software’s in the system. Virus can perform certain activities like stealing CPU (or) memory cycles. Corrupting data, accessing private information

**Worms:** will use the network to spread from system to system generally warms will have remorelogign and execution capability

3.1 **Detection methods used for Cyber threats :**
Network data will be very huge the main problem will be accessing it and identifying the issue in quick session. Most of the data will be normal only small portion will have attacking pattern identifying that quickly is needed here.

3.1.1 **Big data analytics in cyber security**
Big data analytics focus on analyzing large volumes of data and get the insight in to the data. This uses statistical tools, machine learning and artificial intelligence algorithms. Some known tools like Hadoop, Flume, Casandra [2]. Belloware the key steps in big data analytics

3.1.1.1 **Threat visualization** : Weight the attack by means of origin zone and pattern. Statical tools used on historical data to see which patterns are genuine, which are malicious [3]

3.1.2 **Predictive models** : Here machine learning and Artificial intelligence techniques are used to predict which one is an attack pattern which one is normal.

3.2 **Machine learning in detecting cyber threats :**
ML learns by itself without programming. ML algorithms will be fed with network monitoring logs containing client IP address, server IP address, services used, and time of each of services. ML will help us to get future traffic trends and identify anomalies in the existing network traffic. There is significant research done on detecting botnet threats using ML, flow-based botnet detection system was developed by [4]. ML uses mathematical models to predict new incoming data. These algorithms can be classified into two types: predictive (supervised) learning always has one (or) more target variables present like number of requests made to webpage, time interval between two successive requests. The algorithms like Linear regression, Support Vector Machine (SVM) will fall under this. Pattern discovery (unsupervised learning) will not contain any target variable. Here algorithms will be focusing on association present in the data sets. Example, the malicious program will have similar patterns. Below are different Machine learning methods.

3.2.1 Extreme Learning Machine

This algorithm uses feed-forward neural networks with multiple layers of hidden nodes. The weight of the node was determined by analytical methods. As per research paper, this can learn thousand times faster than conventional algorithms [6].

![Extreme Learning Machine (ELM)](image)

3.2.2 Random Forest (RF)

It uses decision trees to decide on the classification tasks. This was taken from the concept of majority voting of multiple trees. The output was obtained as aggregative result of individual tree results. As per the research, this approach was very good in malware detection [7].

![Random Forest (RF)](image)
3.2.3 Gradient Boosting:
This model gives prediction set by using decision trees and prediction models. It adds week learners to loss function, this will be continued up to stage where residuals don’t have any patterns [8]

![Gradient Boosting Model](image)

Fig 8 Gradient Boosting:

3.2.4 Logistic Regression:
This uses binary classification. This are well suited for the problems where there are two known outcomes. This can be used to identify whether network traffic is malicious (or) not [9]

![Logistic Regression Example](image)

Fig 9 Logistic Regression

3.2.5 Machine learning in cyber security
As the threats frequency and complexity is increasing cyber world looking at machine learning to mitigate the attacks [10]. By using ML techniques routine tasks can be automated and prediction of attack can be performed.
For future prediction and pattern analysis machine learning will be clubbed with cluster classification algorithms [11].

a) Detecting threats: This is like detecting the threats before they actually appear. [18] Here vast majority of data will be analyzed from that malicious actives will be identified. The attack patterns data set is given as input from that machine learning techniques can identify any threat that is coming in future

b) Network risk scoring: This is to classify the network in to different zones based on the risk associated with it. Machine learning can be used to identify risk in each zone by analyzing the network traffic. There are famous algorithms used in this area like SVM, Random forest algorithm [12]

c) Automating the routine security tasks: Some of the routine activities in the security world like network data analysis and malware assessments can be automated[13]. Also used to reduce the time spend on reducing false positive [14]

3.3 AI in cyber security:
Artificial intelligence can be used to do the reputative human tasks and automating the tasks.

3.3.1 Authentication
Traditional usage of passwords and usernames are becoming easy targets so using biometric and face recognition is going to be future for the authentication. As per Apple research findings unlocking Apple gadget success rate of opening with different face recognition possible only a one-in-a-million chance [15]

3.3.2 AI in Phishing Detection:
As per surveys out of 99 mails one mail will be phishing. AI techniques used to identify the fake mail id’s, fake websites, fake news. AI uses methods like score of webpage/site, Weight fact will understand the heading and body to find other sites reporting simmialr information. scan for the sensational trapping words that used in phishing [16]

3.3.3 Behavioral Analysis with AI:
AI algorithms will be used in identifying the behavioral patterns of the system like browsing history, login times any thing deviating from that can be recognized as abnormal activity. The abnormal activity can include quick shopping, fast key board rate than usual visiting un usual websites frequently etc. [17]

III. CONCLUSION
In this paper, we have discussed different cyber threats and new techniques like data analytics, machine learning artificial intelligence techniques. There is catch here even cyber criminals also will use the advanced techniques to make attacks so it was very important to come out of this. This paper used as base paper to study the existing techniques Going forward we are working on scalable cyber secure defense systems that can leverage best of the abilities of IDS, Data analytics and machine learning techniques can over come the existing and upcoming threats in effective manner

REFERENCES
[6] Shifei Ding, Han Zhao, Yanan Zhang Extreme learning machine: algorithm, theory and applications, Neural Computing and Applications ISSN 0941-0643
[8] Gerard BlauBenol Cadre etc , Optimization by gradient boosting Statistics Theory (math.ST); Machine Learning (cs.LG)


[15] Stefano Arca, Paola Campadelli, Elena Casiraghi, Raella Lanzarotti, "An Automatic Feature Based Face Authentication System", 16th Italian Workshop on Neural Nets (WIRAN), 2005, pp. 120-126

