Detecting cyber attacks on web applications by applying Different Machine learning techniques

Dr. Manam Vamsi Krishna[1], Dr .B Priyanka[2]
#1 Currently Working as Assistant Professor in CSE dept. Malla Reddy Institute of Technology, Hyderabad..
#2 Currently Working as Assistant Professor in CSE dept. Malla Reddy Institute of Technology, Hyderabad.

Abstract:

Increased use of cloud services, a growth in the number of online application users, changes in network architecture that connects devices running mobile operating systems, and continually expanding network technologies all provide new cyber security challenges. As a result, network security methods, sensors, and protection schemes must change to accommodate the needs and issues of users in order to counter emerging threats. Because growing application layer cyber attacks are recognised as top risks and the key problem for network and cyber security, we will focus on countering them in this essay. The article's main contribution is the suggestion of a machine learning approach for modelling normal application behaviour and detecting cyber assaults. Patterns (in the form of Perl Compatible Regular Expressions (PCRE) regular expressions) are obtained utilising a graph-based segmentation technique and dynamic programming to create the model. The model is based on data gathered from client-generated HTTP requests to a web server. On the CSIC 2010 HTTP Dataset, we tested our technique and found it to be effective.

1.INTRODUCTION

The number of security incidents reported around the world has recently risen. The number of attacks has increased dramatically in comparison to prior years, according to national CERTs (for example, CERT Poland [1]). According to the report [1,] there were 1082 incidents in 2012, an increase of about 80% over the previous year, owing primarily to malware and phishing.

The rise in occurrences is directly tied to the growing number of mobile device users, who make up the population of connect-from-anywhere terminals and routinely test the traditional network security boundaries. The so-called BYOD (bring your own device [4]) trend also exposes many businesses' traditional security to novel and emerging dangers. Many modern malwares, such as ZITMO (Zeus In The Mobile), are focused on obtaining information on users, their personal data, and gaining access to remote services such as banks and web services, rather than on the mobile device itself.

2.LITERATURE SURVEY

2.1 R. Christopher, "Port scanning techniques and the defense against them," SANS Institute, 2001.

Port Scanning is one of the most popular techniques attackers use to discover services that they can exploit to break into systems. All systems that are connected to a LAN or the Internet via a modem run services that listen to well-known and not so well-known ports. By port scanning, the attacker can find the following information about the targeted systems: what services are running, what users own those services, whether anonymous logins are supported, and whether certain network services require authentication. Port scanning is accomplished by sending a message to each port, one at a time. The kind of response received indicates whether the port is used and can be probed for further weaknesses. Port scanners are important to network security technicians because they can reveal possible security vulnerabilities on the targeted system. Just as port scans can be ran against your systems, port scans can be detected and the amount of information about open services can

be limited utilizing the proper tools. Every publicly available system has ports that are open and available for use. The object is to limit the exposure of open ports to authorized users and to deny access to the closed ports.

2.2 S. Staniford, J. A. Hoagland, and J. M. McAlerney, "Practical automated detection of stealthy portscans," Journal of Computer Security, vol. 10, no. 1-2, pp. 105–136, 2002.

Portscanning is a common activity of considerable importance. It is often used by computer attackers to characterize hosts or networks which they are considering hostile activity against. Thus it is useful for system administrators and other network defenders to detect portscans as possible preliminaries to a more serious attack. It is also widely used by network defenders to understand and find vulnerabilities in their own networks. Thus it is of considerable interest to attackers to determine whether or not the defenders of a network are portscanning it regularly. However, defenders will not usually wish to hide their portscanning, while attackers will. For definiteness, in the remainder of this paper, we will speak of the attackers scanning the network, and the defenders trying to detect the scan. There are several legal/ethical debates about portscanning which break out regularly on Internet mailing lists and newsgroups.

2.3 M. C. Raja and M. M. A. Rabbani, "Combined analysis of support vector machine and principle component analysis for ids," in IEEE International Conference on Communication and Electronics Systems, 2016, pp. 1–5.

Compared to the past security of networked systems has become a critical universal issue that influences individuals, enterprises and governments. The rate of attacks against networked systems has increased melodramatically, and the strategies used by the attackers are continuing to evolve. For example, the privacy of important information, security of stored data platforms, availability of knowledge etc. Depending on these problems, cyber terrorism is one of the most important issues in today's world. Cyber terror, which caused a lot of problems to individuals and institutions, has reached a level that could threaten public and country security by various groups such as criminal organizations, professional persons and cyber activists. Intrusion detection is one of the solutions against these attacks. A free and effective approach for designing Intrusion Detection Systems (IDS) is Machine Learning. In this study, deep learning and support vector machine (SVM) algorithms were used to detect port scan attempts based on the new CICIDS2017 dataset Introduction Network Intrusion Detection System (IDS) is a software-based application or a hardware device that is used to identify malicious behavior in the network [1,2].

3.PROPOSED SYSTEM

In this paper author is describing concept to detect attack perform on Web Applications using Graph-based approach and Estimating dissimilarities between two components Needleman–Wunsch algorithm.

In graph based approach a graph will form using vertex (circle in graph) and edges are the line connection between two vertexes. Vertex will contains http request data which is coming from client to server, this http data will contains normal or attack data and by analyzing such data we can detect whether request is normal or attack.

All requests which are normal will have similarity and will be goes into same group by adding edges between those two similar http request and attacker will modify request data to perform some malicious behaviour and there will be not much similarity left (due to request data modification) which can indicate us that this request contains attack.

We can check similarity between two request data using Needleman–Wunsch algorithm.

3.1 DATASET DESCRIPTION

To implement this project author is using CSIC dataset and below is the dataset example
GET http://localhost:8080/tienda1/index.jsp HTTP/1.1

User-Agent: Mozilla/5.0 (compatible; Konqueror/3.5; Linux) KHTML/3.5.8 (like Gecko)
Pragma: no-cache
Cache-control: no-cache
Accept: text/xml,application/xml,application/xhtml+xml,text/html;q=0.9,text/plain;q=0.8,image/png,*/*;q=0.5
Accept-Encoding: x-gzip, x-deflate, gzip, deflate
Accept-Charset: utf-8, utf-8;q=0.5, *;q=0.5
Accept-Language: en
Host: localhost:8080
Cookie: JSESSIONID=1F767F17239C9B670A39E9B10C3825F4Connection: close

Above is the normal request data in bold format and from above dataset just we need to look for http data (**GET** http://localhost:8080/tienda1/index.jsp HTTP/1.1) and we extract only http data from above dataset using REGULAR EXPRESSION concept. Below is request data which contains SQL Injection Attack

GET http://localhost:8080/tienda1/publico/anadir.jsp?id=2&nombre=Jam%F3n+Ib%E9rico&precio=85&cantidad=%27%3B+DROP+TABLE+usuarios%3B+SELECT+*+FROM+datos+WHERE+nombre+LIKE+%27%25&B1=A%F1adir+al+carrito HTTP/1.1
User-Agent: Mozilla/5.0 (compatible; Konqueror/3.5; Linux) KHTML/3.5.8 (like Gecko)

4.RESULTS AND DISCUSSIONS

Pragma: no-cache
Cache-control: no-cache
Accept: text/xml,application/xml,application/xhtml+xml,text/html;q=0.9,text/plain;q=0.8,image/png,*/*;q=0.5
Accept-Encoding: x-gzip, x-deflate, gzip, deflate
Accept-Charset: utf-8, utf-8;q=0.5, *;q=0.5
Accept-Language: en
Host: localhost:8080
Cookie: JSESSIONID=B92A8B48B9008CD29F622A994E0F650D
Connection: close

In above http request data we can see attacker is performing SQL Injection attack

http://localhost:8080/tienda1/publico/anadir.jsp?id=2&nombre=Jam%F3n+Ib%E9rico&precio=85&cantidad=%27%3B+<u>DROP+TABLE+usuarios%3B+SELECT+*+FROM+datos+WHERE+nombre+LIKE+%27%25&B1=A</u>%F1adir+al+carrito HTTP/1.1

For clarity you can see above request data in underline text attacker is trying to execute SQL Drop query. All normal requests may not contain that drop query due to which dissimilarity will occur to detect it as attack.

CSIC dataset comes in training and test data, training dataset contains all normal possible request data which can check with test (newly arrived request) data to detect it as normal or attack.
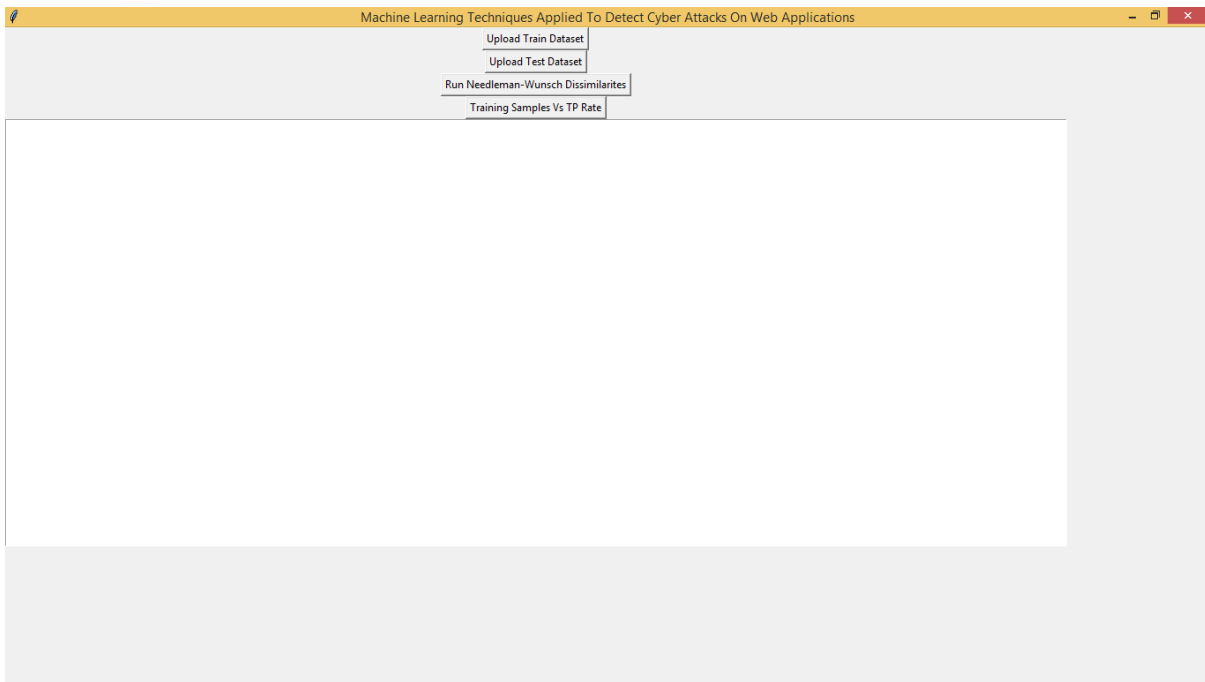
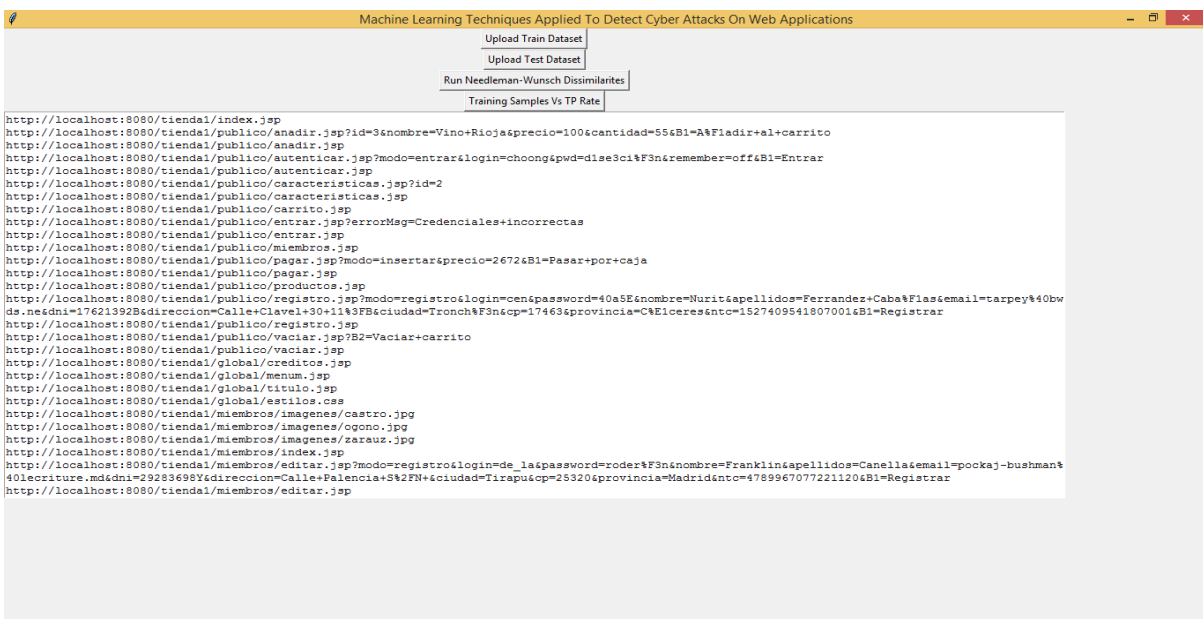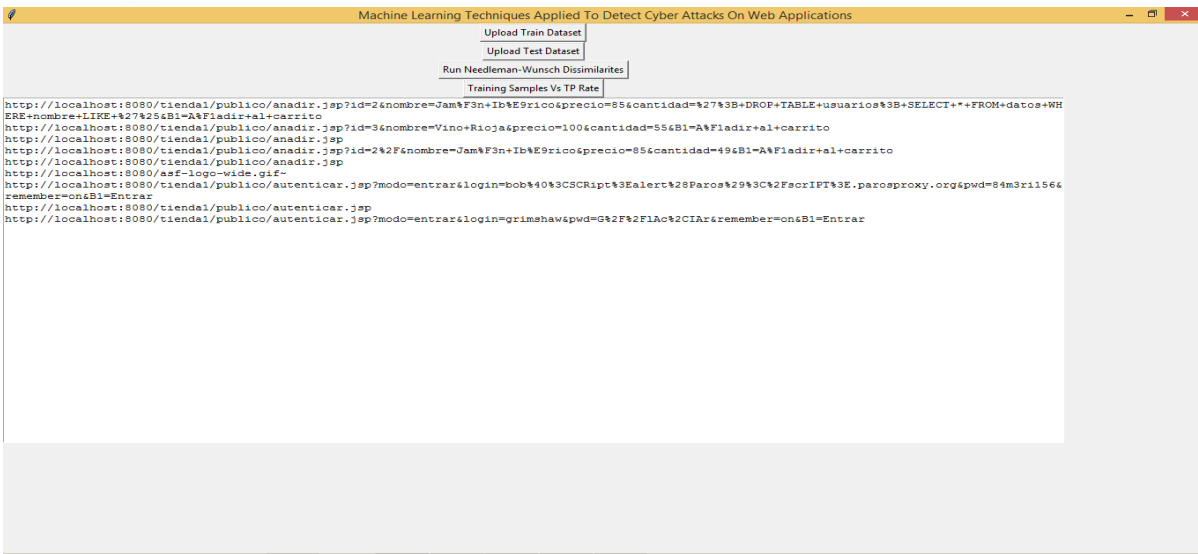**Fig 4.1: 'Upload Train Dataset' button to upload normal training data**

Fig 4.2 After uploading we can see only http request URL data is extracted using regular expression from training data and this will apply on test data to get result. Now upload test data

International Journal For Advanced Research
In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362
IJARST

Fig 4.3 Above are some test request data, now click on 'Run Needleman-Wunsch Dissimilarites' button to check similarity between train and test request data
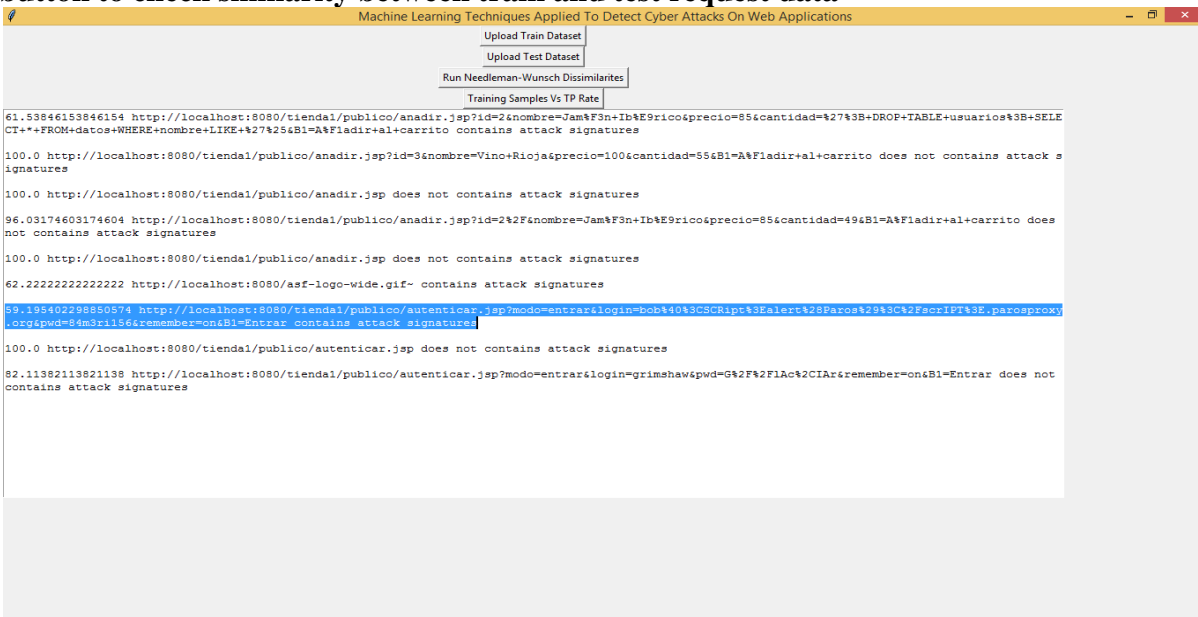
Fig 4.4 In above screen in selected text u can see first contains similarity score between train request data and test request data and then request data is displaying and then showing whether its normal or contains attack signatures.
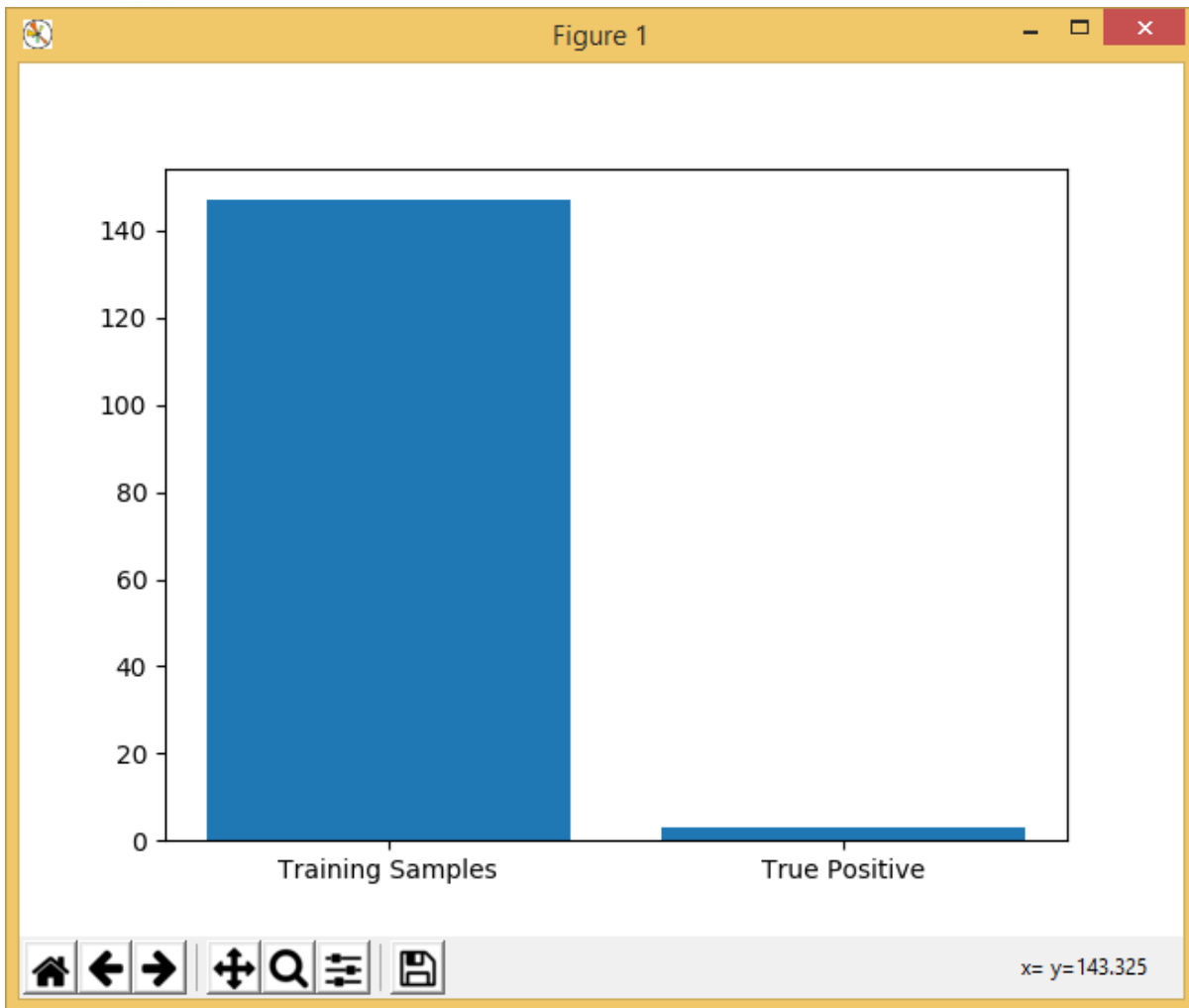
Fig 4.5 In above graph x-axis contains total train dataset size and true positive detection rate and y-axis contains length

5.CONCLUSION

The method for application layer attack detection based on machine learning was proposed in this paper. Patterns (in the form of PCRE regular expressions) are obtained utilising a graph-based segmentation technique and dynamic programming to create the model. Modeling the true behaviour of programmes and detecting cyber assaults are both done with regular expressions. We also presented findings that demonstrate the efficacy of the proposed approach, which may be utilised to detect application layer attacks. Experiments on CSIC'10 reveal that the proposed approach can obtain a detection ratio of 94.46 percent while maintaining a low error rate.

REFERENCES

[1] CERT Polska Annual Report 2012. http://www.cert.pl/PDF/Report_CP_2012.pdf

[2] SOPHOS homepage http://www.sophos.com

[3] Cisco Annual Report 2013. http://www.cisco.com/web/about/ac49/ac20/ac19/ar2013/docs/ 2013_Annual_Report.pdf

[4] BYOD: Bring Your Own Device. http://www.vs.inf.ethz.ch/publ/papers/rohs-byod-2004.pdf

[5] OWASP Top 10 2013. https://www.owasp.org/index.php/Top_10_2013-Top_10

[6] NSG. http://www.ijcst.com/vol31/4/sridevi.pdf

[7] LESG. http://www.cs.northwestern.edu/~ychen/Papers/ LESG-ICNP07.pdf

[8] A. Shabtai, E. Menahem and Y. Elovici. F-Sign: automatic, function-based signature generation for malware, systems, man, and cybernetics, Part C: applications and reviews. Transactions on IEEE, 41, 494–508, 2011.

[9] D. Kong, J. Gong, S. Zhu, P. Liu and H. Xi. SAS: semantics aware signature generation for polymorphic worm detection. International Journal of Information Security, 50, 1–19, 2011.

Author's Profile:

Dr. MANAM VAMSI KRISHNA.
Received his PhD degree in computer science and engineering department from Sri Satya Sai University Technology & Medical Sciences in July 2021. Currently Working as Assistant Professor in CSE dept. Malla Reddy Institute of Technology, Hyderabad. His Research interests in cryptography & Network security, Cyber Security, computer networks, Data mining.

Dr B.Priyanka

Completed here B. Tech. In Computer Science Engineering (2014) Completed M.Tech. in Computer Science Engineering (2016) Received her PhD degree in computer science and engineering department from Sri Satya Sai University Technology & Medical sciences in September 2021.

Research interest: Big data, Machine learning, Data mining Working As Assistant Professor in Mallareddy Institute of Technology