DETECTION OF CYBER ATTACKS ON WEB APPS THROUGH DIFFERENT MACHINE LEARNING TECHNIQUES

[1]Dr. Shiva Kumar B, [2]Ashok M, [3]Kirankumar Nellutla, [4]Yeldi Srilekha
[1,2,3]Assistant Professor, [4]UG Student, [1,2,3,4]Department of Computer Science Engineering, Rishi MS Institute of Engineering and Technlogy for Women, Kukatpally, Hyderabad.

ABSTRACT
New cyber security difficulties are brought on by increased usage of cloud services, an increase in users of online applications, modifications to the network architecture connecting devices running mobile operating systems, and rapidly evolving network technology. As a result, in order to address user wants and concerns, network security techniques, sensors, and protection strategies must adapt. We will concentrate on preventing escalating application layer cyber attacks in this article since they are acknowledged as top threats and the main issue for network and cyber security. The recommendation of a machine learning technique for modeling typical application behavior and identifying cyber attacks is the article's primary contribution. Regular expressions in the form of Perl Compatible Regular Expressions (PCRE) are retrieved as patterns utilizing 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.

INTRODUCTION
Recently, there has been more security issues recorded globally. According to national CERTs, assaults have drastically grown in comparison to previous years (for example, CERT Poland [1]). In 2012, there were 1082 instances, up almost 80% over the previous year, mostly as a result of malware and phishing, according to the study [1].
Increased use of mobile devices, which make up the majority of connect-from-anywhere terminals and frequently test the traditional network security perimeters, is directly related to an increase in incidents. Bring your own device (BYOD) trends also puts many firms' traditional security at risk from innovative and emerging threats. 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 deviceitself.

LITERATURE SURVEY
R. Christopher, “Port scanningtechniques and the defense againstthem,” SANS Institute, 2001.Port Scanning is one of the most popular techniques attackers use to discover services that they can exploit to breakinto 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, whetheranonymous logins are supported, and whether certain network services require authentication. Port scanning is accomplished by sending a message toeach 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 theproper tools. Every publicly available system has ports that are open andavailable for use. The object is to limitthe exposure of open ports to authorized users and to deny access to the closedports.
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 usedby 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 regularlyon Internet mailing lists and newsgroups.

M. C. Raja and M. M. A. Rabbani,"Combined analysis of support vectormachine and principle component analysis for ids," in IEEE InternationalConference on Communication andElectronics Systems, 2016, pp. 1–

5. Compared to the past security of networked systems has become a critical universal issue thatinfluencesindividuals, enterprises and governments.The rate of attacks against networkedsystems has increased melodramatically,and the strategies used by the attackersare continuing to evolve. For example, the privacy of important information,security of stored data platforms, availability of knowledge etc. Dependingon these problems, cyber terrorism is oneof the most important issues in today'sworld. Cyber terror, which caused a lot ofproblems 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 solutionsagainst these attacks. A free and effectiveapproach for designing IntrusionDetection Systems (IDS) is MachineLearning. In this study, deep learning andsupport 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 behaviorin the network [1,2].

PROPOSED SYSTEM

In this paper author is describing concept to detect attack perform on WebApplications using Graph-based approach and Estimating dissimilaritiesbetween 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 httprequest 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 requestis 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 notmuch 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–Wunschalgorithm.

DATASET DESCRIPTION

To implement this project author is usingCSIC dataset and below is the dataset example

GET
http://localhost:8080/tienda1/index.jspHTTP/1.1
User-Agent: Mozilla/5.0 (compatible; Konqueror/3.5; Linux) KHTML/3.5.8(like Gecko)
Pragma: no-cache Cache-control: no-cacheAccept:
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:8080Cookie: JSESSIONID=1F767F17239C9B670A3 9E9B10C3825F4Connection: 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 requestdata which contains SQL Injection Attack
GET
http://localhost:8080/tienda1/publico/anadir.jsp?id=2&nombre=Jam%F3n+I
b%E9rico&precio=85&cantidad=%27%3B+DROP+TABLE+usuarios%3B+SELECT+*+FROM+datos+W
HERE+nombre+LIKE+%27%25&B1=A%F1 adir+al+carrito 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: enHost: localhost:8080
Cookie: JSESSIONID=B92A8B48B9008CD29F622A994E0F650D Connection: close
In above http request data we can seeattacker is performing SQL Injectionattack
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

For clarity you can see above request data in underline text attacker is trying toexecute SQL Drop query. All normal requests may not contain that drop querydue to which dissimilarity will occur to detect it as attack. CSIC dataset comes in training and test data, training dataset contains all normalpossible request data which can check with test (newly arrived request) data to detect it as normal or attack.

RESULTS AND DISCUSSIONS

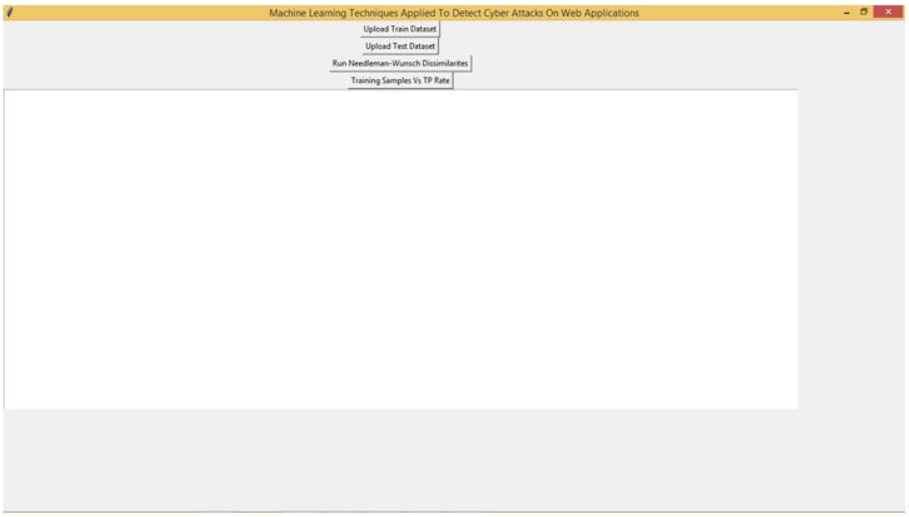

Fig 4.1: „Upload Train Dataset“ button to upload normal training data

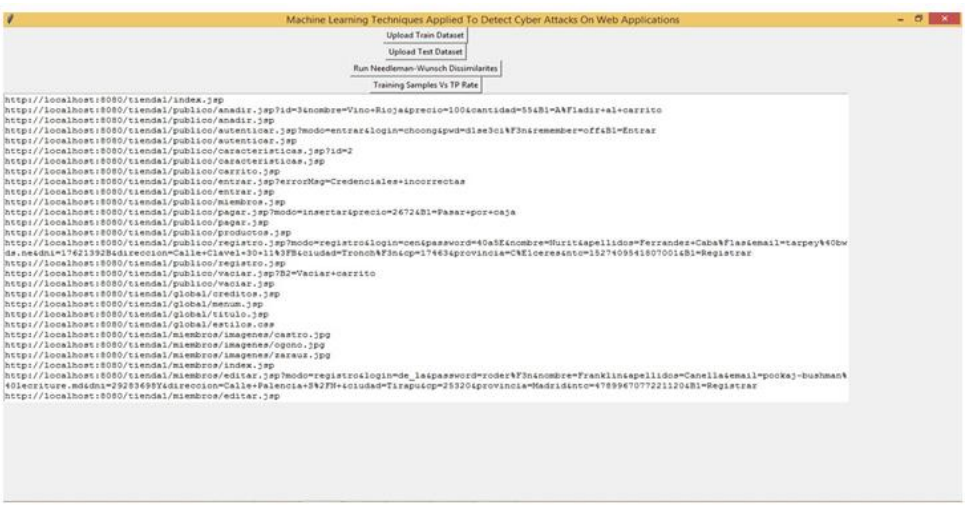

Fig 4.2 After uploading we can see only http request URL data is extracted using regular expression from training data and this will apply on test data toget result. Now upload test data

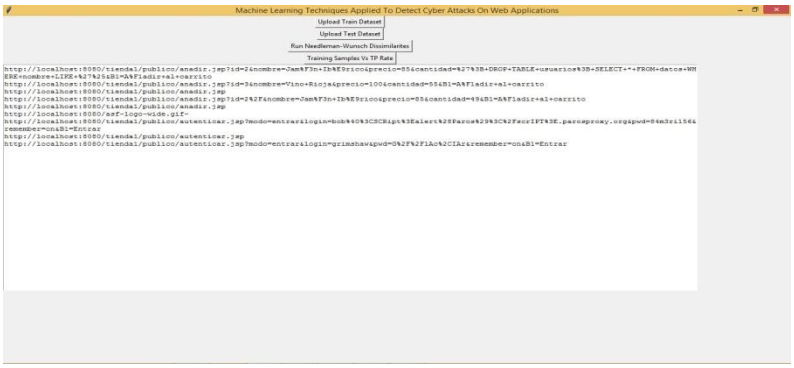

Fig 4.3 Above are some test request data, now click on „Run Needleman- WunschDissimilarites" button to check similarity between train and test request data

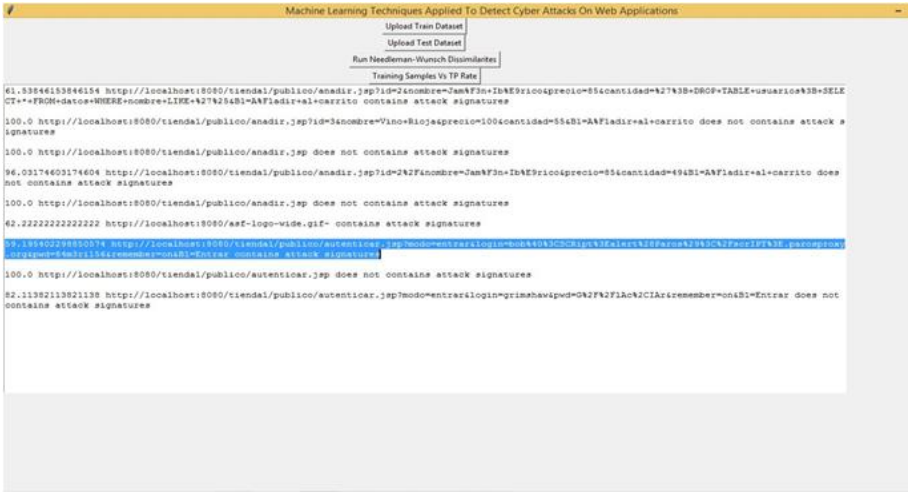

Fig 4.4 In above screen in selected text u can see first contains similarity score between train request data and test request data and then request data is displayingand then showing whether its normal or contains attack signatures.

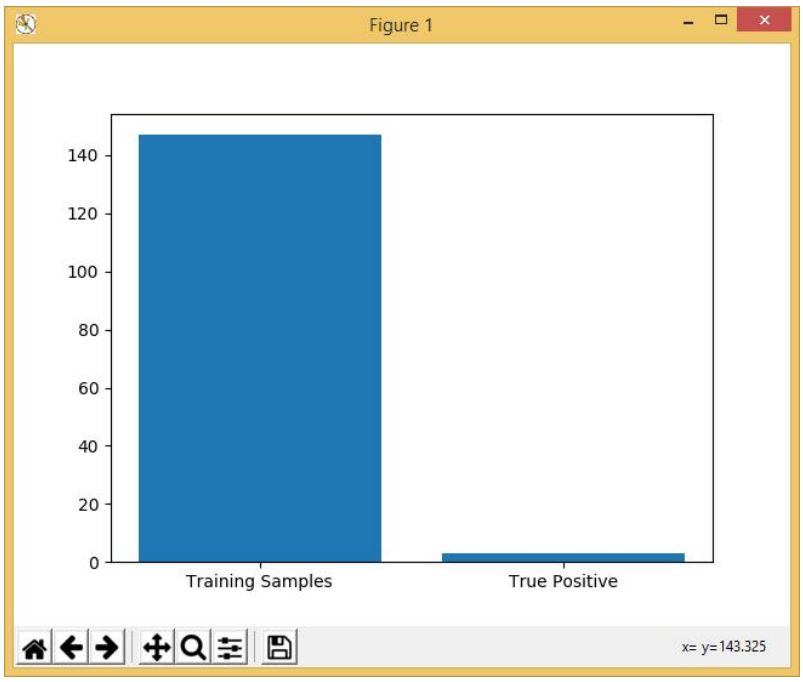

Fig 4.5 In above graph x-axis contains total train dataset size and true positive detection rate and y-axis contains length

CONCLUSION

In this article, a machine learning-based approach to application layer attack detection was developed. To develop the model, a graph-based segmentation approach and dynamic programming are used to acquire patterns (in the form of PCRE regular expressions). Regular expressions are used to both identify cyber attacks and model the real behavior of programmers. We also provided data that shows how effective the suggested method is for identifying application layer threats. The suggested method can achieve a detection ratio of 94.46 percent while retaining a low error rate, according to tests performed on CSIC'10.

REFERENCES

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

2. SOPHOS homepagehttp://www.sophos.com Cisco Annual Report 2013. http://www.cisco.com/web/about/ac49/ac20/ac19/ar2013/docs/ 2013_Annual_Report.pdf

3. BYOD: Bring Your Own Device. http://www.vs.inf.ethz.ch/publ/papers/rohs-byod-2004.pdf OWASP Top 10 2013. https://www.owasp.org/index.php/Top_1 0_2013-Top_10 NSG. http://www.ijcst.com/vol31/4/sridevi.pdf LESG. http://www.cs.northwestern.edu/~ychen/ Papers/LESG-ICNP07.pdf

4. 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 onIEEE, 41, 494–508, 2011.

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