IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 7, July 2016

A Practical Approach of Neural Network Based Detection and Prevention of Malicious URL using MATLAB

Varuna[1], **Preeti Gupta**[2]

M.Tech, CSE, JCDM College of Engineering, Sirsa, India[1]

Asst Professor (CSE), JCDM College of Engineering, Sirsa, India[2]

Abstract: The Web contains the mixed type of URL which contains the Malicious as well as normal URLS. The research work is about the study of Malicious URLS techniques and provides the approach of security along with the Detection and Prevention. The System information need to be secure and should be confidential. The system information protection is main aspect of this research work. This paper proposed the approach for prevention and Detection of the Webpage URL using Neural Network. This research work provides a new way of securing the information to avoid hassle in transmission over network. The Dictionary has been created of which are counted as the malicious and input data of URL with parameters such as IsURLKeywords, Toxic Factor based on auto Downloading Links and the third parameter is No of Download links. Based on these parameters, the weight has been assigned and equation has been created. The threshold values have been settled and generated results in MATLAB Simulation Tool.

Keywords: Botnets, Denial-Of-service attacks, IDS, SQL Injection Attack, Penetrations.

I. INTRODUCTION

The web has become the medium of selection for folks to look for info, conduct business, and luxuriate in recreation. At an equivalent time, the net has conjointly become the first platform employed by miscreants to attack users. as an example, drive-by-download attacks area unit a preferred selection among larva herders to grow their botnets. in an exceedingly drive-by-download attack, the assailant infects a (usually benign) computer with malicious code that eventually ends up in the exploitation of vulnerabilities within the net browsers (or plug-ins) of unsuspecting guests. If in, the exploit generally downloads and executes a malware binary, turning the host into a larva. Additionally to drive-by-download exploits, cybercriminals conjointly use social engineering to trick victims into putting in or running untrusted package. As associate example, contemplate an internet page that asks users to put in a faux video player that's presumptively necessary to point out a video (when, in fact, it's a malware binary). Another example is faux anti-virus programs. These programs area unit unfold by web content that scare users into thinking that their machine is infected with malware, attractive them to transfer associated execute an actual piece of malware as a remedy to the claimed infection. the net may be a terribly giant place, and new pages (both legitimate and malicious) area unit extra at a frightening pace. Attackers unrelentingly scan for vulnerable hosts that may be exploited and leveraged to store malicious pages, that area unit than organized in advanced malicious meshes to maximise the changes that a user can land on them. As a result, it's a difficult task to spot malicious pages as they seem on the net.

However, it's crucial to succeed at this task so as to safeguard net users. as an example, one will leverage info regarding web content that compromise guests to make blacklists. Blacklists stop users from accessing malicious content within the 1st place, and became a preferred defense resolution that's supported by all major browsers. Moreover, the flexibility to quickly realize malicious pages is critical for vendors of anti-virus product World Health Organization have to be compelled to acquire, as quick as attainable, fresh free malware samples to update their signature databases. checking out malicious web content may be a three-step method, during which URLs area unit 1st collected, then quickly inspected with quick filters, and eventually examined full mistreatment specialized analyzers. a lot of exactly, one has got to 1st collect tips that could web content (URLs) that area unit continue to exist the net. to gather URLs, one generally uses net crawlers, that area unit programs traversing the net in an exceedingly systematic fashion. ranging from a collection of initial pages, this program follows hyperlinks to seek out as several (different) pages as attainable.

Web Applications

The web may be a extremely programmable surroundings that permits mass customization through the immediate readying of an oversized and various vary of applications, to lots of international users. Two necessary parts of a contemporary web site area unit versatile internet browsers and internet applications; each obtainable to all or any and varied at no expense. Web browsers area unit code applications that permit users to retrieve knowledge and

DOI 10.17148/IJARCCE.2016.5712

move with content set on websites at intervals a web site. Today's websites area unit a way cry from the static text and graphics showcases of the first and mid-nineties: fashionable websites permit personalised dynamic content to be force down by users per individual preferences and settings. moreover, websites can also run client-side scripts that "change" the internet browser into Associate in Nursing interface for such applications as web mail and interactive mapping code (e.g., Yahoo Mail and Google Maps).

Most significantly, fashionable websites permit the capture, processing, storage and transmission of sensitive client knowledge (e.g., personal details, mastercard numbers, Social Security data, etc.) for immediate and repeated use. And, this is often done through internet applications.

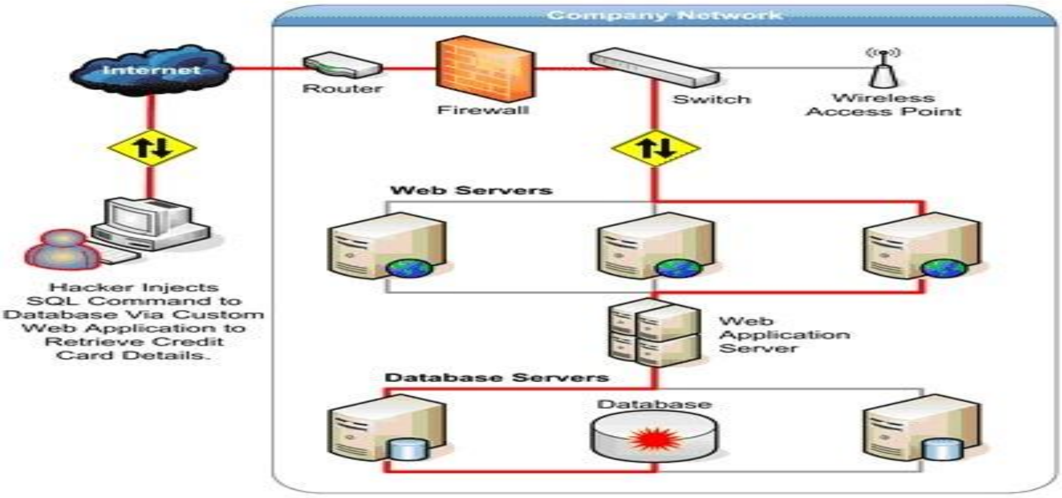


Figure 1: company network

Such features as webmail, login pages, support and product request forms. These are all common examples of web applications. The Malicious URLs detection system works the following:

a. It allows organizations to protect their systems from the threats that come within systems.
b. To prevent and to reduce the range of unauthorized range usage of the system resources.
c. To detect attacks and other security violations when they are getting compromised.
d. To act as a quality contest for security design and administration, especially of large and Complex enterprises.
e. To detect computer break-ins, penetrations, Denial-Of-service attacks, SQL injection attack, port scan vulnerabilities or flows in system.
f. To provide useful information, recovery, and correction of causative factors.
g. To detect deviations in usage and to recognize types of security violations.

II. OBJECTIVES

The Proposed work can be extended using neural Network technique. The Objectives are mentioned as for new proposed Work:

1. To Study the Types and Collection of Malicious Keywords.
2. URLs and Keywords Dataset Initialization.
3. To Develop the Efficient Algorithm for improve the Searching and detection using Neural Network and detect the malicious URL's.
4. To implement the algorithm in MATLAB Simulation Tool and generate results.

III. PROBLEM FORMULATION AND METHODOLOGY

Malicious transactions might injury the information integrity and convenience. However, in spite of the pertinency of the detection of malicious information transactions, the truth is that only a few sensible mechanisms that area unit ready to determine users death penalty malicious transactions has been planned up to now. The Malicious URLs causes the safety mechanisms that area unit incorrectly organized allowing potential intruders (hackers) to access the information.

Unauthorized users "still" the credentials of licensed users so as to access the information server. This analysis planned a brand new mechanism for the detection of malicious URLs of Webpages and prevents the user to urge access within the web content. As during a typical information surroundings it's potential to outline the profile (sequence of SQL commands) of all the transactions that every user is allowed to execute, the planned mechanism uses that profile of valid transactions to spot user's makes an attempt to execute invalid sequences of commands. The sequences of commands that represent every valid group action area unit obtained by analysis of the execution profile of the information shopper applications.

Research Methodology

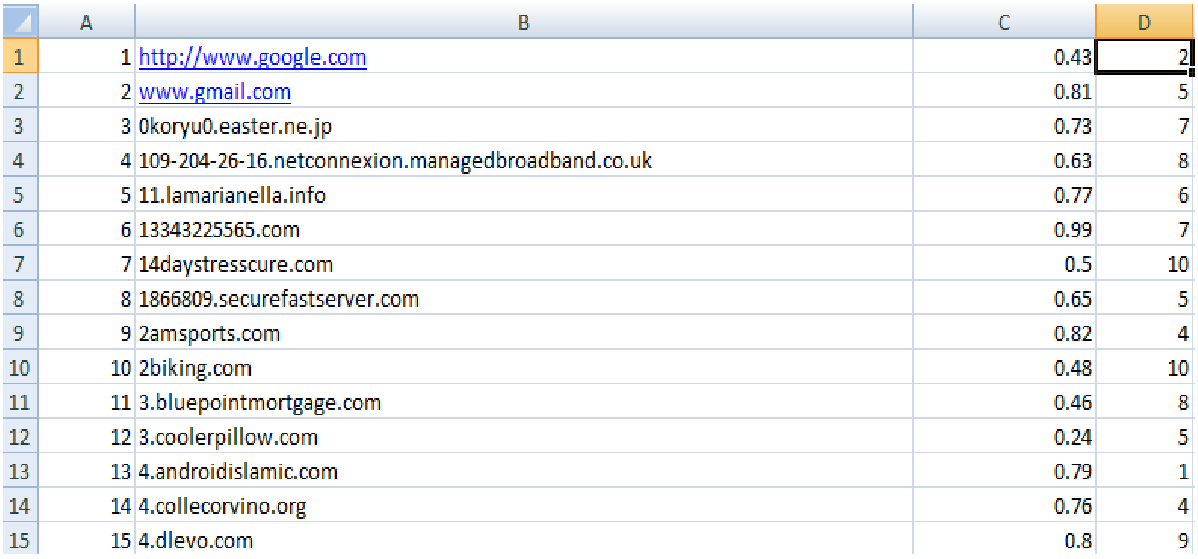
1. Start
2. Initially Malicious keywords dictionary has been initialized.
3. The Collection of URL's has been initialized with the parameters consideration such as Toxicity and Downloading Links.
4. Develop an equation of Neural Network with Weights assigning to the input values.
5. Computation of equation value with equation F+4*Toxic+2*Links
6. Settle the Threshold Value and compare the equation values.
7. Categorize the URL in High, Low, Medium category based on the Threshold values.
8. Generate the classification results using Graphical representation.
9. Stop

IV. RESULTS & DISCUSSIONS

In this chapter for the implementation of the research work a algorithm has been proposed. This proposed algorithm when applied on the collected URL's, predict the risk level and number of malicious words found in URL.

Input

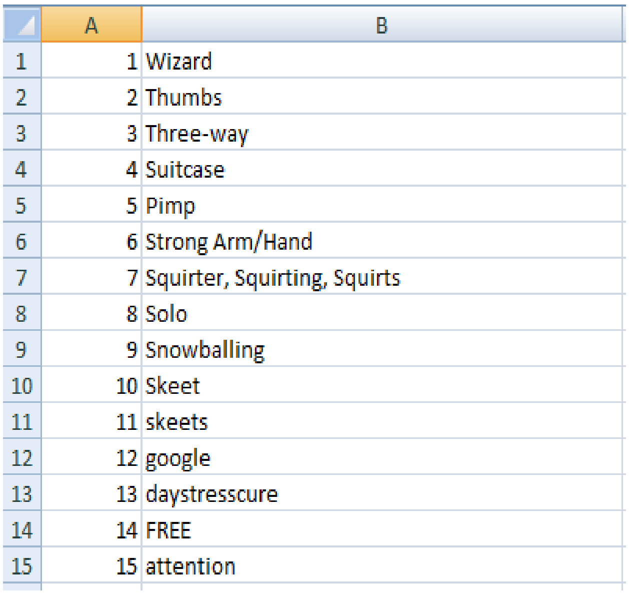
INPUT URL's. With number of downloading links on Home Page.



	A	B	C	D
1	1	http://www.google.com	0.43	2
2	2	www.gmail.com	0.81	5
3	3	0koryu0.easter.ne.jp	0.73	7
4	4	109-204-26-16.netconnexion.managedbroadband.co.uk	0.63	8
5	5	11.lamarianella.info	0.77	6
6	6	13343225565.com	0.99	7
7	7	14daystresscure.com	0.5	10
8	8	1866809.securefastserver.com	0.65	5
9	9	2amsports.com	0.82	4
10	10	2biking.com	0.48	10
11	11	3.bluepointmortgage.com	0.46	8
12	12	3.coolerpillow.com	0.24	5
13	13	4.androidislamic.com	0.79	1
14	14	4.collecorvino.org	0.76	4
15	15	4.dlevo.com	0.8	9

Figure 2: collection of URL's

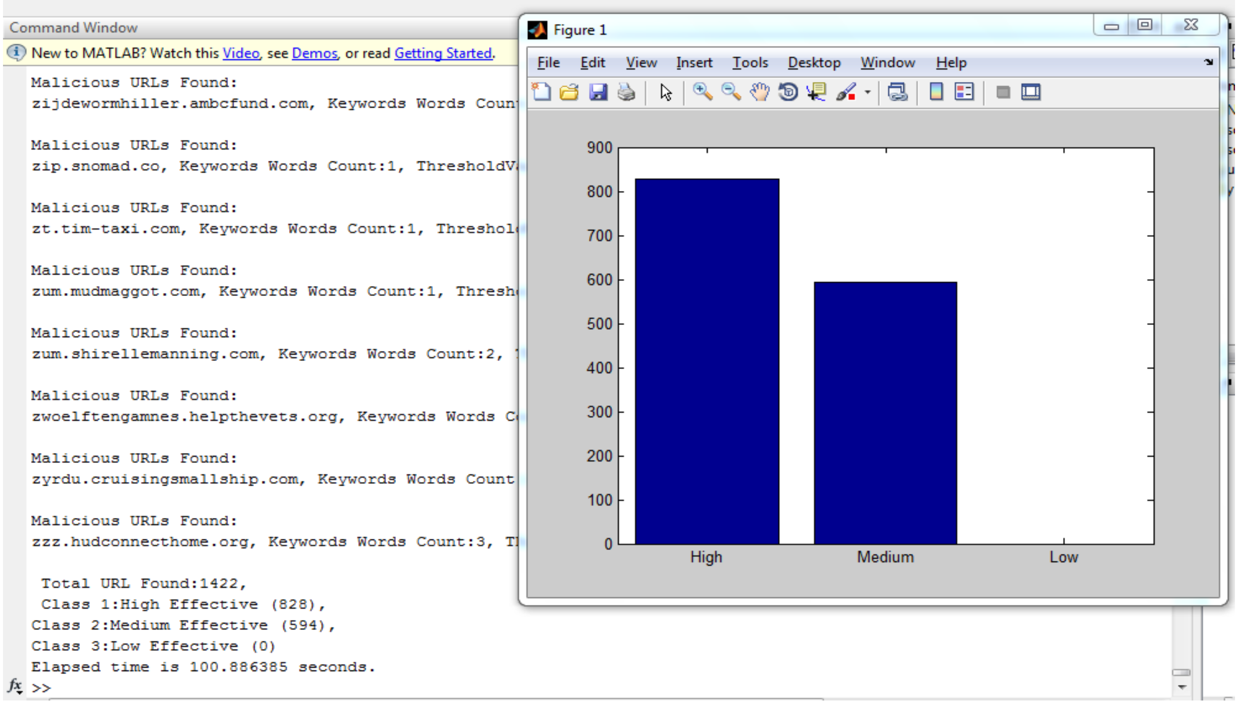
If the input Words found in URL's, they considered as malicious Links.



	A	B
1	1	Wizard
2	2	Thumbs
3	3	Three-way
4	4	Suitcase
5	5	Pimp
6	6	Strong Arm/Hand
7	7	Squirter, Squirting, Squirts
8	8	Solo
9	9	Snowballing
10	10	Skeet
11	11	skeets
12	12	google
13	13	daystresscure
14	14	FREE
15	15	attention

Figure 3: expected words to detect malicious

Output


Command Window
New to MATLAB? Watch this Video, see Demos, or read Getting Started.
Figure 1
File Edit View Insert Tools Desktop Window Help
900
800
700
600
500
400
300
200
100
0
High
Medium
Low

```
Malicious URLs Found:
zijdewormhiller.ambcfund.com, Keywords Words Coun

Malicious URLs Found:
zip.snomad.co, Keywords Words Count:1, ThresholdV

Malicious URLs Found:
zt.tim-taxi.com, Keywords Words Count:1, Threshol

Malicious URLs Found:
zum.mudmaggot.com, Keywords Words Count:1, Thresh

Malicious URLs Found:
zum.shirellemanning.com, Keywords Words Count:2,

Malicious URLs Found:
zwoelftengamnes.helpthevets.org, Keywords Words C

Malicious URLs Found:
zyrdu.cruisingsmallship.com, Keywords Words Count

Malicious URLs Found:
zzz.hudconnecthome.org, Keywords Words Count:3, T

 Total URL Found:1422,
 Class 1:High Effective (828),
Class 2:Medium Effective (594),
Class 3:Low Effective (0)
Elapsed time is 100.886385 seconds.
>>
```

Figure 4: risk level

Output- Number of Words Found

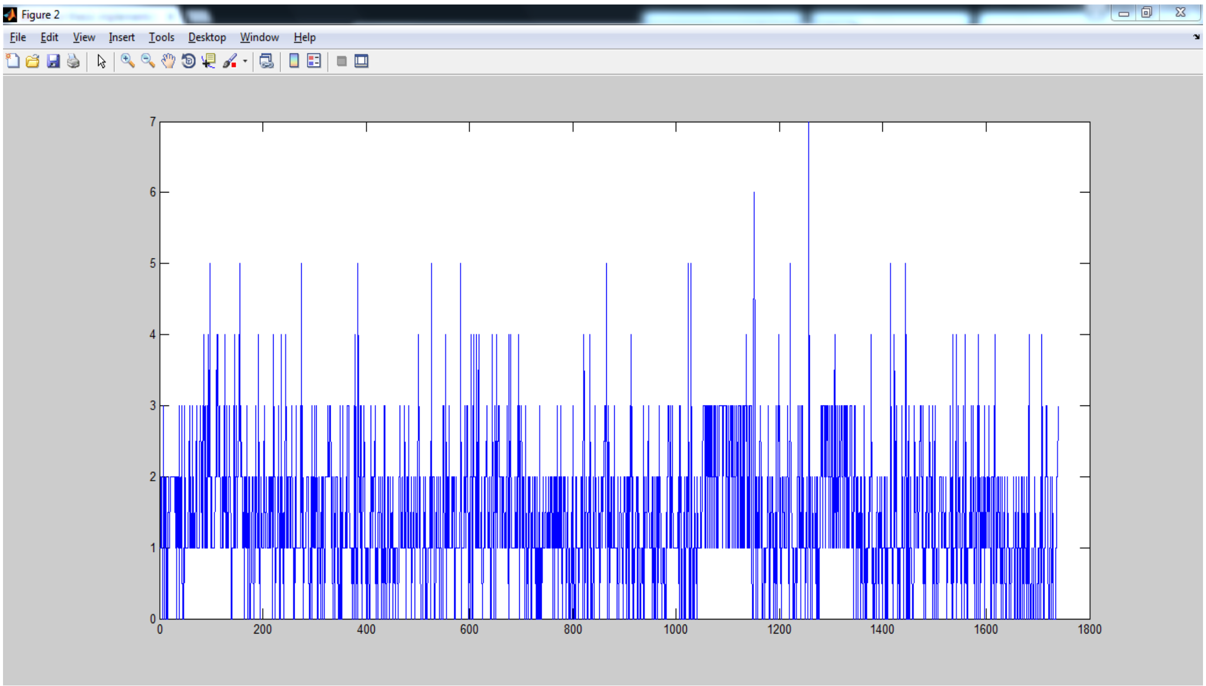


Figure 5: number of words found

V. CONCLUSION

The research work has been implemented for detection and prevention of the Malicious URL's. The web contains large dataset of the URL's and collection contains the mixed type of url such as Malicious, Spam and valid as well. The Malicious URL expands the virus in the system during accessing of URLs and it defects the system's activities. The sharing of URL is most used in the social networking sites in which the user share the URL with people and user don't have any information about the type of URL.

DOI 10.17148/IJARCCE.2016.5712

In such scenarios, the particular URL should be detect based on the parameters. This paper has been explained the different parameters such as toxicity with range 0 to 1, IsWordsCount Spam with true, false condition and number o downloading links which can be any number. Based on these parameters, the weight has been assign correspondingly and developed the equation. The equation result has been then compared by the threshold value. This complete procedure has been implemented in the MATLAB 2010 and results has been plotted in the form of graphs. The categorize URL is counted as High, Low and Medium category. The Proposed work shows accuracy in terms of detection and prevent from unauthorized URLs.

In future, the ANFIS (Adaptive Neuro-Fuzzy Inference system) can be implement with this work and results efficiency can be improved with reducing the time complexity.

REFERENCES

[1] Qing Zhao (2008), "Study on Security of Web-Based Database", Computational Intelligence and Industrial Application, 2008. PACIIA '08, Page (s):902 - 905

[2] Bertino, E.; Sandhu, R. (2005), "Database security - concepts, approaches, and challenges", Dependable and Secure Computing, IEEE Transactions,Page(s): 2 – 19

[3] Ke Wei (2006), "Preventing SQL injection attacks in stored procedures", IEEE.

[4] Gahlaut, Himanshu (2008), "Prevention of Malicious Transactions in DBMS", M.Tech Thesis, Department of computer Science and Engineering, NIT Rourkela.

[5] Baohua Wang (2008), "A Formal Multilevel Database Security Model",IEEE,Page(s):252 - 256

[6] Asmawi, Aziah (2008), "System architecture for SQL injection and insider misuse detection system for DBMS",Information Technology, 2008. ITSim, International Symposium, Page(s):1-6

[7] Elia, I.A. ; Fonseca, J. ; Vieira, M.(2010), "Comparing SQL Injection Detection Tools Using Attack Injection: An Experimental Study", Software Reliability Engineering (ISSRE), IEEE 21st International Symposium, Page(s): 289 – 298.

[8] Ghorbanzadeh, P. et al. (2010), "A survey of mobile database security threats and solutions for it", Information Sciences and Interaction Sciences (ICIS), Page(s): 676 – 682

[9] Mudzingwa, D. (2012), "A study of methodologies used in intrusion detection and prevention systems (IDPS)", Proceedings of IEEE, Page(s):1 - 6

[10] Jianhua Lu et. al. (2012), "A Design of Solution to Database Security Based on Multi-Layer Intrusion Tolerance", Industrial Control and Electronics Engineering (ICICEE), Page(s): 1571 – 1574.

DOI 10.17148/IJARCCE.2016.5712