Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 6, June : 2023

Machine Learning Techniques to Detect Cyber Attacks on Web Applications

B. Hari Kumar
Professor & Dean, ECE
Geethanjali College of Engineering and Technology
Hyderabad,India
harikumar.ece@gcet.edu.in

M. Shivani
Student, IV ECE
Geethanjali College of Engineering and Technology
Hyderabad,India
19r11a04h2@gcet.edu.in

R. Vaishnavi
Student, IV ECE
Geethanjali College of Engineering and Technology
Hyderabad,India
20r15a0417@gcet.edu.in

***Abstract*—** Cyber security faces difficulties as a result of rising cloud service usage, increase in the number of people using web apps, modifications to the network architecture connecting mobile devices, and rapidly advancing network technology. As a result, in order to fulfill the needs and issues of the users, network security methods, sensors, and protection schemes must also be developed accordingly. Preventing new application layer cyber attacks is concentrated in this paper as they are the biggest danger to network and cyber security. The main contribution is the suggestion of using machine learning to model the typical behaviors of applications and to identify cyber attacks. The model is made up of patterns that were discovered through the use of dynamic programming and a graph-based segmentation technique. The model is built using data gathered from HTTP requests made by clients to web servers.

1 Introduction

Commercial organizations are now increasingly using advanced cyber-warfare to harm, interrupt, or block information content in computer networks. It is important to protect network protocols from attacks by strong challengers who can even control a small portion of network participants. The parties under their control have the ability to carry out both passive and active attacks such as jamming, data crashing, . Attack detection is the practice of continuously observing activity within a computer system or network, determining it for indicators of potential incidents, and frequently blocking unauthorized entry. This is often done by automatically gathering data from various systems and network sources, processing the data, and then looking for any security issues. Traditional methods of detecting and preventing intrusions, such as antivirus programmes, access control systems, and encryptions, have certain limits in their ability to completely defend networks and systems against more complex attacks like interruption of service. Moreover, systems developed using these approaches typically have significant inaccurate positive and negative detection rates and also a lack of ongoing adjustment to developing malicious behaviours. To improve detection rates and adaptability, however, a number of Machine Learning (ML) approaches have been applied to the detection of attacks issue during the last decade. In order to keep the attack knowledge bases complete and up to date, multiple techniques are frequently employed. Recent days have seen a rise in the importance of cyber security and protection against multiple cyber attacks. It is therefore clear that developing precise protection strategies such as a computerized Intrusion Detection System (IDS) based on machine learning is important for the security of the system. A system to detect intrusions is, in general, a system or piece of software that monitors a system or network for signs of malicious activity and policy hacking.

1.1 Software Requirements

- Python IDLE
- Virtual Studio Code

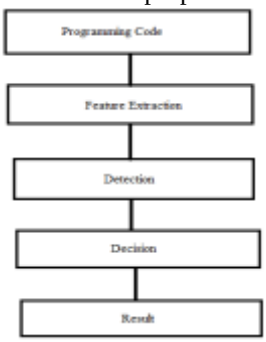
2 Previous Techniques

Large network information volumes, extremely unequal data distribution, the challenge of recognising decision boundaries between normal and abnormal behaviours, and the need for ongoing ability due to an environment that is constantly shifting are the challenges that an IDS often has to deal with. Typically, the difficulty is successfully gathering and classifying various behaviours in a computer network. The two main kinds of network behaviour classification strategies are harm detection and identification of hazards. Applying signature matching algorithms, like harm detection approaches scan network and system activity for cases of known misuse. When used to identify known attacks, this strategy is successful. The intrusion detection systems can send off alerts, but responding to each one expenditures both time and money resulting in the system vulnerable. To solve this issue, IDS must be cooperative enough to gather alerts and make decisions based on their association, rather than beginning the elimination process as soon as the first symptom is identified. But because of how advanced the various cyber security fields are, it will require a lot of effort. The inner inspections authority, however, does not include a large number of cyber security capabilities. This particular structure addresses the need for security, which will be gained through

management reviews, cyber risk evaluations, information management and protection, risk data analytics, crisis handling, and resilience arrangement. It additionally deals with the risk/compliance management, development life cycle, security programme, third-party management, information/asset management, access management, threat/vulnerability management, and need to implement cyber security controls as part of an overall.

2.1 Flow Chart

The flow chart of the proposed technique is shown below:



3 Methodology

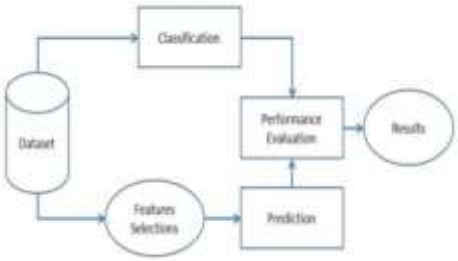


Figure 3.1 Block Diagram

The proposed strategy applies to the framework of machine learning. The identified data is required for establishing the model parameters of the typical application behaviour throughout the learning phase.

Anomaly-Based Methods:

The anomaly-based methods for the cyber attack detection typically build a model that is to describe normal and abnormal behaviour of network traffic. Commonly such methods use two types of algorithms equipped from machine learning theory, they are unsupervised and supervised approach.

Supervised approach:

To create a set of regular expressions that represent regular HTTP requests made by clients to a web application, we recommend applying a graph-based method. The HTTP request's vertices (HTTP Request type, URL, parameters) are represented. Creating a set of regular expressions that describe a usual HTTP request is a recognised issue known as supervised segmentation, where parameters that are comparable to one another are assigned to the same component. We want to combine similar HTTP requests altogether and describe them using a single pattern, to put it another way. The technique is really not limited to the HTTP web page protocol and is easily adaptable to other types of written information, such as various record files produced by the software programme or databases. Needleman-Wunsch method has been designed to calculate the differences between two components.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 6, June : 2023

Dataset Description:

This experiment made use of the dataset collected. It contains a large number of dumps for HTTP protocol queries. The information dumps are in the form of an HTTP protocol which includes data on the methods used. The HTTP header contains parameters such as cache-control, acceptance of charset, etc. and more. This dataset needed to be trained by the use of algorithms and then examine and predict the output status of the HTTP url given.

4 Results

Results are shown below.

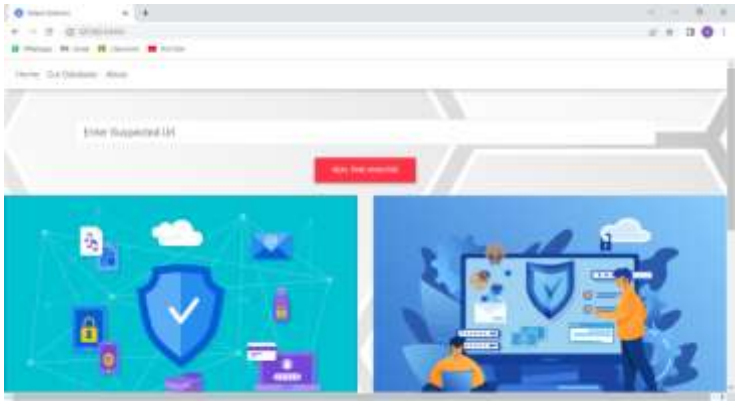


Figure 4.1 Web Page

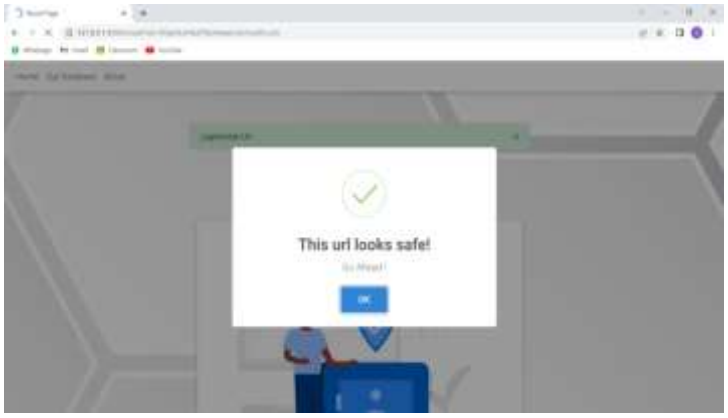


Figure 4.2 Safe URL Output

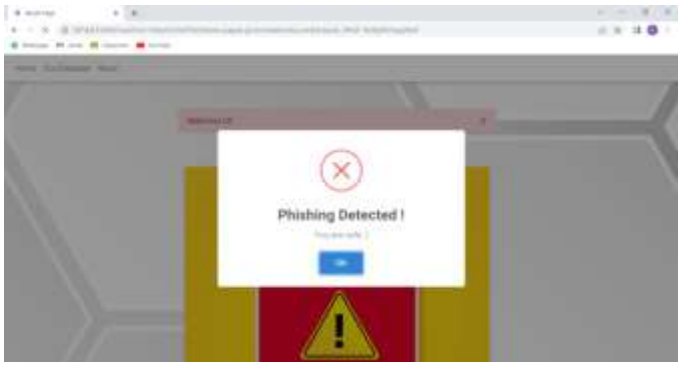


Figure 4.3 Unsafe URL Output

5 Future Scope

In order to use the models that they built for real-time systems in order to benefit from circumstances like attack detection and avoidance in real life, some researchers aim to do this in the future through their research. Online learning and online forecasting are the two phases of real-time Machine Learning (ML) Real-time online prediction refers to such activities. Technology is capable of absorbing current information and instantly modify the representation through online learning. As an outcome, additional researchers may consider transforming intelligent algorithms into systems that operate in real time to be an important field of research.

6 Conclusions

It was suggested to use machine learning as the approach for application layer identification of attacks. A graph-based segmentation approach and dynamic programming methods are used to obtain the patterns which make the model. The prediction of real application behaviours and the identification of online threats and attacks are both done with the use of regular patterns. The efficiency of the suggested algorithm's details which can be successfully applied for the application phase recognition of attacks, was further proven by the information we provided.

7 References

[1] Dipankar Dasgupta. Immunity-based intrusion detection system: A general frame-work. In Proceedings of the 22nd National Information Systems Security Confer-ence (NISSC). Arlington, Virginia, USA, 1999.

[2] Jonatan Gomez and Dipankar Dasgupta. Evolving fuzzy classi_ers for intrusion detection. In Proceedings of the 2002 IEEE Workshop on Information Assurance,West Point, NY, USA, 2002.

[3] Steven A. Hofmeyr, Stephanie Forrest, and Anil Somayaji. Intrusion detection using sequences of system calls. Journal of Computer Security, 6(3):151{180, August 1998}.

[4] Peter Mell Karen Scarfone. Guide to intrusion detection and prevention systems (idps). National Institute of Standards and Technology, NIST SP - 800-94, 2007.

[5] Jamie Twycross, Peter J. Bentley, Uwe Aickelin, Julie Greensmith, and Jungwon Kim. A review of immune system-based intrusion detection techniques was published in Natural Computing, 6(4):413–466, in December 2007.

[6] H. Nguyen, K. Franke, G. Alvarez, S. Petrovic, and C. Torrano-Gimenez. use of the general feature selection method for web assault detection. 2011; LNCS 6694, pp. 25–32; Proceedings of the International Workshop on Computational Intelligence in Security for Information Systems.

[7] E. Menahem, Y. Elovici, and A. Shabtai. F-Sign: Part C: applications and evaluations, automated, function-based signature creation for malware, systems, people, and cybernetics. IEEE Transactions, vol. 41, p. 494-508, 2011.

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

[9] D. Toshniwal and M. Sharma. Utilising variable size buckets, the pre-clustering approach for anomaly detection and clustering. Information Technology Recent Advances, 515–519, 2012.

[10] T. Abrao, M. L. Proenca, J. J. P. C. Rodrigues, M. F. Lima, and M. H. A. C. Adaniya. DSNS and the FireflyHarmonic Clustering Algorithm are used to discover anomalies. (2012) Communications (ICC), 1183–1187.

[11] Y. Labit, P. Owezarski, P. Casas, and J. Mazel. Unsupervised network anomaly detection using sub-space clustering, inter-clustering results association, and anomaly correlation. 24–28 October 2011, Network and Service Management (CNSM), 1–8.

[12] W. Houbowicz, R. Renk, L. Saganowski, and M. Choras. Network traffic identification for anomaly detection using statistics and signals. 2012, Expert Systems, 29, 232-245.

[13] Adewale Olumide S., Adetunmbi Adebayo O., Falaki Samuel O, and K. Boniface. k-nearest neighbour and rough set-based network intrusion detection. 2008, pages 60–66 of International Journal of Computing and ICT Research.

[14] P. Huttenlocher and F. Felzenszwalb. Graph-based picture segmentation that is effective. International Journal of Computer Vision, September 2004, Volume 59, Pages 167–181.

[15] D. Wunsch and B. Needleman Saul Christian a generic approach that may be used to compare the amino acid sequences of two proteins. Molecular Biology Journal, 48, 443–453, 1970.

[16] Z. Zhang, J. Li , C. Manikopoulos, J. Jorgenson and J. Ucles. HIDE: a hierarchical network intrusion detection system using statistical preprocessing and neural network classification. In Proceeding of IEEE Workshop on Information Assurance and Security, 2001.