2013 International Conference on Communication Systems and Network Technologies

Ontology for Detection of Web Attacks

Ashwini D. Khairkar
Department of IT
Bharati Vidyapeeth College of Engineering, Pune
ashkhairkar@gmail.com

Deepak D Kshirsagar
Research Scholar (CSE),
Department of Electronics & Computer Engineering, IIT, Roorkee
kdeepak83@gmail.com

Sandeep Kumar
Deptt of Electronics & Computer Engg
Indian Institute of Technology,
Roorkee, Uttarakhand,India.
sandeepkumargarg@gmail.com

Abstract-Intrusion Detection System (IDS) must reliably detect malicious activity. The expansion of web application also exponentially increases cyber threats. Current survey shows that application layer is more vulnerable to web attacks. There are more than 75% of attacks are deployed at application layer and out of that 90% are vulnerable to attacks. In this paper, we address issues of existing IDS i.e. low false positive rate, low false negative rate and data overload. We discuss about use of semantic web in the Intrusion Detection Systems. This article presents a proposition of using Semantic Web and Ontology concepts to define an approach to analyze Security logs with the goal to identify possible security issues. It extracts semantic relations between computer attacks and intrusions in an Intrusion Detection System. Ontology provides to enable, reuse of domain knowledge and it is also easier to understand and update legacy data.

Keywords: Attacks, Semantic security, Ontology security, Web attacks, Intrusion Detection System

I. INTRODUCTION

A. *Intrusion Detection System*

Intrusion Detection System (IDS) can be either software or hardware based that monitor's network traffic and delivers an alert if it notices malicious activity [1]. Intrusion Detection was introduced in late 1980s [2]. Today, intrusion detection is one of the highest challenging tasks for network administrators and security professionals. As the use of e-business and social networking site increased, so it is exponentially increase in cyber threats. Intrusion Detection System performs three tasks: monitor, detect and responds to malicious activity. Intrusion Detection System use policies to define certain events, if detected will issue alert automatically to the event. Such a response might include logging off a user, disabling a user account and launching of scripts. Any intrusion detection system has some inherent requirements. Its prime purpose is to detect as many attacks as possible with minimum number of false alarms, i.e. the system must be accurate in detecting attacks with good detection rate. However, an accurate system that cannot handle large amount of network traffic and is slow in decision-making will not fulfill the purpose of an intrusion detection system. We wants a system that detect most of the attacks, gives very few false alarms, for large amount of data and its fast enough to make real time decision.

B. *Types of IDS*

There are two types of intrusion detection system: Host based intrusion detection system (HIDS) and Network based intrusion detection system (NIDS)[3].

HIDSs evaluate information found on a single or multiple host systems, including contents of operating systems, system and application files. A HIDS are run on individual hosts or devices on the network. A HIDS monitors the inbound and outbound packets from the device only and will alert the user or administrator if suspicious activity is detected. It consists of an application, generally software, on a machine that is designed to inspect input actions that are internal to the machine like system calls, application and audit logs, file-system modifications, and other host activities and states [4]. For example: Tripwire – It is commercial type of HIDS product that can ensure the integrity of critical data on a wide variety of servers and network devices (*e.g.,* routers, switches, firewalls, and load balancers) called nodes. It does this by gathering system status, Configuration settings, file content, and file metadata on the nodes and checking gathered node data against previously stored node data to detect modifications [4]. NIDS are placed at a strategic point or points within the network to monitor traffic to and from all devices on the network. Ideally, one would scan all inbound and outbound traffic.

For example: Sourcefire 3D Sensors – It is commercial types of product. It is purpose-built network security appliances available with throughputs from 5 Mbps up to 10 Gbps. 3D Sensors running software can be deployed to protect all areas of a network—the perimeter, the DMZ, the core, and critical internal network segments[4].

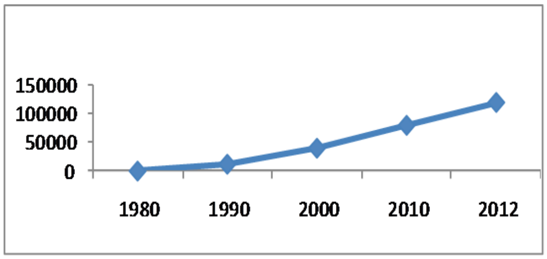

Figure 1.Growth rate of cyber threats

C. *Web Attacks*

According to The National Vulnerability Database (NVD), there are over 18,500 vulnerabilities in the web-based applications, which include 2,147 cross-site scripting (XSS), 2,757-buffer overflow and 1,600 SQL injection vulnerabilities [5]. The vulnerability caused by unchecked input would lead the hacker to inject malicious code to bypass or modify the originally intended functionality of

978-0-7695-4958-3/13 $26.00 © 2013 IEEE
DOI 10.1109/CSNT.2013.131

the program to gain information or unauthorized access to a system. Directory traversal- It is one of the most common attacks. It aims to traverse the directory structure of a Web server to access files that may not be public [6]. For example:

```
http://host/cgibin/vuln.cgi?page=../../../../bin/ls%20-
```

In above example, request for a full directory listing of the “etc” directory within a UNIX system. It is possible to make different variations on these attacks in order to fool conventional IDS. Ontology system will provide solution to this type of problem, if system will detect input consist of .. and / then it will classified as malicious activity and error message is generated. Cross-Site Scripting (XSS)-It can be used by an attacker to compromise the same origin policy of client-side scripting languages, as JavaScript. XSS occurs when a Web application unknowingly gathers malicious data on behalf of a user, usually in the form of a hyperlink. Usually the attacker will encode the malicious portion of the link to the site so the request looks less suspicious. After the data is collected by the Web application, it usually creates an output page for the user containing the malicious data that was originally sent to it, but in a manner to make it appear valid content from the website [6].

```
http://www.vulndev.net/<script>document.location='htt
p://www.repository.com/cgi-bin/cookie.cgi?
'%20+document.cookie </script>
```

hich void vide solution. If input consist of < , > , script , document and cookie it will classified as malicious activity and error message is generated.

SQL Injection- It is a kind of vulnerability where the attack tries to manipulate data base applications by issuing crafted SQL queries [6].

For example:

```
SELECT * FROM users WHERE username = ‘abc’
AND password = ‘anything’ OR ‘x’ = ‘x’; the
expression ‘x’ = ‘x’
```

In above example x=x is true so OR return true and if intruder know only the username still he get the access to database. Now he get access to database either performs UPDATE, DELETE, MODIFY operation through malicious input. Ontology system will detect this type of attack, it will classify as malicious activity and error message is generated.

II. LITERATURE SURVEY

There are various approaches to implement an intrusion detection system based on its type and mode of deployment. New techniques keep emerging, which will remove the limitation of the previous methods of implementation. There are many IDS currently available and some of them are listed as follows:

Traditionally, intrusion detection system can be classified as signature based and anomaly based system –

Signature or Misuse detection based system - It is a rule-based approach that uses stored signatures of known intrusion instances to detect an attack. This approach successfully detects previously know attacks. However, it fails to detect new attack types and variants of known attacks whose signatures are not stored. When new attacks occur, the signature database has to be manually modified for future use.

For example: Snort is configured with over 2500 signature rules to detect scans and attacks [8] novel approaches to re-structure the signature rules. It is a free and open source network intrusion detection and prevention system, was introduced by Martin Roesch, 1998 and now developed by Source fire. Snort detects thousands of worms, vulnerability exploit attempts, port scans, and other suspicious behavior Anomaly detection approach - In this approach, first it established a profile for normal behavior [7]. Then deviants from the normal profile are identifying as anomalies. In some cases, these anomalies may be just normal operations even though if there are some deviants from normal profile. Therefore, in such cases the anomalies may be showing high false positives rate.

Data Mining - Lee et al. introduced data mining approaches for intrusions detection in [9],[10], and [11]. It includes association rules and frequent episodes, which are based on building classifiers and by discovering relevant patterns of program and user behavior. Association rules and frequent episodes are used to learn the record patterns that describe user behavior. These methods can deal with symbolic data, and the features can be defined in the form of packet and connection details. However, mining of features is limited to entry level of the packet and requires the number of records to be large and sparsely populated; otherwise, they tend to produce a large number of rules and it increase the complexity of the system.

Data Clustering Algorithm - Data clustering methods such as the k-means and the fuzzy c-means have also been applied for intrusion detection [12] and [13]. One of the main drawbacks of the clustering technique is that it is based on calculating numeric distance between the observations. Hence, the observations with symbolic features cannot be easily used for clustering so observations must be numeric. So it results in inaccuracy. In addition, the clustering methods consider the features independently and are unable to capture the relationship between different features of a single record, which further degrades attack detection accuracy.

Naive Bayes - It used for intrusion detection [14]. However, they make strict independence assumption between the features in an observation resulting in lower attack detection accuracy when the features are correlated, which is often the case for intrusion detection.

Decision Tree - It used for intrusion detection [15]. The decision trees select the best features for each decision node during the construction of the tree based on some well-defined criteria. i.e. to use the information gain ratio. Decision trees have high speed of operation with high attack detection accuracy. Artificial Neural Network - Debar et al. [16] and Zhang et al. [17] discuss the use of ANN for network intrusion detection. As a pattern recognition technique, Pattern recognition can be implemented by using a feed-forward neural network that has been trained accordingly. During training, the neural network parameters are optimized to associate outputs (each output represents a class of computer network connections, like normal and attack) with corresponding input patterns (every input pattern is represented by a feature vector extracted from the characteristics of the network connection record). When the neural network is used, it identifies the input pattern and tries to output the corresponding class. When a connection record that has no output associated with it is given as an input, the neural network gives the output that corresponds to a taught input pattern that is least different from the given pattern. Though the neural networks can work effectively with noisy data, they require large amount of data for training and it is often hard to select the best possible architecture for a neural network.

Support vector Machine - Support vector machines have also been used for detecting intrusions [18]. Support vector machines map real valued input feature vector to a higher dimensional feature space through nonlinear mapping and can provide real-time detection capability, deal with large dimensionality of data, and can be used for binary-class as well as multiclass classification. Genetic Algorithm - Genetic algorithms were started in the field of computational biology. After that, they have been applied in various fields with promising results. Das Gupta and Gonzalez [19] used a genetic algorithm, they were examining host based IDSs only. Li [20] used GA to detect anomalous network intrusion .In this approach it includes both quantitative and categorical features of network data for deriving classification rules. However, no experimental results are available as the inclusion of quantitative feature can also increase detection rate. Goyal and Kumar [21] used a GA based algorithm to classify all types of smurf attack. It used the training dataset with very low false positive rate (at 0.2%) and almost 100% detection [22]. Gong et al. [23] presented an implementation of GA based approach to Network Intrusion Detection using GA and showed software implementation. The approach derived a set of classification rules and utilizes a support-confidence framework to judge fitness function. Ontology Based IDS - Ontology based IDS solutions are used in information security. Raskin et al [24] developed the ontology for data integrity of web recourses and Denkeretal [24] drive the control access through ontology developed in DAML+OIL[24] but these Ontology has not been fully utilize due to simple representation of attack attributes so inefficient for intrusion detection. In [6] a better approach developed through ontology, for grasping the domain knowledge of application and [27,28] adopted the good approach but carrying overhead due to lack of search space reduction. Ontology based IDS using Bayesian Filter is effective in detecting the vulnerabilities and zero day attacks as compare with other traditional system along with low false rate but response time of system is negligible high as compare with other IDS

Most existing intrusion detection systems suffer from the following problems:

Data overload: This aspect, which does not relate directly to misuse detection but is extremely important, what amount of data an analyst can efficiently analyze. The amount of data to be checked is growing rapidly and there is the possibility for size of logs to reach millions of records per day.

False positives: A common complaint is that more number of false positives IDS will generate. A false positive occurs when normal attack is mistakenly classified as malicious.

False negatives: This is the case where an IDS does not generate an alert when an intrusion is actually taking place. Ontology with semantic can help to improve intrusion detection by addressing each and every one of the above mentioned problems.

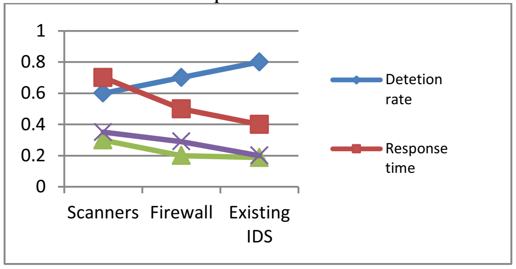

Figure 2: Comparison of Detection rate , False Positive Rate ,False Negative Rate and Response time with traditional IDS and existing systems.

III. ONTOLOGY FOR INTRUSION DETECTION

There are various definitions found in literature about what is ontology. Originally, the term was born in the field of philosophy, being a word of Greek origin, which deals with the nature of being and its existence. Below are some of the most used definitions for the term ontology:

- According to Gruber [29], "ontology is a formal and explicit specification of a shared conceptualization."
- The W3C consortium [30] defines ontology as: "the definition of terms used to describe and represent an area of knowledge."

The basic components of ontology are classes (organized in taxonomy), relations (used to connect the domain concepts), axioms (used to model sentences that are always true), properties (describe characteristics common to the instances of a class or relationships between classes) and instances (used to represent specific data). Ontology is used for modeling data from specific domains and also allows inferences to discover implicit knowledge in these. More specifically, in this work, we are interested in building ontology for the representation of data available in Security logs of web applications. In this context, Ontology can be useful for improving the classification of the attacks occurred and the identification of related events. An ontological representation of knowledge provides many benefits over simple string matching techniques and mitigates the attack through reason and intelligent decision. Ontology driven software system are capable to show a shared understanding of structured information about the concepts within specific domain and provide the reasoning and greater ability to analyze the information automatically. Ontology also specifying the various semantic relationships among different concepts, mitigating the interoperability issue and being reused and evolve overtime. Ontology file is stored with .owl extension which is accessible in Java platform through Jena API.

CONCLUSIONS

In this paper, we have presented the brief overview of various security techniques. The basic of identifying and recognizing threat with high accuracy and active response mainly concerns enabling comprehensive attack coverage available that must exists at present. It has been observed that these intrusion detection systems are not adequate for protecting from intruders efficiently and figured out that, a semantic based intrusion detection system capable of making intelligent decision based on the context of the target domain is required. This study aimed to take the steps in using ontology for identifying web attacks. Furthermore, improvement with Ontology system, testing and benchmarking it with others in real network traffic, will be made in future works. The amount of recognized threat is proposed to rise with correlation accuracy alarm, risk rating and active response. It is believed that Ontology system could be an effective solution for building an integrated system in the industrial world by combining Firewall and IDS features.

REFERENCES

[1] K.K. Gupta, B. Nath, R. Kotagiri, and A. Kazi, “Attacking Confidentiality: An Agent Based Approach,” Proc. IEEE Int’l Conf. Intelligence and Security Informatics (ISI ’06), vol. 3975, pp. 285-296, 2006.

[2] J.P. Anderson, Computer Security Threat Monitoring and Surveillance,http://csrc.nist.gov/publications/history/ande80.pdf, 2010.

[3] R. Bace and P. Mell, Intrusion Detection Systems, Computer Security Division, Information Technology Laboratory, Nat’l Inst. of Standards and Technology, 2001.

[4] Revision by Tzeyoung Max Wu, Information Assurance Technology Analysis Center (IATAC),6th Edition, Defense Technical Information Center, *OMB No. 0704-0188*

[5] National Vulnerability Database (NVD), http://nvd.nist.gov

[6] Abdul Razzaq, Ali Hur, Hafiz Farooq Ahmad, Nasir Haider “Ontology Based Application Level Intrusion Detection System by using Bayesian Filter” The 2nd IEEE International Conference on Computer, Control & Communication (IEEE-IC4) 2009, PNEC Karachi, Pakistan.

[7] C.Kruegel, T.Toth, and E.Kirda. ”Service-specific Anomaly Detection for Network Intrusion Detection”.

[8] M.Roesch. “Snort – Lightweight Intrusion Detection for Networks”. *Proceedings of the USENIX LISA’99 Conference*, November 1999.

[9] W. Lee and S. Stolfo, “Data Mining Approaches for Intrusion Detection,” Proc. Seventh USENIX Security Symp. (Security ‟98), pp. 79-94, 1998.

[10] Lee, S. Stolfo, and K. Mok, “Mining Audit Data to Build Intrusion Detection Models,” Proc. Fourth Int‟l Conf. Knowledge Discovery and Data Mining (KDD‟98), pp. 66-72, 1998.

[11] W. Lee, S. Stolfo, and K. Mok, “A Data Mining Framework for Building Intrusion Detection Model,” Proc. IEEE Symp. Security and Privacy (SP ‟99), pp. 120-132, 1999.

[12] L. Portnoy, E. Eskin, and S. Stolfo, “Intrusion Detection with Unlabeled Data Using Clustering,” Proc. ACM Workshop Data Mining Applied to Security (DMSA), 2001.

[13] H. Shah, J. Undercoffer, and A. Joshi, “Fuzzy Clustering for Intrusion Detection,” Proc. 12th IEEE Int‟l Conf. Fuzzy Systems (FUZZ-IEEE ‟03), vol. 2, pp. 1274-1278, 2003.

[14] N.B. Amor, S. Benferhat, and Z. Elouedi, “Naive Bayes vs. Decision Trees in Intrusion Detection Systems,” Proc. ACM Symp. Applied Computing (SAC ‟04), pp. 420-424, 2004.

[15] Y.B. Reddyl, R. Guha, “Intrusion Detection using Data MiningTechniques,”Artificial Intelligence and Applications (AIA-2004), pp. 232-241, 2004.

[16] H. Debar, M. Becke, and D. Siboni, “A Neural Network Component for an Intrusion Detection System,” Proc. IEEE Symp. Research in Security and Privacy (RSP ‟92), pp. 240-250, 1992.

[17] Z. Zhang, J. Li, C.N. Manikopoulos, J. Jorgenson, and J. Ucles, “HIDE: A Hierarchical Network Intrusion Detection System Using Statistical Preprocessing and Neural Network Classification,” Proc. IEEE Workshop Information Assurance and Security (IAW ‟01),pp. 85-90, 2001.

[18] D.S. Kim and J.S. Park, “Network-Based Intrusion Detection with Support Vector Machines,” Proc. Information Networking, Networking Technologies for Enhanced Internet Services Int‟l Conf. (ICOIN ‟03), pp. 747-756, 2003.

[19] Dasgupta, D. and F. A. Gonzalez, ”An intelligent decision support system for Intrusion detection and response”, . In Proc. Of International Workshop on Mathematical Methods, Models and Architectures for Computer Networks Security (MMM-ACNS),St.Petersburg. Springer- , 21-23 May,2001.

[20] W. Li, “Using Genetic Algorithm for Network Intrusion Detection”. “A Genetic Algorithm Approach to Network Intrusion Detection”. SANS Institute, USA, 2004.

[21] Anup Goyal, Chetan Kumar, “GA-NIDS: A Genetic Algorithm based Network Intrusion Detection System”, 2008.

[22] B.Abdullah, I. Abd-alghafar, Gouda I. Salama, A. Abd-alhafez, “Performance Evaluation of a Genetic Algorithm Based Approach to Network Intrusion Detection System”, 2009.

[23] R. H. Gong, M. Zulkernine, P. Abolmaesumi, “A Software Implementation of a Genetic Algorithm Based Approach to Network Intrusion Detection”, 2005.

[24] Raskin, C.F. Hempelmann, K.E. Triezenberg, Nirenburg,“Ontology in Information Security: A Useful Theoretical Foundation and Methodological Tool,” *Proceedings of the 2001 Workshop on New Security Paradigms (NSPW-2001)*,pp. 53-59, 2001.

[25] Raskin, C.F. Hempelmann, K.E. Triezenberg, Nirenburg, “Ontology in Information Security: A Useful Theoretical Foundation and Methodological Tool,” *Proceedings of the 2001 Workshop on New Security Paradigms (NSPW-2001)*, pp. 53-59, 2001.

[26] G. Denker, L. Kagal, T. Finin, M. Paolucci, K. Sycara, ”Security for DAML Web Services: Annotation and Matchmaking,” *The Semantic Web (ISWC 2003)*, LNCS 2870, Springer, 2003.

[27] DAML+OIL.Availableat:http://www.daml.org/2000/12/daml+oil.dam.

[28] J. Undercoffer, J., Pinkston, A. Joshi, T. Finin, “Target- Centric Ontology for Intrusion Detection,” *IJCAI Workshop on Ontologies and Distributed Systems (IJCAI'03)*, August,2003.

[29] T. Gruber and R.Toward, principles for the design of ontologies used for knowledge sharing. In Formal Ontology in Conceptual Analysis and Knowledge Representation. Kluwer Academic Publishers, 1996

[30] Stanford Jambalaya plug-in, http://protege.stanford.edu/plugins/jambalaya/jambalaya-simplebackup.htm, Accessed on 10/10/2011.