Journal of Computer Science

Original Research Paper

A Survey of Anomaly Detection Using Data Mining Methods for Hypertext Transfer Protocol Web Services

[1]Mohsen Kakavand, [1]Norwati Mustapha, [1]Aida Mustapha, [1]Mohd Taufik Abdullah and [2]Hamed Riahi

[1]*Faculty of Computer Science and Information Technology, University Putra Malaysia, 43400 UPM Serdang, Selangor Darul Ehsan, Malaysia,*
[2]*Exact Asia Development Center*

Article history
Received: 26-04-2014
Revised: 10-07-2014
Accepted: 01-11-2014

Corresponding Author:
Mohsen Kakavand
Faculty of Computer Science and Information Technology,
University Putra Malaysia, 43400 UPM Serdang, Selangor Darul Ehsan, Malaysia
Email: kakavandirit@gmail.com

Abstract: In contrast to traditional Intrusion Detection Systems (IDSs), data mining anomaly detection methods/techniques has been widely used in the domain of network traffic data for intrusion detection and cyber threat. Data mining is widely recognized as popular and important intelligent and automatic tools to assist humans in big data security analysis and anomaly detection over IDSs. In this study we discuss our review in data mining anomaly detection methods for HTTP web services. Today, many online careers and actions including online shopping and banking are running through web-services. Consequently, the role of Hypertext Transfer Protocol (HTTP) in web services is crucial, since it is the standard facilitator for communication protocol. Hence, among the intruders that bound attacks, HTTP is being considered as a vital middle objective. In the recent years, an effective system that has attracted the attention of the researchers is the anomaly detection which is based on data mining methods. We provided an overview on four general data mining techniques such as classification, clustering, semi-supervised and association rule mining. These data mining anomaly detection methods can be used to computing intelligent HTTP request data, which are necessary in describing user behavior. To meet the challenges of data mining techniques, we provide challenges and issues section for intrusion detection systems in HTTP web services.

Keywords: Data Mining, Intrusion Detection Systems, Anomaly Detection, Hypertext Transfer Protocol (HTTP), Web Services

Introduction

Web application is another name of the software inside the server that produces real time content of the web (Igino Corona, 2010). Many of the online activities and networks such as, internet banking, social networks, email and search engines are nowadays utilized as web applications or browsers which communicate with web services by making use of HTTP protocols. Actually HTTP client is supposed to ask for a certain response from the web service or HTTP server and receive the response (Kapodistria *et al*., 2011). Consequently, the standard communication protocol of web services is making a widespread use of HTTP (Torrano-gimenez *et al*., 2010). Then it is clear that HTTP protocol is an important medium objective for intruders that bound attacks for web services, since this HTTP-based services administer the online operations that deal with the extensive load of data (Yang *et al*., 2009). That is why hackers have been able to penetrate into many web services, as well as, the networks of the companies with high profile (Barot and Toshniwal, 2012).

The fact is that many of the web applications have been created not by experts, but by people with poor security abilities (Robertson *et al*., 2006). Based on data provided by CERT/CC, vulnerabilities and number of virtual attacks have highly raised up from 1998 to 2002. Malek and Hamantiz (2004), indicated that in spite of

Science Publications

the reports claiming that there have been little intrusion in the early 1990s, the reports show 25,000 of intrusions happened in 2000 which means the intrusions have highly increased. The number of attacks happened from 1990 to 2010 are illustrated in Fig. 1. Both the list of Common Vulnerabilities and Exposures (CVE) (Christey and Martin, 2007) and the recent research on security issues in the digital network of the world show that 25 percent of the total security threats were related to web application vulnerabilities (Vasudevan *et al*., 2011).

In contrast to traditional Intrusion Detection Systems (IDSs), data mining anomaly detection methods/techniques has been widely used in the domain of network traffic data for intrusion detection and cyber threat. Data mining is widely recognized as popular and important intelligent and automatic tools to assist humans in big data security analysis and anomaly detection over Intrusion Detection Systems (IDSs). This study reviews the data mining methods used by HTTP web services anomaly detection, concentrating on which aspects of the HTTP traffic are analyzed and what selection methods have been used, which attempts to classify HTTP traffic as normal or anomalous.

Intrusion Detection System (IDS) And Web Services

Anderson (2002) proposed an Intrusion Detection Systems (IDSs) to the world to confront the increasing number of cyber attacks. Intrusion detection and firewall come next to each other in security issues since that time. Intrusion detection system controls networks and users, monitors vulnerabilities and configurations of the systems and informs the administrators in the case of attacks. IDSs are divided into two groups according to the type of intrusion detection and the type of system. For example, misuse detection versus anomaly detection systems are different in their intrusion detection and network-based versus host-based detection systems are divided according to their type of systems.

The network-based intrusion detection systems (NIDS) identify the intrusions in the networks while the computers are connected to internet and other networks at the same time. However, data may be lost during the detection procedure and hence, accuracy is a big issue. Moreover, since data cannot be decrypted at the network level, encrypted data in NIDS can be an issue (Nadiammai and Hemalatha, 2012; Du, 2006; Khalilian *et al*., 2011).

In the host-based intrusion detection Systems (HIDS), every single host at host level will go through the intrusion detection. Comparing to network-based ones, host-based systems have less problems, although for each single host there should be an attached IDS. The advantage of host-based system is that it has the capability of detecting the attacks that was not possible through network-based intrusion detection systems and also, it can work in an encrypted network setting (Nadiammai and Hemalatha, 2012; Du, 2006; Khalilian *et al*., 2011).

Regarding the intrusion detection type of systems, IDSs are divided into two groups of Anomaly-based Detection (AD) and signature or Misuses-based Detection (MD). Misuse detection systems create a special model or 'signature' for each known attack and any new activity that looks like a 'signature' in the list would be caught by the system. These systems are working well with the already known intrusions, however, they fail to identify the new ones (Wang and Stolfo, 2004).

The research came up with a new generation of IDS, that is, anomaly detection systems which work based on the assumption that every anomalous activity is doubted and needs to be checked. These systems are already planned with a normal or expected model or activity and will detect any deviation of the normal model as a possible attack or security threat. Anomaly-based detection has the privilege of detecting novel or unknown attacks and hence, is usually considered a more effective and powerful system (Estévez-Tapiador *et al*., 2004). The cons and pros of the anomaly detection systems are shown in Table 1 (Saboori *et al*., 2010).

Intrusions detections of web services are not that detailed and complicated and are one of the activities of Web Application Firewalls (WAF). HTTP or Application level protocols are usually utilized by web services for data transmission between service providers and users. Communication between applications is done through XML messages via HTTP (Vorobiev and Han, 2006). Therefore, the corresponding protocols ports (i.e., port 80) should be easily accessed so that the activities and services could be run.

Table 1. Misuse detection Vs. anomaly detection

Approaches	Advantages	Disadvantages
Misuse detection	Accurately and generate must fewer false alarm	Cannot detect novel attacks and threats
Anomaly detection	Is able to detect unknown attacks based on audit	High false alarm and limited by training data

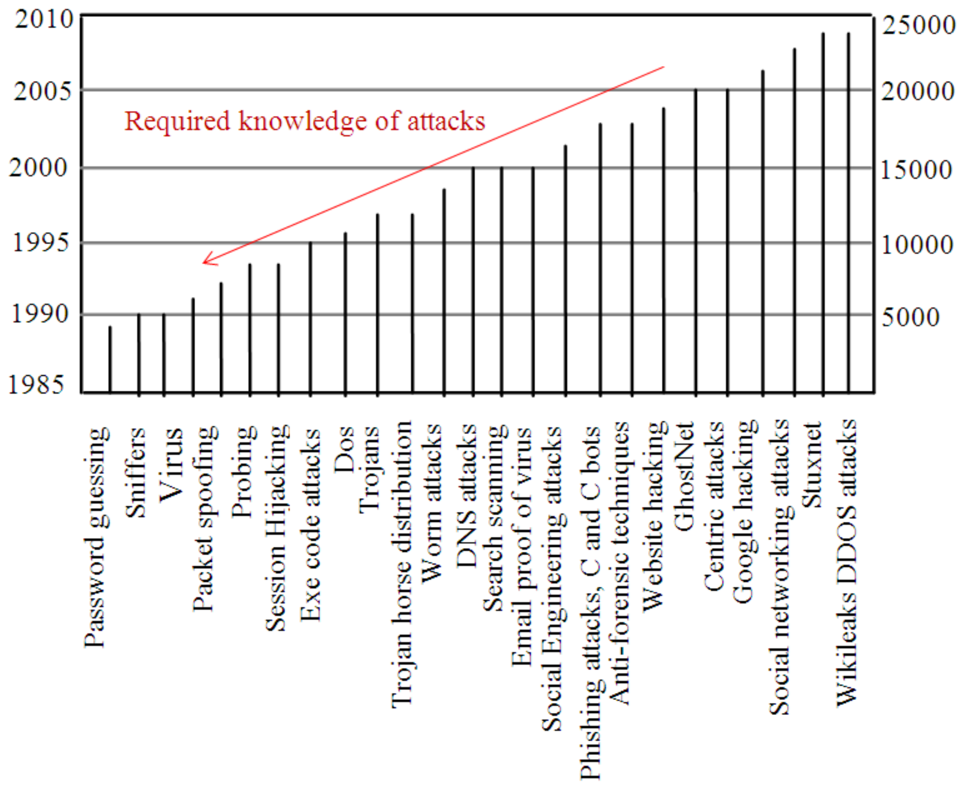


Fig. 1. Security Vulnerabilities and Threats (Malek and Harmantzis, 2004; Vasudevan *et al*., 2011)

That is the reason hackers make use of ports to easily find a way through the network. Actually, security services such as, normal network intrusion prevention or detection systems, firewalls and content filters many times cannot prevent the intruders' access to web services.

The main technology of web services is XML that can be understood by web service IDSs, meaning that, web services are considered as the main parts of today's web applications and hence, intrusion detection systems are helpful in protecting the web services (Najjar and Azgomi, 2010). Firewalls can secure the organizations network, but they have two open TCP ports that are utilized for transmitting requests in web services (for HTTP port 80 and for HTTPS port 447) (Karnwal *et al*., 2012). Data mining approaches for intrusion detection system is designed as in-depth defense mechanism and acts behind the firewalls within the security structure of an enterprise.

Data Mining Based Anomaly Detection Methods

Data mining is a multidisciplinary field refers to a technique to intelligently and automatically assist humans in analyzing the large volumes of data to identify valid, novel and potentially useful patterns in data. Data mining, also popularly referred to as Knowledge Discovery from Data (KDD), is the automated or convenient extraction of patterns representing knowledge implicitly stored or captured in large databases, data warehouses, the Web, other massive information repositories, or data streams (Han *et al*., 2011).

Data mining takes advantage of advances in the fields of Artificial Intelligence (AI), machine learning and statistics in order to help in pattern recognition and classification. Other disciplines used in data mining include rule-based and case-based reasoning, fuzzy logic and neural networks, knowledge-based systems, high-performance computing, information retrieval and data visualization. The Data mining methods/techniques used in anomaly detection systems include classification, clustering, semi-supervised and association rule mining and so on. Various supervised and unsupervised algorithms used by researchers for intrusion detection with varying degree of accuracy are reviewed here. The following is an overview of data mining-base anomaly detection methods that was utilized to detect anomalous attacks.

Classification-based Anomaly Detection

By simple definition, in classification analyze a set of data and generate a set of grouping rules which can be used to classify future data or predict future data trends that sometimes called supervised learning. Several major kinds of classification method including decision tree induction, Bayesian networks, k-nearest neighbor classifier. In this case, main idea is build a classification model for normal and anomalous events based on labeled training data with require knowledge of both normal and anomaly (attacks) class. The learned model is then applied on the test dataset in order to classify unlabeled records into normal and anomalous records in order to classify each new unseen event (Amer *et al.*, 2013) (Fig. 2).

A RIPPER classifier method was suggested (Lee and Stolfo, 2001; Lee *et al.*, 2002) to induce rules from the data by employing a divide- and-conquer approach and involving either discarding or pruning some of the learnt rules is carried out to increase the classifier accuracy. RIPPER classifier has been successfully used in data mining based anomaly detection algorithms to classify incoming audit data and detect intrusions.

Clustering-based Anomaly Detection

Is second learning approach which called unsupervised learning. Here, the data does not contain any labeling information and no separation into training and testing phase is given. Unsupervised learning algorithms assume that only a small fraction of the data is anomaly and that the attacks exhibit a significantly different behavior than the normal records.

In many practical application domains, the unsupervised learning approach is particularly suited when no labeling information is available. Moreover, in some applications the nature of the anomalous records is constantly changing, thus obtaining a training dataset that accurately describe anomaly is almost impossible. On the other hand, unsupervised anomaly detection is the most difficult setup since there is no decision boundary to learn and the decision is only based on intrinsic information of the dataset (Amer *et al.*, 2013) (Fig. 3).

An automatic and unsupervised payload-based anomaly detector is called PAYL (Wang and Stolfo, 2004). PAYL model is data mining-based anomaly detection and very efficient fashion for HTTP traffic data for get effectiveness of accuracy and false positive rate for port 80 traffic. Wang and Stolfo (2004) describe the use of clustering learning for anomaly detection based on Mahalanobis Distance Map (MDM). The procedure is first to measure a profile during a training phase by the frequency distribution and their standard deviation of the application payload moving to a single port and host and then, to use of MDM during the detection stage to compare the new and pre-computed data. Character distributions, buffer-overflow attacks usually have unique character distributions. A character distribution metric was applied on similarly-sized packages by Wang and Stolfo (2004).

Some new designs were introduced by Dokas and Ertoz (Dokas *et al.*, 2005), some data mining method for network intrusion detection systems that were called density based local outliers (LOF) and unsupervised support vector machine (SVM). This anomaly detection made clustering models for identifying known intrusions and anomaly detection schemes for detecting unknown cyber attacks. The research on real-time data suggested that unsupervised SVMs were working wonderfully in identifying the new intrusions; however they used to give too many false alarms rate. Also, the LOF technique proved to be the best in identifying new intrusions.

In another research (Jamdagni *et al.*, 2013), proposed a new anomaly detection system, RePIDS model (an efficient payload-based anomaly intrusion detection system), which is based on pattern recognition technique used in image processing. Unsupervised classification learning as Mahalanobis Distance Map (MDM) is used to discover hidden correlation between the features and among the packet payloads and then Principal Component Analysis (PCA) applied to reduce the dimensionality of feature space and false positive rate by efficient pre-processing of packet payload data. This data mining unsupervised has the important advantage of being simple and fast to compute and this payload-based anomaly intrusion detection system generally quite good at detecting web-based traffic data or HTTP web services.

Semi-Supervised Cluster Analysis

In contrast with supervised learning classification, clustering be without direction from users or classifiers (such as class label value) and therefore may not generate highly worthwhile clusters. The quality of unsupervised clustering can be significantly improved using some weak form of supervision. Such a clustering process based on user feedback or guidance constraints is called semi-supervised clustering.

Methods for semi-supervised clustering can be categorized into two classes: Constraint-based semi-supervised clustering and distance-based semi-supervised clustering. Constraint-based semi-supervised clustering relies on user-provided labels or constraints to guide the algorithm toward a more appropriate data partitioning. This includes modifying the objective function based on constraints, or initializing and constraining the clustering process based on the labeled objects. Distance-based semi-supervised clustering employs an adaptive distance measure that is trained to satisfy the labels or constraints in the supervised data (Han *et al*., 2011).

In semi-supervised anomaly detection approach, where the algorithm models the normal records only. Records that do not comply with this model are labelled as outliers in the testing phase. Advantages of this semi-supervised anomaly detection can be easily understood of Models as well as normal behavior can be accurately learned but possible high false alarm rate - previously unseen (yet legitimate) data records may be recognized as anomalies (Amer *et al*., 2013).

In recent years, some techniques of anomaly detection have been suggested (Perdisci *et al*., 2008) that are semi-supervised anomaly detection. These new techniques of anomaly-based network IDSs are able to identify (anonymous) zero-day attacks, however, the load of false positives that are produced by detection system needs to be taken care of and controlled. A new data mining based anomaly detection system for packet payload data is called McPAD, which is a set of one-class Support Vector Machine (SVM).

Multiple-classifier Payload-based Anomaly Detection (McPAD) is a new accurate data mining-based anomaly detection system that consists of an ensemble of one-class classifiers. It shows that semi-supervised learning is very accurate in detecting HTTP web service attacks that bear some form of shell-code in the malicious payload. This holds true even in the case of polymorphic attacks and for very low false positive rates.

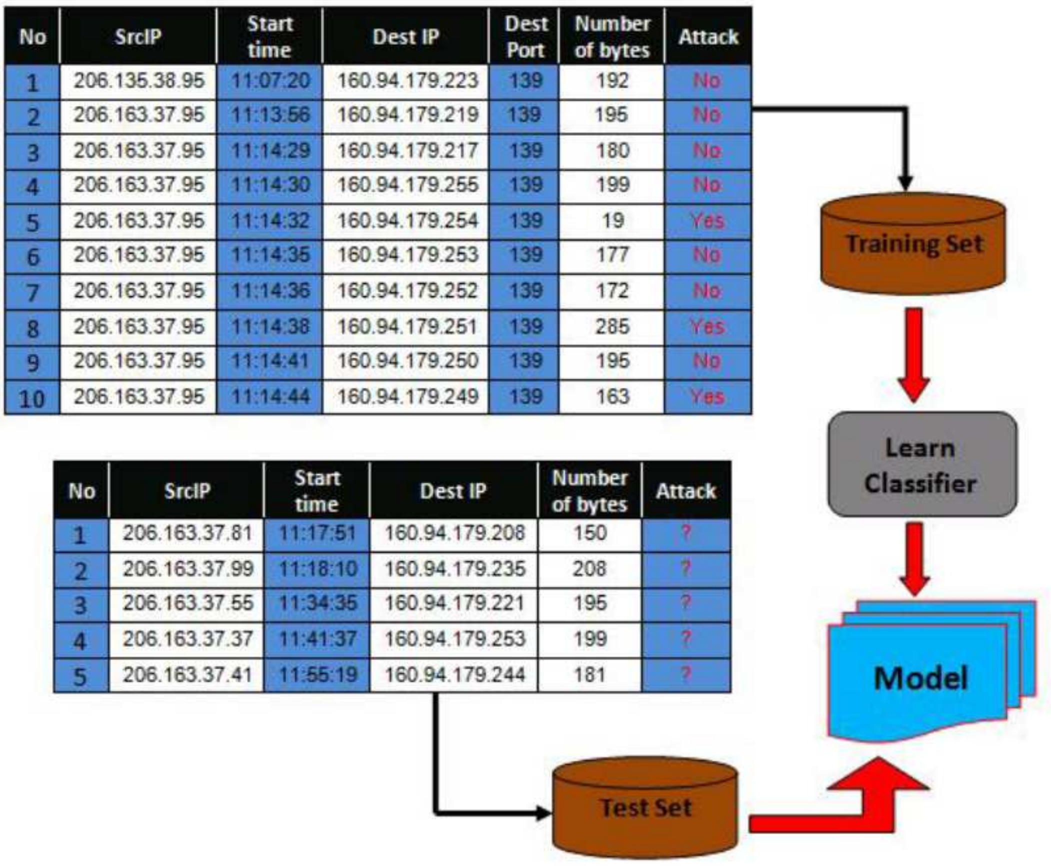
Association Rule Mining

Agrawal *et al*. (1993) are one of many data mining techniques that describe events that tend to occur together. The concept of association rules can be understood as follows: Given a database D of transactions where each transaction T ∈ D denotes a set of items in the database, an association rule is an implication of the form X => Y, where X ⊂ D , Y ⊂ D and X ∩ Y = ∅. The rule X =>Y holds in the transaction set D with confidence c if c% of transactions in X also contain Y. Two important concepts when dealing with association rules are *rule confidence* and *rule support*. The probability of rule confidence is defined as the conditional probability P (Y⊆T | X⊆T) The rule X = >Y has support s in the transaction database D if s% of transactions in D contain X∪Y. Association rules have been successfully used to mine audit data to find normal patterns for anomaly detection in network traffic data. They are particularly important in the domain of anomaly detection because association rules can be used to construct a summary of anomalous connections detected by the intrusion detection system. There is evidence that suggests program executions and user activities exhibit frequent correlations among system features. These consistent behaviors can be captured in association rules.

The model that is used to identify the correlations of the features achieved after pre-processing, is called MADAM ID, for Mining Audit Data for Automated Models for Intrusion Detection (Lee and Stolfo, 2001; Lee *et al*., 2002). MADAM ID consists of classification and meta-classification programs, association rules and frequent episodes program. The model used data mining association rules technique for building intrusion detection systems. In this method, some frequent association rules are used as a sign to find out whether the audit data is sufficient. Then, RIPPER classifier is used to sets of labeled data and the intrusions are figured out. MADAM ID model is type of the systems reconstruct the network packets and extract features that describe the higher level interactions between end hosts. This approach is generally excellent at detecting cyber treats in valid connections.

FARM model was developed by Chan *et al*. (2013) for Simple Object Access Protocol (SOAP) or Extensible Markup Language (XML)-based attacks over HTTP web services. Most research in host and network-based anomaly detection system are only able to detect attacks on the low level of network (computer system) while web service technologies running on higher application level. A Fuzzy Association Rule Model (FARM) is a new data mining anomaly detection system proposed to network security problems, especially for web service based e-commerce applications.

Some new data mining anomaly detection methods introduced for web mining intrusion detection system that called Sensor Web IDS to confront the web intrusions (Ezeife *et al*., 2008). This web mining IDS was an algorithm based on theories of mean and standard deviation and used to compute the greatest possible value length of input parameters. In order to detect the misuse and anomaly intrusions, a method was used for mining the list of frequent parameters and their continuous order called "association rule mining". This data mining model, web mining IDS, which combines power of anomaly and misuse detection by applies association rule mining to find apart from traditional network attacks such as user to root-U2R (e.g., buffer overflow attacks), remote to user attacks-R2U, Denial of Service (DoS) and probes attacks, which are also applicable to web applications, e.g., cross-site scripting attack (XSS), SQL-Injection, session hijacking and cookie poison attacks.



No	SrcIP	Start time	Dest IP	Dest Port	Number of bytes	Attack
1	206.135.38.95	11:07:20	160.94.179.223	139	192	No
2	206.163.37.95	11:13:56	160.94.179.219	139	195	No
3	206.163.37.95	11:14:29	160.94.179.217	139	180	No
4	206.163.37.95	11:14:30	160.94.179.255	139	199	No
5	206.163.37.95	11:14:32	160.94.179.254	139	19	Yes
6	206.163.37.95	11:14:35	160.94.179.253	139	177	No
7	206.163.37.95	11:14:36	160.94.179.252	139	172	No
8	206.163.37.95	11:14:38	160.94.179.251	139	285	Yes
9	206.163.37.95	11:14:41	160.94.179.250	139	195	No
10	206.163.37.95	11:14:44	160.94.179.249	139	163	Yes

No	SrcIP	Start time	Dest IP	Number of bytes	Attack
1	206.163.37.81	11:17:51	160.94.179.208	150	?
2	206.163.37.99	11:18:10	160.94.179.235	208	?
3	206.163.37.55	11:34:35	160.94.179.221	195	?
4	206.163.37.37	11:41:37	160.94.179.253	199	?
5	206.163.37.41	11:55:19	160.94.179.244	181	?

Fig. 2. Data mining anomaly detection model

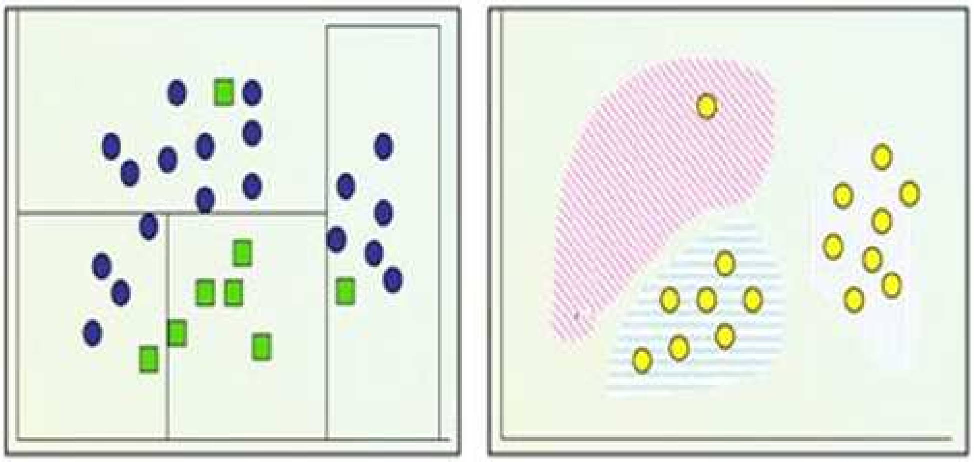


Fig. 3. Classification Vs. clustering

Table 2. Performance of anomaly detection algorithms

Main algorithm	Methodology	Dataset	Detection rate (%)	False positive	Ref.
RIPPER classifier (JRIP)	NIDS/MD	DARPA 1998	80.2	NA	Lee *et al*. (2002)
Mahalanobis Distance Map(MDM)	NIDS/AD	DARPA 1999	98.0	0.1%	Wang and Stolfo (2004)
Density based local (LOF)	NIDS/AD	DARPA 1998	73.7	1%	Dokas *et al*. (2005)
Unsupervised SVM			84.2	4%	
Association Rule Mining (ARM)	HIDS/AD/MD	Private	98.3	NA	Ezeife *el al*. (2008)
One-Class SVM	NIDS/AD	DARPA 1999	95.0	0.01%	Perdisci *el al*. (2008)
Mahalanobis Distance Map (MDM)	NIDS/AD	DARPA 1999	100.0	0.087%	Jamdagni *et al*. (2013)
Fuzzy Association Rule Mining (FARM)	NIDS/AD/MD	Private	99.0	0.1%	Chan *el al*. (2013)

However, intrusion detection systems are still exposed to new problems. The highlighting features available algorithms of intrusion detection types in literature are shown in Table 2.

Discussion

The following data mining anomaly detection algorithms are illustrated in Table 2, the results showed that unsupervised models such as Mahalanobi Distance Map (MDM) (Jamdagni *et al*., 2013) and one-class Support Vector Machine (SVM) (Perdisci *et al*., 2008) had considerably higher accuracy and false positive rate in same dataset. Moreover, data mining association rule technique such as Fuzzy Association Rule Model (FARM) (Chan *et al*., 2013) proved to have high accuracy in recognize anomaly attacks as well as considerable low false positive rate. In spite of the fact that the above approaches proved to have high accuracy, still they all failed to show a high percentage of false positive which is considered a parallel criteria of accuracy as well.

Going through the history of data mining anomaly detection algorithms, one finds out that the two data mining techniques as semi-supervised and unsupervised classification, that is, one-class Support Vector Machine (SVM) and Mahalanobi Distance Map (MDM) have showed rewarding results in same dataset (DARPA 1999 dataset). Consequently, the researcher relies on these two techniques as the standard algorithms for HTTP web service anomaly detection. Furthermore, the researcher assumes that the efficiency of these two techniques should be checked out through private datasets as well. Actually, DARPA/MIT Lincoln Laboratories created and reported the most important datasets for testing IDS (1998-1999) (Haines *et al*., 2001) and is a public dataset that many intrusion detection system researchers make use of its data due to its high amount of data, as well as, its ability to compare directly with the original lab test. Nevertheless, this dataset is reported to have some problems regarding its outdated data and the increasing growth of web behaviors in the course of time. Therefore, unsupervised data mining (Jamdagni *et al*., 2013) and semi-supervised anomaly detection (Perdisci *et al*., 2008) which has shown acceptable results, can be safe for HTTP Web Service request anomaly detection, but the quality of anomalous records is routinely and continuously changing which makes the outliers accurately unattainable for detections over HTTP web service request data. Moreover, high dimensional data always have some possible inaccuracy because of the limitations in quadratic computational complexity (Amer *et al*., 2013).

We provided an overview on four general data mining techniques that such as classification, clustering, semi-supervised and association rules. These data mining anomaly detection methods can be used to computing intelligent HTTP request data, which are necessary in describing user behavior. To meet the challenges of data mining techniques, we provide challenges and issues section for intrusion detection systems in HTTP web services. The challenges and issues related to data mining methods and intrusion detection systems are summarized as follows:

- Research review show that unsupervised and semi-supervised data mining anomaly detection methods tested on public dataset (DARPA1999) resulted in acceptable accuracies; however, detection model was not checked out on different sources of private datasets. Thus, one could hardly ensure the reliability and accuracy of these algorithms because of the public dataset limitations
- Although some data mining clustering techniques in anomaly detection have proven to be successful according to the literature review, many of these studies ignore the continuous alteration in web service data. Therefore, it is not easy to correctly identify the intrusions on HTTP web service request data. Moreover, the possibility of inaccuracy always exists in dealing with high dimensional data due to limitation of quadratic computational complexity
- Most of the studies that reported high accuracy in detection failed to show high rate in false positive. Hence, since false positive is an integrated measure for achieved accuracies, the amount of it that is produced by detection system must be carefully controlled
- Majority of anomaly-based intrusion detections can only identify network layer and computer systems attacks. Moreover, the patterns and behaviors of HTTP web services are changing and growing in the course of time, as well as, the types of attacks

(SOAP/XML over HTTP intrusions) and target features (i.e., the frequency, size, variety). No delicacy testing has been done on payload packet features in HTTP web services and especially the SOAP-based on XML intrusions at the moment

Conclusion

In this review study, we proposed a comprehensive survey of anomaly detection systems using data mining methods/techniques for HTTP web services in the recent past and present. The key ideas are to review data mining techniques to discover consistent and useful patterns of system features over HTTP web services and introduce the set of classifiers learning such as classification, clustering, semi-supervised and association rule mining that can identify anomalies and known attacks. We also generally discussed the review on intrusion detection systems and specially the issues regarding data mining-based anomaly detection. The discussion showed that old data sets are no longer in general use in many cases, since there is an increase in the variety, frequency, complexity and amount of the attacks, while target dataset cannot be publicly accessed to compare and evaluate. Although public data in unsupervised methods have proved high percentage of accuracy in majority of intrusion detection studies, these studies were not repeated with a different set of data to let for comparison and also many of detection accuracies failed to show high percentage of false positive.

Furthermore, in spite of many research on detection of web application attacks, SOAP or XML-based attacks are not fully covered and a large number of data mining anomaly-based detection models can only identify the network layer and computer systems attacks. There is an attempt in this research to introduce a new data mining method specifically for anomaly-based web service intrusion detection on HTTP traffic, based on the discussed issues in this regard. The rate of success obtained in many areas shows the value of investigating the possibility to increase the accuracy and performance of the intrusion detection model while handling SOAP/XML-based attacks.

In our future work we aim to implement data mining anomaly detection methods such classification, clustering and semi-supervised within a particular context of HTTP web services. We also plan to experiment these data mining techniques on 1999 DARPA/MIT Lincoln labs and new HTTP datasets for packet payload.

Acknowledgement

This research work supported by the Fundamental Research Grant Scheme (FRGS) under Ministry of Higher Education, Malaysia and it is gratefully appreciated.

Author's Contributions

All authors equally contributed in this work.

Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and no ethical issues involved.

References

Amer, M., M. Goldstein and S. Abdennadher, 2013. Enhancing one-class support vector machines for unsupervised anomaly detection. Proceedings of the ACM SIGKDD Workshop on Outlier Detection and Description, Aug. 11-14, Chicago, IL, New York, pp: 8-15. DOI: 10.1145/2500853.2500857

Anderson, J.P., 2002. Computer Security Threat Monitoring and Surveillance. 1st Edn., Washington, pp: 1-56.

Agrawal, R., Tomasz Imieliński and Arun Swami, 1993. Mining association rules between sets of items in large databases. Proceedings of the 1993 ACM SIGMOD international conference on Management of data, May 25-28, ACM Press, Washington, DC, pp: 207-216. DOI: 10.1145/170035.170072

Barot, V. and D. Toshniwal, 2012. A new data mining based hybrid network Intrusion Detection model. Proceedings of the International Conference on Data Science and Engineering, Jul. 18-20, IEEE Xplore Press, Cochin, Kerala, pp: 52-57. DOI: 10.1109/ICDSE.2012.6282310

Chan, G.Y., C.S. Lee and S.H. Heng, 2013. Discovering fuzzy association rule patterns and increasing sensitivity analysis of XML-related attacks. J. Netw. Comput. Applic., 36: 829-842. DOI: 10.1016/j.jnca.2012.11.006

Christey, A.S. and R.A. Martin, 2007. Vulnerability type distributions in CVE. MITRE, Common Weakness Enumeration (CWE^{TM}).

Dokas, P., L. Ertoz, V. Kumar, A. Lazarevic and P. Tan *et al.*, 2005. Data mining for network intrusion detection. csee.umbc.edu, 10: 21-30.

Du, W., 2006. Intrusion Detection System. Syracuse University.

Estévez-Tapiador, J.M., P. García-Teodoro and J.E. Díaz-Verdejo, 2004. Measuring normality in HTTP traffic for anomaly-based intrusion detection. Comput. Netw., 45: 175-193. DOI: 10.1016/j.comnet.2003.12.016

Ezeife, C.I., J. Dong and A.K. Aggarwal, 2008. SensorWebIDS: A web mining intrusion detection system. Int. J. Web Inform. Syst., 4: 97-120. DOI: 10.1108/17440080810865648

Haines, J.W., R.P. Lippmann, D.J. Fried, M. Zissman and E. Tran, 2001. 1999 DARPA intrusion detection evaluation: Design and procedures.

Han, J., M. Kamber and J. Pei, 2011. Data Mining: Concepts and Techniques: Concepts and Techniques. 3rd Edn., Elsevier, Burlington, ISBN-10: 0123814804, pp: 744.

Igino Corona, G.G., 2010. Detection of server-side web attacks. JMLR: Workshop Conf. Proc. 11: 160-166. DOI: 10.1145/331499

Jamdagni, A., Z. Tan, X. He, P. Nanda and R.P. Liu, 2013. RePIDS: A multi tier real-time payload-based intrusion detection system. Comput. Netw., 57: 811-824. DOI: 10.1016/j.comnet.2012.10.002

Kapodistria, H., S. Mitropoulos and C. Douligeris, 2011. An advanced web attack detection and prevention tool. Inform. Manage. Comput. Security, 19: 280-299. DOI: 10.1108/09685221111188584

Karnwal, T., T. Sivakumar and G. Aghila, 2012. A comber approach to protect cloud computing against XML DDoS and HTTP DDoS attack. Proceedings of the IEEE Students' Conference on Electrical, Electronics and Computer Science, Mar. 1-2, IEEE Xplore Press, Bhopal, pp: 1-5. DOI: 10.1109/SCEECS.2012.6184829

Khalilian, M., N. Mustapha, N. Sulaiman and A. Mamat, 2011. Intrusion detection system with data mining approach: A review. Global J. Comput. Sci. Technol., 11: 187-187. DOI: 10.11109/ITNG.2010.187

Lee, W. and S.J. Stolfo, 2001. A framework for constructing features and models for intrusion detection systems. ACM Trans. Inform. Syst. Security, 3: 227-261. DOI: 10.1145/382912.382914

Lee, W., S.J. Stolfo and K.W. Mok, 2002. Algorithms for mining system audit data. Data Mining Rough Sets Granular Comput., 95: 166-189.

Malek, M. and F.C. Harmantzis, 2004. Data mining techniques for security of web services. Proceedings of the International Conference on E-Business and Telecommunication Networks, Aug. 24-28, Portugal, pp: 1-14.

Nadiammai, G.V. and M. Hemalatha, 2012. An evaluation of clustering technique over intrusion detection system. Proceedings of the International Conference on Advances in Computing, Communications and Informatics, Aug. 03-05, ACM Pressm Chennai, India, pp: 1054-1060. DOI: 10.1145/2345396.2345565

Najjar, M.S.A. and M.A. Azgomi, 2010. A distributed multi-approach intrusion detection system for web services. Proceedings of the 3rd International Conference on Security of Information and Networks, Sept. 07-11, ACM Press, New York, pp: 238-244. DOI: 10.1145/1854099.1854147

Perdisci, R., D. Ariu, P. Fogla, G. Giacinto and W. Lee, 2008. McPAD: A multiple classifier system for accurate payload-based anomaly detection. Comput. Netw., 5: 864-881. DOI: 10.1016/j.comnet.2008.11.011

Robertson, W., G. Vigna, C. Kruegel and R.A. Kemmerer, 2006. Using generalization and characterization techniques in the anomaly-based detection of web attacks.

Saboori, E., S. Parsazad and Y. Sanatkhani, 2010. Automatic firewall rules generator for anomaly detection systems with Apriori algorithm. Proceeding of the 3rd International Conference on Advanced Computer Theory and Engineering, Aug. 20-22, IEEE Xplore Press, Chengdu, pp: 57-60. DOI: 10.1109/ICACTE.2010.5579365

Torrano-gimenez, C., A. Perez-villegas and G. Alvarez, 2010. An anomaly-based approach for intrusion detection in web traffic. J. Inform. Assurance Security, 5: 446-454.

Vorobiev, A. and J. Han, 2006. Security attack ontology for web services. Proceedings of the 2nd International Conference on Semantics, Knowledge and Grid, Nov. 1-3, IEEE Xplore Press, Guilin, pp: 42-42. DOI: 10.1109/SKG.2006.85

Vasudevan, A.R., E. Harshini and S. Selvakumar, 2011. SSENet-2011: A Network Intrusion Detection System dataset and its comparison with KDD CUP 99 dataset. Proceedings of the Second Asian Himalayas International Conference on Internet, Nov. 4-6, IEEE Xplore Press, Kathmandu, pp: 1-5. DOI: 10.1109/AHICI.2011.6113948

Wang, K. and S.J. Stolfo, 2004. Anomalous Payload-based Network Intrusion Detection. In: Recent Advances in Intrusion Detection, Jonsson, E., A. Valdes and M. Almgren (Eds.)., Springer Berlin Heidelberg, ISBN-10: 978-3-540-23123-3, pp: 203-222.

Yang, X., M. Sun, X. Hu and J. Yang, 2009. Grammar-Based Anomaly Methods for HTTP Attacks. Proceedings of the Chinese Conference on Pattern Recognition, Nov. 4-6, IEEE Xplore Press, Nanjing, pp: 1-5. DOI: 10.1109/CCPR.2009.5344007