

Anomaly Detection based on Fuzzy Rules

Wenjiang Jiao^{a,*}, Qingbin Li^b

^a*School of Computer Science and Technology, Shandong University, Jinan, 250100, China*

^b*Shandong University, Jinan, 250012, China*

Abstract

Essentially, the fuzzy assert rule library is the fuzzy decision tree. A fuzzy decision tree growth algorithm based on local dynamic optimization is present. Following the idea of the greedy strategy, the algorithm ensures that once a continuous attribute is chosen as a branch node, the membership functions of this attribute after fuzzifying is dynamically optimized. On the other hand, according to fuzzy logic, enhanced Apriori algorithm is present to all the fuzzy frequent item sets composed of fuzzified attributes of multiple events. Then, the fuzzy frequent item sets are transformed into fuzzy association rules that compose the fuzzy association rule library. As for a multiple event sequence, eight different detection algorithms are provided and tested on the same platform. Experiments show that two new algorithms using the fuzzy decision tree and fuzzy association rule library detection models get the highest score.

Keywords: web application; network security; intrusion detection; anomaly detection

(Submitted on November 28, 2017; Revised on December 29, 2017; Accepted on January 19, 2018)

© 2018 Totem Publisher, Inc. All rights reserved.

1. Introduction

The data cleansing technology based on outlier detection define central bias function of compound events and dissimilarity function of event collection. Hence, the central deviation priority exception set growing algorithm is proposed for constructing the exception set, avoiding the NP difficult problem by controlling algorithm complexity within polynomial complexity [14,2,4]. It solves the cleaning problem of the training data. Secondly, according to the own feature of Web application invasion detection, the classification detection algorithm based on the fuzzy proposition rule and the clustering detection algorithm based on the fuzzy association rule are raised. They introduce fuzzy logic [6,8]; on the basis of discretization of compound event attributes, map such attributes to fuzzy set; define appropriate fuzzy set membership function; then the local dynamic optimal fuzzy decision tree growing algorithm is proposed. With the fuzzy decision tree, indirect classifying detection is achieved based on the fuzzy proposition rule; the mining algorithm based on fuzzy association rule is proposed, which realizes the abnormal detection based on the fuzzy association rule with Web application request as unit [15,1,3]. With genetic algorithm, parameters of fuzzy set membership function and the support thresholds of frequent fuzzy item sets are optimized. Finally, multiple detection methods based on fuzzy rule are implemented [12]. Experiments are conducted about abnormal detection on the compound event sequences of the intrusion into various Web applications.

Anomaly detection has five basic elements: object detection, data source, data model, detection model and detection algorithm. Detection object is a kind of intrusion detection system that needs to be monitored and protected, such as file systems, network services, user accounts, keyboards, etc.

Users can access these resources locally or remotely, and the user's access to resources is called user behavior. The core idea of anomaly detection is the modeling of user behavior, which is based on the behavior anomaly.

If a system resource is designated as a test object, the test data source includes all users of the resource access log. Log

* Corresponding author.

E-mail address: wenjiangjiao2017@126.com

files are often used to record the user's access to resources, as well as add new audit methods to track the more detailed information generated by these visits.

For example, the keyboard as the background keyboard monitors the keystrokes, or through the application process of a strace tracking system call, and so on. The keystroke sequences of system calls are recorded as well as the object detection keyboard and application data source. Commonly used data sources also include network packets, Web service log, login user Shell command sequence, etc.

In order to get the training data set employed to construct the detection model, it is necessary to abstract all kinds of heterogeneous data sources. The data model is relatively uniform. In the field of anomaly detection, the event is usually used as a data model. The event can be a TCP connection, an HTTP service access, a user login, a system call or a Shell command, etc.

Formally, events can be divided into an attribute vector. Simple events can be divided into a single dimensional vector. For example, the system call event only calls the system ID an attribute dimension, and the complex event requires multiple attribute dimensions. For example, to describe a TCP connection, you usually need the destination port number, duration, number of bytes sent, and so on.

Based on the data model, the detection model can be trained. The detection model describes the behavior characteristics of the user when gaining access to the object. The establishment of efficient detection model is at the core task of intrusion detection.

A set of data models typically used to train the detection model is called the training data set. The set of test performance is referred to as the test data set, and sometimes it is necessary to use the intrusion data source to invade the data set.

Detailed experimental plan, testing data and analysis is given [13,7,10]. In order to avoid confusion of concepts, the event or compound event mentioned in the paper are formally attribute vectors consisting of several attributes. Event collection is formed by all events in the compound event sequence when time sequential information is not taken into account; when the context stresses time sequential information, it is called event sequence or compound event sequence [5, 9].

2. The compound event

The compound event defined here is used to describe HTTP request sent by the Web application user. Abstract HTTP request is transformed to an attribute vector by: exploring the relationship between internal attributes of the vector and quantifying such relationship to fuzzy rule and use the fuzzy rule bank as a foundation to design proper algorithm to detect any possible abnormality, which is the basis for doing abnormal detection based on fuzzy rule [15].

For anomaly detection of Web application, defining a proper compound event requires certain knowledge in the field. Here HTTP request is reduced to four types: elementary attribute, statistical attribute, parameter attribute and Boolean attribute. According to that, compound event is defined [11]. For the convenience of studying internal association between event attributes, we divide attributes to value attributes, discrete attributes and compound discrete attributes as per the type of attribute value. Compound discrete attribute is composed of some relevant discrete attributes. The aim of defining compound discrete attribute is to reduce the number of attributes as much as possible and avoid from mining too much of the hardly explainable rules in the premise of guaranteeing the descriptive ability of the compound event. For instance, a dubious character can be contained and used as a compound discrete attribute, which can then be designed as a sub attribute of Boolean type like: whether it contains hexadecimal number or binary character string. It is shown in Table 1.

3. Anomaly detection based on rule

In the field of knowledge discovery, a variety of data classification methods have been proposed. Classification analysis of data generally includes three steps:

The first step is to set up the training data set, which is an attribute vector. Composite events defined in this paper are the data tuples. Training data sets of data tuples are called training samples.

In the second step, according to the training data set of the learning data model, the data model is the abstraction of the training sample, which contains a top that is found from the training data set of knowledge. These models include decision trees, Bias and Bias probability network;

Table 1. Composite event properties and types

Attribute class	Attribute description	Attribute value type
Basic properties	Request URL	discretization
	HTTP response code	discretization
	HTTP response byte count	value
	Request string length	value
	Request method (POST, GET, HEAD, TRACE...)	discretization
Statistical properties	Number of 30s requests in the past with the IP host	value
	Number of failed 30s requests with IP host	
	30s request failure rate with IP host	
	The number of bytes sent to the IP host in the past 30 seconds	
	The total number of bytes sent by the past 30s to the same IP host	
Parameter properties	Number of parameters	value
	Parameter1	discretization
	Parameter2	discretization
	...	discretization
Boolean property	Whether the string contains sixteen binary coded	Composite dispersion
	Request string contains binary code	

The third step is to construct a set of test data. The test sample is classified by the data model, and the classification accuracy is calculated.

In the field of intrusion detection, it is very difficult to establish a training dataset with class labels because:

- (1) It needs human labor to give class labels to tremendous training datasets; in the abnormal detection, class label is an additional “behavior” attribute; it needs someone to judge whether event behavior is abnormal based on knowledge in the field.
- (2) It’s hard to decide the percentage of the normal and abnormal events in the training dataset, which directly influences the judging accuracy of the data model. For Web application intrusion detection, since Web attack means are diversified, attribute sources of events will diversify as well. It’s rather difficult to build an accurate training dataset with labels. Instead, it requires a method to get the data model by means of unlabeled the training dataset and classifying whether an event is abnormal; such kind of unguided learning process is clustering in data mining.

To analyze the difficulty in Web application intrusion detection based on compound events, the paper introduces a rule-based analysis method. The method is based on rules and has the following features:

- (1) Rules describe object behavior from a higher level, which are easily explainable and understandable. The rule bank is the foundation to construct expert systems. Rules of high credibility can be directly regarded as ground for the misused detection to improve the precision of detection.
- (2) The rule bank itself is a good data model learning process indicator; in the training process, the growing curve of rule bank scale tends to be mild, and it’s judged on whether the training is sufficient and complete.
- (3) Rules have metric parameters, e.g. support degree and confidence of an association rule. Through rule measuring, noisy data and influence by abnormal data can be removed. Rule-based detection has benign robustness.

3. 1. Propositional rule

Proposition rule is the most easily understood rule with the best representation. In the field of machine learning, proposition rule is also named the if-then rule. Suppose compound event e is the attribute vector formed by n discrete attributes. $e = \langle a_1, a_2, \dots, a_i, \dots, a_n \rangle$. Choose one attribute as a class label and mark a_c ; then, the class label collection relative to a_c is C . Propositional rules as showed in (1):

$$f(a_1 = v_1) \wedge (a_i = v_i) \dots \wedge (a_k = v_k) \text{ then } (a_c = c_1) \quad (1)$$

To confirm whether an attribute is the best at classifying attribute on the current node, the basis is: if the attribute is employed to divide the sample, the attribute classifies the relative sample of the node on whether the required information by the labeling classification is the least. We use information growth to verify whether the candidate attribute is the current best classifying attribute. Suppose collection T is formed by t samples; samples are all attribute vectors; class label attribute has m discrete values, i.e. collection T can be portioned to m classes C_i according to the attribute of a class label; it is number of samples in the class C_i . After division of samples as per one classifying attribute, the amount of information required by labeling classification of sub collection with the same attribute value is derived by equation 2.

$$I(t_1, t_2, \dots, t_m) = - \sum_{i=1}^m \frac{t_i}{t} \ln \left(\frac{t_i}{t} \right) \quad (2)$$

Each path from root node to leaf node of the decision tree corresponds to a piece of proposition rule. Proposition rule collection can be derived through the simple decision tree traverse algorithm. In the decision tree in Figure 1, a sample attribute vector has four attributes: A, B, C, D, where C is the class labeling attribute. From that decision tree, the following five pieces of proposition rules can be derived:

$$\begin{aligned} &f(A = a_1) \wedge (B = b_1) \text{ then } (C = c_1) \\ &f(A = a_1) \wedge (B = b_2) \text{ then } (C = c_2) \\ &f(A = a_1) \text{ then } (C = c_1) \\ &f(A = a_3) \wedge (D = d_1) \text{ then } (C = c_2) \\ &f(A = a_3) \wedge (D = d_2) \text{ then } (C = c_1) \end{aligned}$$

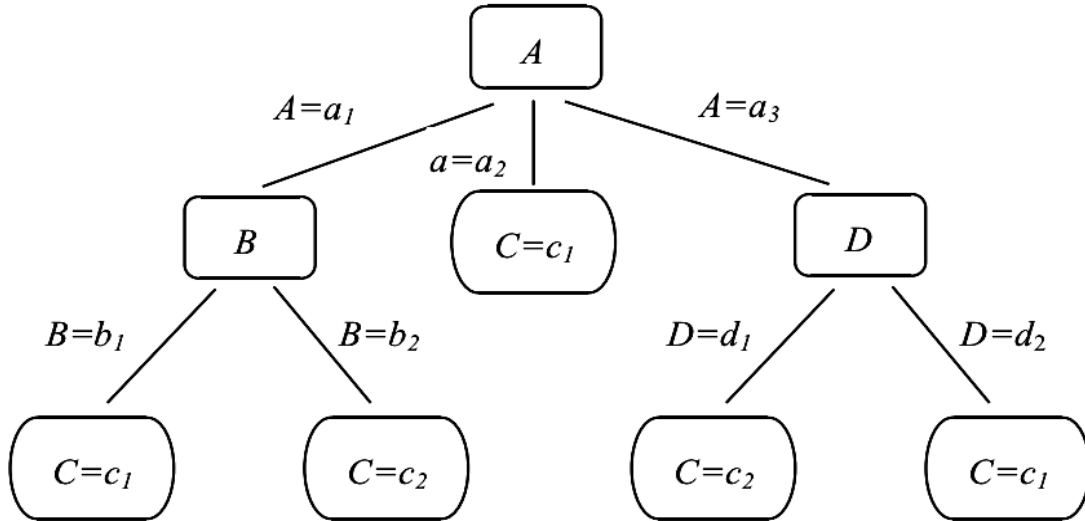


Figure 1. Example of a decision tree

According to the association rule, we use the way of Boolean clustering to go into the detection of intrusion. Boolean clustering needs only to divide pending data into normal and abnormal type, with rather high precision of partition. The computing method of similarity between association rule banks proposed by Wang et al. can be employed to implement Boolean clustering of an association rule. Set two pieces of association rule $R_1 : X_1 \rightarrow Y_1[s_1, c_1]$ and $R_2 : X_2 \rightarrow Y_2[s_2, c_2]$; the similarity function that defines an association rule is put in Equation 3:

$$Similarity(R_1, R_2) = \begin{cases} \max(0, 1 - \max(\frac{|c_1 - c_2|}{c_1}, \frac{|s_1 - s_2|}{s_1})) \\ 0 \end{cases} \quad (3)$$

Based on the above analysis, the rule-based detection method of invasion can easily judge whether an invasion exists by effectively combining field knowledge. In order for the rule-based approach to detect Web application anomaly more effectively, two questions should be dealt with: discretion reasonably numerical attribute and realize anomaly judgment based on request.

3.2. Fuzzy association rule clustering

Association rule is an important tool for the field of data mining. The internal relation of association rules is used to describe the normal behavior of the system. In this paper, the behavior of the system is abstracted as a compound event sequence, and the association rule library records the dependencies between the attributes in the attribute vector of the corresponding event. The use of association rules for anomaly detection needs to solve three problems: the avoidance of the boundary sharpening effect, the detection based on the demand and the optimization of the subset of attributes and parameters.

In the attribute-value pair of an association rule, the value of discrete attribute does not change, but the numerical attribute is mapped in the form of set to interval label, which represents the subinterval of sample attribute value. The value of a numerical attribute is subinterval label. Compared with the general association rule, the basic concept of fuzzy association rule doesn't change; what differs is the attribute-value of the event value attribute is no longer one-to-one mapping to interval label, but maps to the fuzzy set, which stands for the fuzzy concept. Fuzzy association rule accords fully with fuzzy proposition rule in terms of discretization and fuzzification of the numerical attribute.

Give event attribute vector instance e , which contains discrete attribute (including compound discrete attribute) and value attribute; after e is uniformly fuzzier, the fuzzy item set is derived like: $f_e = \{f_d_i \mid f_d_i \leq a_k, v_j(f_e) > a_k\}$.

Since the concept of fuzzy set and membership is introduced, the calculation of support and confidence of fuzzy association rules differ. Make sample event sequence D as training set; D contains t event attribute vector instances. Fuzzily all examples in D ; then get the collection F_E , $F_E = (f_e_1, f_e_2, \dots, f_e_t)$ of fuzzy item set.

Assume that D is a training set for the two-dimensional event vector e , which contains 2 sample instances. Examples of D are set by sharp set mapping and fuzzy set mapping. It is shown in Table 2.

Table 2. Sharp set mapping and fuzzy mapping

Instance			Sharp set mapping: term	Fuzzy set mapping: Fuzzy term	
1	a_1	15	$\langle a_1, \text{low} \rangle$	$\langle a_1, \text{"F Low"}, 0.7 \rangle$	
				$\langle a_1, \text{"F middle"}, 0.3 \rangle$	
				$\langle a_1, \text{"F High"}, 0 \rangle$	
	a_2	yes	$\langle a_2, \text{yes} \rangle$	$\langle a_2, \text{"Yes"}, 1 \rangle$	
2	a_1	50	$\langle a_1, \text{middle} \rangle$	$\langle a_2, \text{"No"}, 0 \rangle$	
				$\langle a_1, \text{"F Low"}, 0.2 \rangle$	
				$\langle a_1, \text{"F middle"}, 0.7 \rangle$	
	a_2	no	$\langle a_1, \text{"F High"}, 0.1 \rangle$	$\langle a_2, \text{"Yes"}, 0 \rangle$	
3	a_1	80	$\langle a_2, \text{no} \rangle$	$\langle a_2, \text{"No"}, 1 \rangle$	
				$\langle a_1, \text{high} \rangle$	$\langle a_1, \text{"F Low"}, 0 \rangle$
				$\langle a_1, \text{"F middle"}, 0.2 \rangle$	
	a_2	yes	$\langle a_1, \text{"F High"}, 0.8 \rangle$	$\langle a_2, \text{"Yes"}, 1 \rangle$	
			$\langle a_2, \text{yes} \rangle$	$\langle a_2, \text{"No"}, 0 \rangle$	

Algorithm 1 Frequent fuzzy item set mining algorithm

Input: use fuzzy set for training set and map it to get fuzzy item set sample collection F_E ; the lower threshold of fuzzy item membership is:

Support degree f_sup of μ_{min} fuzzy association rule; maximum item set number I_{max} of fuzzy item set

Output: frequent fuzzy item set collection F_L

(1) Construct a collection which contains all 1-fuzzy item sets, and write as F_L1 ; the average membership of all 1-fuzzy item sets in F_L1 to the sample set F_E is bigger than μ_{min} , which is called 1-frequent fuzzy item set collection;

For each a_k in A

For each v_j in V_{a_k} {

Select $f_i = \{f_d\}$ in F_L1 ;

}

(2) Construct return value collection F_L ; assign F_L1 to F_L , i.e. $F_L = F_L1$; set integer $P = 2$;

(3) When $p-1 = I_{max}$, or F_L_{p-1} is null, the algorithm stops, back to F_L ;

(4) Do joint operation; connect every two by two of all $(p-1)$ -frequent fuzzy item sets which satisfy conditions in F_L_{p-1} into p -fuzzy item set; enter into $F_L'_p$, which becomes candidate p -fuzzy item set collection;

(5) Check all p -fuzzy item sets in the p -fuzzy item set collection $F_L'_p$; if $(p-1)$ -fuzzy sub item set in p -fuzzy item set doesn't belong to F_L_{p-1} , remove p -fuzzy item set from $F_L'_p$;

(6) Check the support of sample collection F_E from all p -fuzzy item set in the p -fuzzy item set collection $F_L'_p$; if the support of p -fuzzy item set is below f , remove the p -fuzzy item set from $F_L'_p$;

(7) Till now, all p -fuzzy item sets in $F_L'_p$ are frequent; so $F_L'_p$ becomes formal p -frequent fuzzy item set collection;

(8) Add all p -frequent fuzzy item sets in $F_L'_p$ to F_L , p itself increases 1; back to step (3).

(9) End

3.3. Anomaly detection based on Clustering

Similar to the indirect anomaly detection based on proposition rule classification, the association rule can also be used for anomaly detection with event sequence segment as the unit to confirm whether anomaly happens in the segment. The segment detection method of association rule gets the standard association rule bank in accordance to training dataset. During the detection, segments pend event sequences according to the similar association rule mining algorithm and dig pending association rule bank. Later, by using Equation 2, calculate the similarity between the standard bank and the pending one; when the similarity is smaller than the threshold value, confirm that anomaly occurs in the segment.

However, the association rule segment detection method has drawbacks:

- (1) Unable to use Web APP request as unit to accurately confirm abnormality of each request. It is helpless to the invasion of attack request, which mingles in plentiful normal requests.
- (2) The high precision of anomaly detection and good robustness base on association rule are achieved on the foundation of well-established association rule bank; the association bank constructing algorithm involves big expenses, especially after introduction of fuzzy logic. The fuzzy association rule bank becomes bigger and encounters such problems as more complicated attributes subset and parameter optimization; so, during the detection, it's not appropriate to construct the child rule bank by using segment as unit; instead, single request should be regarded as detection unit, and needs an algorithm with small overhead. That algorithm is able to concentrate on the model training stage's complicated and money-consuming tasks like rule bank construction and optimization. After the rule bank is well established, it only needs the simple detection method, which can be the effective implementation of anomaly detection based on request. Fuzzy association rule base is shown in Formula 4:

$$F_RS = \{F_R_l \mid FR_l : X_l \rightarrow Y_l[f_confidence_l], 1 \geq l \geq |F_RS|\} \quad (4)$$

According to Equation 4, when pending fuzzy item sets meet both the former and latter terms of the fuzzy association rule, i.e. the membership degree of all fuzzy items in the former and latter terms of the rule are all bigger than zero, it's positive matching and matching function=1. When pending fuzzy item sets don't meet the former item of the rule, i.e. there are fuzzy items whose membership are zero in the former item, it's zero matching and matching function =0; if pending fuzzy item sets suffice the former item but not the latter, it's negative matching and matching function are -1.

According to matching function, the anomaly score of pending fuzzy item set f_e against fuzzy association rule bank F_RS can be reached by Formula 5:

$$AnomalyScore = \sum_{F_Re F_RS} (-Match(f_e, F_R_l) * f_confidence_l) \quad (5)$$

The higher the abnormal score, the greater the possibility of anomaly. Therefore, it is necessary to calculate the value of the matching function when calculating the abnormal score because the mismatch indicates the exception, and matching means that the abnormal possibility is reduced.

4. Experimental Analysis and Results

After discussing features of the proposed anomaly detection algorithm based on fuzzy association rule clustering, we utilized identical training datasets and testing datasets to compare several methods and analyze the detection efficiencies and performances of each.

For the experiment, client detection program, target Web server and Web app invasion platforms are applied to test different training datasets and testing datasets.

For the anomaly detection algorithm based on rule, the scale of rule bank is a natural indicator of the train model learning process. Figure 2 is the growth curve of the proportion of both the decision tree and fuzzy decision tree to training dataset size. Apparently, for the same training dataset, fuzzy decision tree has bigger information volume, containing almost 2000 decision leaf nodes, and tend to be steady when the scale of training dataset reaches 75000. One thing to note is that the scale of a fuzzy decision tree is not only related to the training set, but also dependent on some parameters of the algorithm. In Figure 2, the growth curve of a fuzzy decision tree corresponds to various parameter values: the lower threshold of node sample percentage $s_{min} = 0.1$. Evolution time of genetic algorithm $n=35$. Gene pool sizes $m=10$. Crossover probability $P_c = 0.8$. Mutation probability $P_m = 0.01$.

The key to the classification anomaly detection algorithm is the classifying accuracy of the decision tree. Generally speaking, classifying accuracy would decrease with increasingly more class labels. Figure 3 curves the relationship between decision tree classifying accuracy and the number of class labels. For fuzzy decision tree, the evolutionary time n of genetic algorithm of membership function parameter also affects the classification accuracy. The more time it evolves, the better the parameter of membership function will become and the higher the classifying efficiency of the fuzzy decision tree will reach.

The basis of decision tree classification is the normal value of each component of the event attribute vector. Suppose Web app has two URLs. If normal values of each component of the event attribute vector to which the two kinds of URL request are fully consistent, the decision tree can't correctly identify the two kinds of event and so the classification result is random. Even if the event attribute vector has enough components, it rarely happens. With increasing number of class labels, it is more likely and classification accuracy rate is declining. In order to ensure the effectiveness of the anomaly detection algorithm based on decision tree classification, the decision tree must have a sufficiently high classification accuracy rate. Figure 3 displays that fuzzy decision tree's accuracy rate reaches up to 70% when it classifies events that contain 12 class labels, while the general decision tree has only 50% accuracy.

Another factor affecting the anomaly detection algorithm based on decision tree classification is the distribution of malicious requests in Web app invasion. The more intensive the malicious request is, the more it can be detected. As a matter of fact, that is a common feature of all anomaly detection algorithms based on request segment. Figure 4 shows the comparison of decision tree's classification accuracy with those of different testing datasets. Testing datasets contain malicious requests of different percentages. All detection datasets only contain events with three class labels.

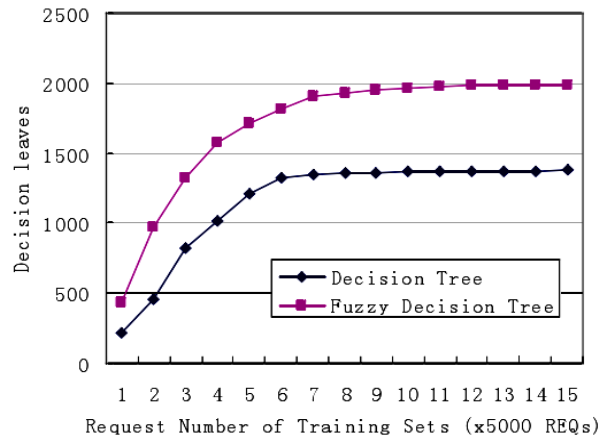


Figure 2. Decision tree sizes growth curve

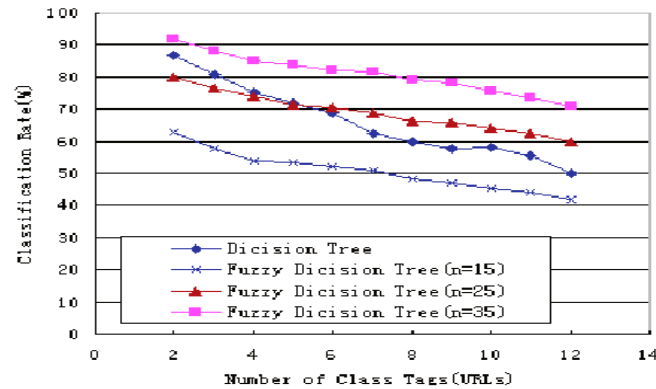


Figure 3. Decision tree classification accuracy

As to the anomaly detection algorithm based on fuzzy association rule clustering, the key is to have the appropriate fuzzy association rule bank through training. The algorithm should build a separate rule bank for each Web application. To control the scale of the rule bank, the algorithm defines the maximum item set number of fuzzy association rule fuzzy item sets to 4. Assume there are four fuzzy item sets A, B, C and D. Then fuzzy association rule has four types: unitary rule A, binary rule $A \rightarrow B$, ternary rule $AB \rightarrow C$, and quaternary rule $ABC \rightarrow D$. Figure 5 plots the cumulative curve of the scale of fuzzy rule bank of the three Web apps: group.php, register.php and login.php. We see that the more complicated the application is, the bigger the rule bank gets. Note in Figure 5, each dot on the curve relates to a certain size of training dataset, which is the number of rules in the fuzzy association rule bank after optimization by genetic algorithm. The rule bank of three applications gets steady when the scale of training set reaches 20000.

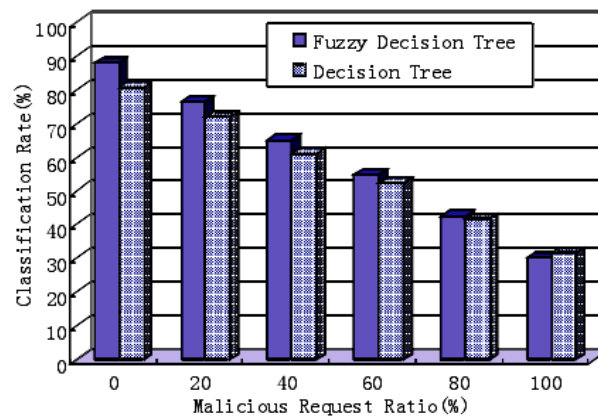


Figure 4. Classification accuracy of decision tree is under different detection data sets

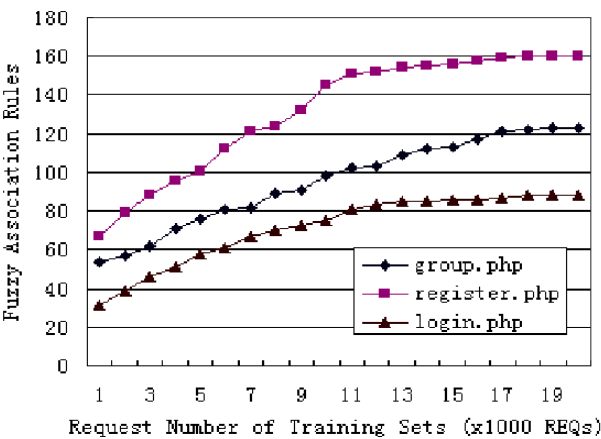


Figure 5. Three kinds of Web application fuzzy association rule base scale growth curve

The scale growing curve in Figure 5 demonstrates the convergence of the fuzzy association rule bank. More importantly, the size of training set required to get steady rule bank is determined. Still, we take for instance the three Web apps: group.php, register.php and login.php. For each of them we get a training set of enough size. Firstly, training is done to get the general association rule banks, which are G, R, and L; then, by the constructing algorithm for fuzzy association rule bank, we get the optimized fuzzy association rule banks, which are F_G, F_R and F_L. For each of the three Web apps, construct detection sets; use six kinds of rule banks (i.e. standard rule banks) acquired through training to do crossing calculation of the average anomaly score of request events in the detection sets. It is shown in Table 3. The anomaly score actually means the degree of difference between pending demand and standard rule bank. The smaller the anomaly score is, the lower the degree of discrepancy.

Table 3. The comparison of cross anomaly scores based on standard rule base

Web application \ Standard rule base	The abnormal scores of each standard rule base(average)			
	Association rule base		Fuzzy association rule base	
Group.php	G	1.73	F_G	-24.6
	R	11.7	F_R	1.67
	L	8.9	F_L	0.45
Register.php	G	14.6	F_G	1.56
	R	2.34	F_R	-25.1
	L	10.4	F_L	2.56
Login.php	G	8.9	F_G	1.56
	R	7.6	F_R	-1.67
	L	1.45	F_L	-24.1

5. Conclusions

In this paper, a fuzzy rule-based anomaly detection algorithm is introduced. Firstly, we propose an algorithm for outlier mining, which is based on the central deviation of outliers. Secondly, fuzzy logic is introduced into the mining of propositional rules and association rules, and a complete mining algorithm is proposed. Two kinds of anomaly detection models are established, which are fuzzy decision tree and fuzzy association rule base. Finally, genetic algorithm is used to optimize the membership function parameters and the subset selection of fuzzy sets.

Real time online detection is the developmental direction of IDS technology. In this paper, the anomaly detection algorithm based on event flow graph has a low cost and can meet the requirements of online detection. The next step is to further decrease the cost of other detection algorithms to achieve a more comprehensive online detection.

References

1. Songling Fu, “Distributed Online Social Network Data Storage and Optimization Research”, *National Defense Science and Technology University*, 2014.
2. Xiaoshi Fan, Ying Lei and Yanan Wang, Intuitionistic Fuzzy Inference Method in Traffic Anomaly Detection”, *Chinese Journal*

- of Electronics and Information Technology*, (4): 2218-2224, 2015.
3. Cunchen Li, "Research and Application of Mass Data Distributed Storage Technology", *Beijing University of Posts and Telecommunications*, 2013.
 4. Cikou Liu, Feng Wang and Mingchuan Yang, "Research on Distributed Storage Technology for Large Data", *Telecommunications Technology*, vol.5, pp.33-36, 2015.
 5. Nerbu Li, "Distributed Anomaly Detection of Data Mining and Multi Stage Intrusion Alert Correlation Based on Research", *Jilin University*, 2010.
 6. Shuai Liu, "Research on Multi-level Anomaly Behavior Analysis and Detection Technology for Network Data Stream", *The PLA Information Engineering University*, 2015.
 7. Yang Pan, "Research on the Application of Hadoop Technology in Distributed Data Storage", *Dalian Maritime University*, 2015.
 8. Yizhou Qian, "Distributed Real-time Database with High Performance Data Storage Cloud Retrieval Mechanism", *Zhejiang University*, 2012.
 9. Shanqi Tao, "Research and Implementation of Intrusion Detection System for Mining Association Rules Based on Snort", *Nanjing University of Aeronautics*, 2009.
 10. Yu Wang, "Data Redundancy and Maintenance Technology in Distributed Storage System", *South China University of Technology*, 2011.
 11. Dongsheng Xu, Xiaoyan Ai, "Anomaly Intrusion Detection Based on Genetic Optimization and Fuzzy Rule Mining", *Computer applications*, vol.6, pp. 2227-2229, 2009.
 12. Zhengmin Xia, "Study on Network Traffic Analysis and Anomaly Detection Based on Fractal", *Shanghai Jiao Tong University*, 2012.
 13. Zhuoluo Yang, "Research and Implementation of Distributed Storage Technology in Data Warehouse", *Kunming University of Science and Technology*, 2012.
 14. Zhenqian Yang and Yongdan Yang, "Development and Application of Distributed Storage Technology for Large Data. *Electronics and Software Engineering*", vol.2, pp.201-210, 2016.
 15. Ling Zhang, "Research on Intrusion Detection Model Based on Rough Set and Artificial Immune", *Beijing University of Posts and Telecommunications*, 2014.
 16. Yuping Zhou, "Research on the Key Technology of Intrusion Detection Based on Intelligent Soft Computing", *Donghua University*, 2010.

Wenjiang Jiao received his M.S degree from Shandong University. He is an engineer in School of Computer Science and Technology, Shandong University. His research interests include computer network and network security.

Qingbin Li is an engineer in China Mobile Group Shandong Co., Ltd. Ji'nan the branch company engaged in computer network maintenance, IT maintenance and software development, the main research areas include computer network, network security and web development.