View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Elsevier - Publisher Connector

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 55 (2015) 231 – 237

Procedia Computer Science

ELSEVIER

Information Technology and Quantitative Management (ITQM 2015)

A New Intrusion Detection Approach using PSO based Multiple Criteria Linear Programming

Seyed Mojtaba Hosseini Bamakan.[a,b], Behnam Amiri[c], Mahboubeh Mirzabagheri[b], Yong Shi[a,b,d*]

[a]Key Laboratory of Big Data Mining and Knowledge Management, University of Chinese Academy of Sciences, Beijing 100090, China;
[b]Research Center on Fictitious Economy & Data Science , Chinese Academy of Sciences, Beijing 100090, China
[c]Department of IT management, University of Tehran, Tehran, Iran;
[d]College of Information Science and Technology, University of Nebraska at Omaha, 68182, NE, USA;

Abstract

Intrusion detection system (IDS) is an inseparable part of each computer networks to monitor the events and attacks, which many researchers proposed variety of models to improve the performance of IDS. In this paper we present a new method based on multiple criteria linear programming and particle swarm optimization to enhance the accuracy of attacks detection. Multiple criteria linear programming is a classification method based on mathematical programming which has been showed a potential ability to solve real-life data mining problems. However, tuning its parameters is an essential steps in training phase. Particle swarm optimization (PSO) is a robust and simple to implement optimization technique has been used in order to improve the performance of MCLP classifier. KDD CUP 99 dataset used to evaluate the performance of proposed method. The result demonstrated the proposed model has comparable performance based on detection rate, false alarm rate and running time compare to two other benchmark classifiers.

© 2015 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license
(http://creativecommons.org/licenses/by-nc-nd/4.0/).
Peer-review under responsibility of the Organizing Committee of ITQM 2015

*Keywords:*Network Intrusion detection; Multiple criteria linear programming; Particle swarm optimization; Classification, Data Mining

1. Introduction

Obviously organizations for doing their daily business depend on Internet and computer networks, therefore protecting their business from potential attacks or abnormal activities aimed at compromising their networks needs to be carefully concerned [1-2]. Different systems such as firewall, user authentication, antivirus and data

[*] Corresponding author. Tel.: 010-8268-0697
E-mail address: yshi@ucas.ac.cn

1877-0509 © 2015 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license
(http://creativecommons.org/licenses/by-nc-nd/4.0/).
Peer-review under responsibility of the Organizing Committee of ITQM 2015

doi:10.1016/j.procs.2015.07.040

encryption designed to protect computers networks [2], however these traditional network intrusion systems have failed to completely protect networks because of increasing and sophisticated nature of new attacks [3].

Therefore, intrusion detection system (IDS) become an inseparable part of each computer networks to monitors and analyses network traffics to detects the threats, attacks and abnormal events before they damage organizations' valuable information assets [1-4]. An intrusion detection system consist of data collection, data clearing and pre-processing, detecting the intrusion, reporting and do a reasonable action which in these process detecting the attack is an essential part [4]. According to [5], IDS defined as a software that have an ability to automate the intrusion detection. High classification accuracy and a low false alarm rate are the two main characteristics of well-developed IDS [1].

In general, IDSs are categorised into two main group based on their detection approaches: anomaly and misuse (signature) detection [2-4]. In anomaly detection systems, it suppose that abnormal behaviour is uncommon and different from normal behaviour, so any deviations from normal activities considered as an intrusive behaviour and the model will be build based on the normal data [1-4]. On the other hand, misused detection system marked the audit data as an intrusion if it is matched with the predefined signature of attacks [4]. Both of these approaches have some pros and cons. Anomaly detection systems has better performance in detecting the previously unknown attacks, but also has high false alarm rates, since any deviation from normal activities classified as abnormal behaviour [1]. Misuse detection systems, however has a low false alarm rate but fail when facing previously unknown attacks [1-4].

Although many efforts has been done on building an effective detection systems, nevertheless it is also an open research area. Because intrusion detection problem has different aspect such as large scale of traffic data, unbalanced data sets, ambiguous boundaries between normal and intrusive behaviour and highly dynamic environment, so each researcher address some of this aspects [4]. Support Vector Machine (SVM) as a well-known classification techniques has been used to improve the accuracy of IDS. [6] used SVM to improve intrusion system in wireless local area network. To improve the performance of SVM, [7] used both SVM and Artificial Neural Networks (ANN). However, to enhance low-frequent attack's detection and detection stability of ANN-based IDS, [8] proposed an approach called FC-ANN, based on ANN and fuzzy clustering. Most recently, an IDS was introduced by integrating intelligent dynamic swarm based rough set (IDS-RS) for feature selection and simplified swarm optimization for classification [3]. As state in [4], application of computational intelligence (CI) in intrusion detection systems attracted the attention of many researches to itself. CI methods such as artificial immune systems, artificial neural networks, swarm intelligence and soft computing showed a better performance compare with the traditional methods in high computational speed, fault tolerance and handling noisy data sets. Since recently could computing has been addressed by researchers, [5] provide a comprehensive review on intrusion detection and prevention systems (IDPS) in cloud computing systems. As proposed in [9], integrating swarm intelligence, especially particle swarm optimization with the other machine learning classifier to decrease the training time and improve the efficiency of IDS is an opening research area.

Multiple Criteria Linear Programming (MCLP) is an optimization-based classification model proposed and extended by Shi et al [10]. MCLP seeks to minimizing the sum of the all overlapping with the separating hyper-plane and maximizing the sum of the distances from a point to the separating hyper-plane [11].This method showed a good performance in different applications such as Credit Card Risk Analysis, Firm Bankruptcy Prediction, VIP E-Mail Behaviour Analysis and so on [11-13]. But the performance of MCLP model will be affected, if the parameters do not select and tune properly [14]. Although many techniques have been proposed for parameters optimizing, grid algorithm and genetic algorithm. Particle swarm optimization shown a better performance in short-time searching, less computationally intensive and ability to find global optimum [14]. So, we apply PSO to tuning the parameters of MCLP model. The remainder of this paper is organized as follow. In section 2, we describe the multiple criteria linear programming method. Section 3 gives an overview on particle swarm optimization and the result of proposed model are shown in section 4. Finally, we conclude and provide future works in Section 5.

2. Multiple criteria linear programming

Multiple criteria linear programming method is an optimization based classification algorithm which has been proposed and developed by Shi since 1998. This method construct a linear programing model based on two main objectives and some constraints to solve a classification problem. Consider X as an n-dimensional attribute vector, $X = (x_1, x_2, \dots , x_n)$ that each sample, X, is assumed to belong to a predefined class that called the class label attribute. Hence, this classifier try to find the best hyper-plane which can accurately separate the training samples as a different classes and then decided whether a data belong to a specific class or not [12]. In this case, if the samples are linearly separable, an appropriate hyper-plane here $w.x = b$ which w defined as weights for a subset of variables $w = (w_1, w_2, \dots, w_n)^T$and b defined as threshold can be find to separate the two classes. Suppose G denoted as class label Good and B denoted as class label Bad. The separable hyper-planes can be fined as:

$$(w.x_i) \leq b,\ \forall\, x_i\ \in G \tag{1}$$

$$(w.x_i) \geq b,\ \forall\, x_i\ \in B \tag{2}$$

Two opposite objective can be consider to define a classification problem based upon a linear programing modelling. The first objective is to minimize the sum of the samples overlapping degree from the separating hyper-plane, which it also called minimizing the sum of the distance (MSD) and the second objective is to maximize the sum of the distances of each samples from the separating hyper-plane which it considered as maximizing the minimum distance (MMD).

In MCLP modeling, ξ_i defined as the overlapping with respect of the training sample x_i which should be minimized, while β_i defined as distance from the training sample x_i to the separating hyper-plane. Hence, the constraint can be formulated as following:

$$(w.x_i) \leq b + \xi_i - \beta_i,\ \forall\, x_i\ , y =\ Good \tag{3}$$

$$(w.x_i) \geq b - \xi_i + \beta_i,\ \forall\, x_i\ , y = Bad \tag{4}$$

These two variable can be rewritten as $Y(Xw - eb) \geq \beta_i - \xi_i$ where e defined as vector of ones. The first Multiple Criteria Linear Programming (MCLP) model can be described as follows [15]:

$$\min \textstyle\sum_{i=1}^{n} \xi_i \tag{5}$$

$$\max\ \textstyle\sum_{i=1}^{n} \beta_i$$

$$\text{s.t. } (x_i\ , w) = b + y_i(\xi_i - \beta_i), i = 1, \dots , n$$

$$\xi, \beta \geq 0$$

According to shi et al [11-16-17], “the best trade-off between -$\sum_{i=1}^{n} \xi_{i}{}_{i}$and $\Sigma_i\beta_i$ is identified for an “optimal” solution. To explain this, assume the “ideal value” of -$\Sigma_i\, \xi_i$ be $\xi* > 0$ and the “ideal value” of $\Sigma_i\beta_i$ be $\beta* > 0$. Then, if -$\Sigma_i\, \xi_i > \xi*$, the regret measure is defined as -$d_\xi{}^{+} = \sum_{i=1}^{n} \xi_i + \xi*$. Otherwise, it is defined as 0. If -$\sum_{i=1}^{n} \xi_{i}{}_{i} < \xi*$, the regret measure is defined as $d_\xi^{-} = \xi* + \Sigma_i\xi_{i;}$ otherwise, it is 0. Thus, the relationship of these measures are (i) $\xi* + \sum_{i=1}^{n} \xi_{i}{}_{i} = d_\xi^{-}$- d_ξ^{+}, (ii) | $\xi* + \sum_{i=1}^{n} \xi_i$|= d_ξ^{-}+ d_ξ^{+}, and (iii) d_ξ^{-}, $d_\xi^{+} \geq 0$. Similarly, we derive

β^* - $\Sigma_i\beta_i = d_\beta^- - d_\beta^+$, $|\beta^* - \Sigma_i\beta_i| = d_\beta^- + d_\beta^+$, and d_β^- , $d_\beta^+ \geq 0$”. An MCLP model for two-class separation is formulated as following[10]:

$$\text{Minimize } d_\xi^{\ -} + d_\xi^{\ +} + d_\beta^- + d_\beta^+ \tag{6}$$

$$s.t.\quad \xi\ ^* + \sum_{i=1}^{n} \xi_i = d_\xi^{\ -} - d_\xi^{\ +},$$

$$\beta^* - \sum_{i=1}^{n} \beta_i = d_\beta^- - d_\beta^+,$$

$$(x_i\ , w) = b + y_i(\xi_i - \beta_i), i = 1, \dots, n$$

$$\xi, \beta, d_\xi^-, d_\xi^+, d_\beta^-, d_\beta^+ \geq 0,$$

Here $\xi*$and β$*$ are two parameters that we optimized them by particle swarm optimization method, w and b are unrestricted. The geometric meaning of model (6) is shown as in Figure.1. For more information about the application of MCLP in classification please refer to [10-11].

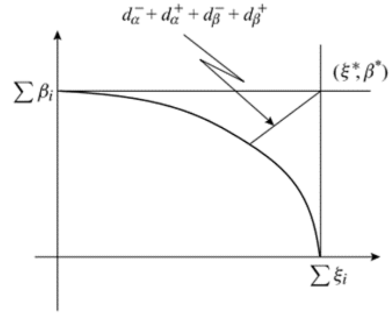

Fig.1. The geometric meaning of MCLP model optimization [11]

3. Particle swarm optimization (PSO)

By inspiration from animal’s collective behaviour such as bird flocking, fish schooling and ant colonies, artificial intelligence researchers proposed some optimization techniques called Swarm intelligence (SI) [4-18]. Swarm intelligence is a distributed solutions to complex problems which intend to solve complicated problems by interactions between simple agents and their environment [4-19]. Particle swarm optimization introduced by Eberhart and Kennedy in 1995 has several advantages compare with the other algorithms in this group such as simple to implement, scalability, robustness, quick in finding approximately optimal solution and flexibility [4-18-19].

In particle swarm optimization, each individual of population called particle. In standard PSO, after the initialization of the population, each particle update its velocity and its position in each iteration based on their own experience (pbest) and the best experience of all particles (gbest) as shown in Eql.(7 & 8). At the end of each iteration the performance of all particles will be evaluated by predefined cost functions.

$$v^i[t+1] = w.v^i\ [t] + c_1\, r_1(p^{i,best}[t] - p^i[t]) +\ c_2\, r_2(p^{g,best}[t] - p^i[t]) \tag{7}$$

$$p^i[t+1] = p^i[t] + v^i[t+1] \tag{8}$$

Where, $i = 1,2, \ldots, N$, N is the a number of swarm population. $v^i[t]$ is the velocity vector in $[t]th$ iteration. $p^i[t]$ represent the is the current position of the ith particle. $p^{i,best}[t]$ is the previous best position of ith particle and $p^{g,best}[t]$ is the previous best position of whole particle. To control the pressure of local and global search, w has been used. c_1 and c_2 are positive acceleration coefficients which respectively called cognitive parameter and social parameter. r_1 and r_2 are random number between 0 and 1.

4. Experiments and Results

For examining the performance of our model in intrusion detection, we used a standard dataset randomly selected from 10% of KDD Cup 99. The KDD99 dataset was derived in 1999 from the DARPA98 network traffic dataset with the purpose of evaluating the performance of intrusion detection systems [3-4]. Although some researchers have proposed that KDD99 dataset has some draw-backs [1], it is the most popular benchmark that has been used for testing network intrusion detection methodologies until now [1-20]. The training set contain 41 features for each TCP connection with a label which determine them as normal or attack [20]. The training set contains normal network traffic and 24 attacks that grouping into one of the four categories: Denial of Service (DoS), Probe, Users to Root (U2R), and Remote to Local (R2L). However, the testing dataset contains 38 attacks which mean 14 of them do not exist in the training set [4].

In this experiment, since the proposed MCLP model in this paper is a binary classification, we build a training data set that contain normal and DoS attack which compare with the other attacks has comparatively large size. Totally 5131 normal record and 6766 DoS attack has been chosen by random selection method. The following steps show the process of experiment and evaluation of proposed model:

Step 1: Preparation of training dataset, in which $X = (x_1, x_2, \ldots, x_n)$ represent an n-dimensional attribute vector, and $y_i \in \{-1, +1\}$ is the class label of which -1 and +1 respectively denoted to normal and attack record.

Step 2: Initialization the PSO parameters, w, c_1 , c_2, r_1 , r_2, n, $\xi*$ range and β∗ range, here, we defined $c_1 = 1$, $c_2 = 1.5$, , $n = 20$, $\mathrm{w}_{max} = 0.85$, $\mathrm{w}_{min} = 0.30$ and since according to [11] $\xi*$ should be negative number, -0.00001 and -0.1 respectively considered as lower and upper bound of $\xi*$. Furthermore, β∗ should be positive number, hence, 10 and 1000 respectively considered as lower and upper bound of β∗.

Step 3: Define the accuracy of model as a fitness function, here the ratio of correct predicted records to the entire records considered as an accuracy, in fact; accuracy = $\frac{(TP+TN)}{TP+FP+FN+TN}$

Step 4: Update the velocity and position of each particle in each iteration based on their own experience (pbest) and the best experience of all particles (gbest) according to equation (7) and (8).

Step 5: Terminate the PSO iteration by satisfying the stopping criteria, which here defined as getting the highest accuracy and acquire the optimal MCLP parameters ($\xi*$ and β∗)

Step 6: Apply the best tuned MCLP model on training dataset.

Two main popular metrics for IDSs' performance evaluation which are accuracy (ACC) and false alarm rate (FAR), has been calculated to compare the performance of PSO-MCLP model with MCLP, Support Vector Machine and C5.0. By considering Table 1 as Confusion matrix, accuracy defined as corrected prediction which means the percentage of True Negatives (TN) as well as True Positives (TP) upon the total records; $\mathrm{ACC} = \frac{(\mathrm{TP+TN})}{\mathrm{TP+FP+FN+TN}}$ and false alarm rate defined as normal records predicted as attacks; $\mathrm{FAR} = \frac{\mathrm{FP}}{\mathrm{FP+TN}}$. Table 2 shows the results of different classifier performance compare with PSO based multiple criteria linear programming model.

Table 1. Confusion matrix

		Predicted Class (Test Result)	
		P	N
Actual Class	P	True positive (TP)	False negative (FN)
	N	False positive (FP)	True negative (TN)

Table 2. result of PSO-MCLP's performance compare with MCLP, SVM and C5.0 in intrusion detection

Classifier	Traning Dataset		Testing Dataset	
	Accuracy	False Alarm Rate	Accuracy	False Alarm Rate
PSO-MCLP	0.9987	0.01724	0.9913	0.01947
MCLP	0.9823	0.03718	0.9746	0.03633
SVM	0.9964	0.02256	0.9897	0.02751
C5.0	0.9777	0.06329	0.9838	0.04268

Table 2 show the detection rate and false alarm rate for training and testing dataset. A model with a higher detection rate and lower false alarm rate has a better performance compare with the others. Here, PSO-MCLP gain 99.13 percent as an accuracy and 1.947 percent as a false alarm rate, hence stand in first position. SVM with accuracy of 98.97 percent and false alarm rate of 2.751 percent stand in the second position compare with the other methods. MCLP model gain 97.46 percent for accuracy which means that the parameters chosen by user were not well tuned.

5. Conclusion

In recent years, many research has been done to develop an effective data mining-based IDS. An effective IDS is defined when it can simultaneously obtain both high classification accuracy and low false alarm rates. In this paper we proposed a model based on multiple criteria linear programming which its parameter has been optimized by particle swarm optimization. The performance of the proposed model has been examined by KDD Cup 99. The experimental study indicated that the proposed PSO-MCLP model get better performance based on accuracy and running time compare with MCLP model which its parameters has been chosen by user or by cross validation.

As future work, in order to extend the applicability of our model, Multi-class classification based PSO-MCLP model will be apply on KDD Cup 99 to examine the performance of our model on detecting the different attacks simultaneously. Furthermore, performance of genetic algorithms for parameter tuning of MCLP model will compare with particle swarm optimization.

References

[1] Kou, G et al, Multiple criteria mathematical programming for multi-class classification and application in network intrusion detection. *Information Sciences* 2009; 179(4): p. 371-381.

[2] Tsai, C.-F., et al., Intrusion detection by machine learning: A review. *Expert Systems with Applications* 2009; 36(10): p. 11994-12000.

[3] Chung, YY. and Wahid N, A hybrid network intrusion detection system using simplified swarm optimization (SSO). *Applied Soft Computing*, 2012; 12(9): p. 3014-3022.

[4] Wu, SX. and Banzhaf W, The use of computational intelligence in intrusion detection systems: A review. *Applied Soft Computing*, 2010; 10(1): p. 1-35.

[5] Patel, A, et al, An intrusion detection and prevention system in cloud computing: A systematic review. *Journal of Network and Computer Applications*, 2013; 36(1): p. 25-41.

[6] Mohammed, MN. and Sulaiman N, Intrusion detection system based on SVM for WLAN. *Procedia Technology*, 2012;1: p. 313-317.

[7] Chen, WH, Hsu H.S, and Shen H.P, Application of SVM and ANN for intrusion detection. *Computers & Operations Research*, 2005; 32(10): p. 2617-2634.

[8] Wang, G, et al, A new approach to intrusion detection using Artificial Neural Networks and fuzzy clustering. *Expert Systems with Applications*, 2010; 37(9): p. 6225-6232.

[9] Kolias, C, Kambourakis G, and Maragoudakis M, Swarm intelligence in intrusion detection: A survey. *computers & security*, 2011; 30(8): p. 625-642.

[10] Shi, Y, Multiple criteria optimization-based data mining methods and applications: a systematic survey. *Knowledge and information systems*, 2010; 24(3): p. 369-391.

[11] Shi, Y, et al., Optimization Based Data Mining: Theory and Applications: Theory and Applications. 2011; Springer.

[12] Peng, Y, et al., Cross-validation and ensemble analyses on multiple-criteria linear programming classification for credit cardholder behavior, *in Computational Science-ICCS 2004*. 2004; Springer. p. 931-939.

[13] He, J, et al., Classifications of credit cardholder behavior by using fuzzy linear programming. *International Journal of Information Technology & Decision Making*, 2004; **3**(04): p. 633-650.

[14] Wang, X., et al., Real estate price forecasting based on SVM optimized by PSO. *Optik-International Journal for Light and Electron Optics*, 2014; 125(3): p. 1439-1443.

[15] Li, A, Y. Shi, and He J., MCLP-based methods for improving “Bad” catching rate in credit cardholder behavior analysis*.* *Applied Soft Computing*, 2008; 8(3): p. 1259-1265.

[16] Shi, Y, et al., Data mining via multiple criteria linear programming: applications in credit card portfolio management. *International Journal of Information Technology & Decision Making*, 2002; 1(01): p. 131-151.

[17] Shi, Y. and Peng Y, Multiple criteria and multiple constraint levels linear programming: concepts, techniques and applications. 2001; World Scientific New Jersey, USA.

[18] J. Kennedy, R.E., Particle swarm optimization*. in: Proceedings of the 1995 IEEE International Conference on Neural Networks*, 1995; Part 4 (of 6) Perth: p. pp. 1942–1948.

[19] Olariu, S. and Zomaya AY, Handbook of bioinspired algorithms and applications. 2005; CRC Press.

[20] Mahoney, M.V. and P.K. Chan. An analysis of the 1999 DARPA/Lincoln Laboratory evaluation data for network anomaly detection. in Recent Advances in Intrusion Detection. 2003; Springer.