OPEN ACCESS

International
Journal
Of Modern Engineering Research (IJMER)

The Combining Approach of Svms with Ant Colony Networks: An Application of Network Intrusion Detection

K.Karthika[a], R.Priya[b] , S.Suresh Kumar[c]
[a]*Assistant Professor, Department of Computer Applications, Saradha Gangadharan College, Puducherry, India.*
[b]*Assistant Professor, Department of Computer Applications, Saradha Gangadharan College, Puducherry, India.*
[c]*Assistant Professor, Department of Computer Applications, Saradha Gangadharan College, Puducherry, India.*

ABSTRACT:- Network security is one of the major concerns of the modern era. With the rapid development and massive usage of internet over the past decade, the vulnerabilities of network security have become an important issue. We introduce a new machine-learning-based data classification algorithm that is applied to network intrusion detection. The basic task is to classify network activities (in the network log as connection records) as normal or abnormal while minimizing misclassification. Although different classification models have been developed for network intrusion detection, each of them has its strengths and weaknesses, including the most commonly applied Support Vector Machine (SVM) method and the Clustering based on Self-Organized Ant Colony Network (CSOACN). Our new approach combines the SVM method with CSOACNs to take the advantages of both while avoiding their weaknesses. Our algorithm is implemented and evaluated using a standard benchmark KDD99 data set. Experiments show that CSVAC (Combining Support Vectors with Ant Colony) outperforms SVM alone or CSOACN alone in terms of both classification rate and run-time efficiency.

Keywords:- Data mining, Data classification, Intrusion detection system (IDS), Machine learning, Support vector machine, Ant colony optimization.

I. INTRODUCTION

In today's information system management, large-scale data clustering and classification have become increasingly important and a challenging area. Although various tools and methods have been proposed, few are sufficient and efficient enough for real applications due to the exponential growing-in-size and high dimensional data inputs. As a particular application area, Intrusion Detection Systems (IDSs) are designed to defend computer systems from various cyber attacks and computer viruses. IDSs build effective classification models or patterns to distinguish normal behaviors from abnormal behaviors that are represented by network data. There are two primary assumptions in the research of intrusion detection: (1) user and program activities are observable by computer systems (e.g. via system auditing mechanisms), and (2) normal and intrusion activities must have distinct behaviors .

1.1. Data-mining-based approaches for IDSs

Researchers have proposed and implemented various models that define different measures of system behavior. IDSs have been developed based on these models. Many of the existing IDSs, however, cannot adequately deal with new types of attack or changing computing environments, and hence the installed IDSs always need to be updated. As it is an energy and time consuming job for security experts to update current IDSs frequently by manual encoding, using data mining approaches to network intrusion detection provides an opportunity for IDSs to learn the behaviors of networks automatically by analyzing the data trails of their activities. Data mining has been widely used in many application areas. Two key advantages of using a data mining approach to IDSs are the following. (1) It can be used to automatically generate the detection models for IDSs, so that new attacks can be detected automatically as well. (2) It is general, so it can be used to build IDSs for a wide variety of computing environments. The central theme of data mining approaches is to take a data centric point of view and consider intrusion detection as a data analysis process [2]. This includes four essential steps.(1) Capturing packets transferred on the network.(2) Extracting an extensive set of features that can describe a network connection or a host session.(3) Learning a model that can accurately describe the behavior of abnormal and normal activities by applying data mining techniques.(4) Detecting the intrusions by using the learned models. In our research, we assume that the steps (1) and (2) have been developed and are already

available for the further training and testing phases. Approaches relevant to step (3) in data mining, in general, are by classification [3], link analysis, and sequence analysis [1]. In the rest of the paper, we will use SVM to denote either the concept or the algorithm when there is no confusion.

1.2. Motivation and contribution

Support Vector Machines (SVMs) have been widely accepted as a powerful data classification method (Section 4). On the other hand, the Self-Organized Ant Colony Network (CSOACN) has been shown to be efficient in data clustering (Section 5). Our work aims to develop an algorithm that combines the logic of both methods to produce a high -performance IDS. One challenge of developing IDSs is to realize real-time detection in high-speed networks. There are two important issues for this problem. First, in order to reduce the cost of deploying a model, we must be able to minimize the amount of clean data that is used by the data mining process. The machine-learning-based SVM method [4,5] is a good choice for learning with little volume of data [6–8]. Second, when new information is added into a system, updating of the old model is required immediately to ensure that the system is properly protected. As retraining may take weeks, or even months, it is impractical to retrain the new model on all available data. Thus, a mechanism is needed to generate an adaptive model that can be updated by cooperation of the old model with the new information. We take advantage of the clustering based Ant Colony Networks [9] in updating the models. Clustering in intrusion detection is used to resolve the multiple classification problems. However, the general method always involves expensive computation, especially if the set of training data is large. In order to save extensive computations, we modify the traditional CSOACN to control the clustering processes by clustering around certain objects (Section 6.2). This significantly reduced the retraining process and therefore the training time. Considering both of the issues for real-time detection problems, an active learning SVM and the modified CSOACN are chosen as the two basic components for our new classification algorithm. The main contributions of this paper include the following.

(1) Modifications to the supervised learning SVM and the unsupervised learning CSOACN so they can be used together interactively and efficiently.

(2) A new algorithm, CSVAC, that combines the modified SVM and CSOACN to minimize the training data set while allowing new data points to be added to the training set dynamically. The idea of combining supervised learning and unsupervised learning was applied previously [10].

However, the combined two algorithms of [10] are closely related by the neural dissimilarity. In our model, the supervised learning and the unsupervised learning algorithms have no relation, and they follow totally different logic. Their combination has the advantages of applying the logic of both sides and providing a more reliable solution to today's data intensive computing processes.

II. RELATED WORK

Issues related to intrusion detection can be categorized into two broad areas: (1) network security and intrusion detection models, and (2) intrusion detection methods and algorithms based on artificial intelligence (mostly machine learning) techniques. In this section we shall briefly review some related work in the second area, and leave area (1) to the next section, when we discuss the background of IDSs. Intrusion detection as a classification problem has been studied for decades using machine learning techniques, including traditional classification methods (single classifier) such as K-Nearest Neighbor (K-NN), Support Vector Machines (SVMs), Decision Trees (DTs), Bayesian, Self-Organized Maps (SOMs), Artificial Neural Networks(ANNs), Generic Algorithms (GAs), and Fuzzy Logic, as well as hybrid classifiers that combine multiple machine learning techniques to improve the performance of the classifier. A review of using these approaches was given in [11], which also included statistics of the use of these techniques reported in 55 research articles during the period 2000–2007. The review indicates that SVM and K-NN were the most commonly used techniques while the use of a hybrid increased significantly after 2004 and became main stream. Another more recent review [12] provided a thorough survey of intrusion detection using computational intelligence. It presented the details of the classification algorithms and swarm intelligence methods to solve problems using the decentralized agents. Most recently, an IDS was introduced by integrating On Line Analytical Processing (OLAP) tools and data mining techniques [13]. It is shown that the association of the two fields produces a good solution to deal with defects of IDSs such as low detection accuracy and high false alarm rate. As stated in [12], as one of the swarm intelligence approaches, Ant Colony Optimization (ACO), has been applied in many fields to solve optimization problems, but its application to the intrusion detection domain is limited. Several methods were reported using ACO for intrusion detection. For example, an ant classifier was proposed in [14] that used more than one colony of ants to find solutions in multiclass classification problem. Another ant-based clustering algorithm applied to detect intrusions in a network presented in [15] showed that the performance was comparable to some traditional classification methods like SVM, DT, and GA. In [16,17], the authors evaluated the basic ant-based clustering algorithms and proposed several improvement strategies to overcome the limitations of these clustering algorithms that would not perform well on clustering large and high-dimensional network data also

used ACO for intrusion detection in a distributed network. The basic ingredient of their ACO algorithm was a heuristic for probabilistically constructing solutions. All these ACO-based intrusion detection approaches are single classifiers as categorized by [11]. Hybrid intrusion detection approaches involving SVM have been studied in the past, such as the one reported in [4] that uses the Dynamically Growing Self-Organizing Tree (DGSOT) algorithm for clustering to help in finding the most qualified points to train the SVM classifier. It starts with an initial training set and expands the set gradually so that the training time for the SVM classifier is significantly reduced. Another hybrid intrusion detection approach was recently reported that combines hierarchical clustering and SVM. The purpose of using the hierarchical clustering algorithm is to provide the SVM classifier with fewer but higher quality training data that may reduce the training time and improve the performance of the classifier. The method presented in this paper is a hybrid classifier that combines SVM and ACO to improve the performance of SVM (particularly, reduce the SVM training time). This is a similar goal as [4] but we use ACO to achieve the goal that is capable of updating the models without a retraining process, as explained in the previous section about motivations. This project started from a preliminary idea of applying an unsupervised clustering algorithm to the supervising SVM learning. As a high-level framework, it was discussed in a short report in Later, during the development of the IDS, we presented the design and implementation of the system by UML diagrams. The work presented in this paper, however, includes the completed algorithms, detailed discussions of the principles, techniques on parameter adjustments, training and testing data sets, experiment design and analysis, and comprehensive performance evaluations among the new system CSVAC, SVM, ACO, and the KDD99 winner[2].

III. LITERATURE VIEW

3.1. A new intrusion detection system using support vector machines and hierarchical clustering

Whenever an intrusion occurs, the security and value of a computer system is compromised. Network-based attacks make it difficult for legitimate users to access various network services by purposely occupying or sabotaging network resources and services. This can be done by sending large amounts of network traffic, exploiting well-known faults in networking services, and by overloading network hosts. Intrusion Detection attempts to detect computer attacks by examining various data records observed in processes on the network and it is split into two groups, anomaly detection systems and misuse detection systems. Anomaly detection is an attempt to search for malicious behavior that deviates from established normal patterns. Misuse detection is used to identify intrusions that match known attack scenarios. Our interest here is in anomaly detection and our proposed method is a scalable solution for detecting network based anomalies. We use Support Vector Machines (SVM) for classification. The SVM is one of the most successful classification algorithms in the data mining area, but its long training time limits its use. This paper presents a study for enhancing the training time of SVM, specifically when dealing with large data sets, using hierarchical clustering analysis. We use the Dynamically Growing Self-Organizing Tree (DGSOT) algorithm for clustering because it has proved to overcome the drawbacks of traditional hierarchical clustering algorithms (e.g., hierarchical agglomerative clustering). Clustering analysis helps find the boundary points, which are the most qualified data points to train SVM, between two classes. We present a new approach of combination of SVM and DGSOT, which starts with an initial training set and expands it gradually using the clustering structure produced by the DGSOT algorithm. We compare our approach with the Rocchio Bundling technique and random selection in terms of accuracy loss and training time gain using a single benchmark real data set. We show that our proposed variations contribute significantly in improving the training process of SVM with high generalization accuracy and outperform the Rocchio Bundling technique

3.2. A Study of Network Intrusion Detection by Applying Clustering Techniques

In information system, security has remained one hard line area for computers as well as networks. In information protection, Intrusion Detection System (IDS) is used to safeguard the data confidentiality, integrity and system availability from various types of attacks. Data mining is an efficient artifice applied to intrusion detection to ascertain a new outline from the massive network data as well as it used to reduce the strain of the manual compilations of the normal and abnormal behavior patterns. This piece of writing reviews the present state of data mining clustering techniques to implement an intrusion detection system such as, Partitioning methods, Hierarchical methods, Model based clustering methods and their various types As a significant application area of data mining is intrusion detection based on data mining algorithms, aims to solve the troubles of analyzing enormous volumes of data IDSs build efficient clustering and classification models to distinguish normal behaviors from abnormal behaviors using data mining techniques. This study makes foundation in this field of research and exploration and implements intrusion detection model system based on data mining technology. In this paper, many data mining techniques have been proposed to improve the detection rate of Intrusion Detection System. In future, we planned to combine more than one clustering technique because different clustering algorithm have different knowledge to solve the problem so combining more than one data

clustering algorithm is used to remove the demerits of one another and a number of trained classifier lead to a superior performance than any single classifier. These techniques provide better performance in Intrusion Detection accuracy rate, faster running time and detecting the false positive rate. To fragment a complex problem into sub problems for which the solutions obtained are simpler to realize, execute, supervise and update.

3.3. Network Intrusion Detection by Support Vectors and Ant Colony

This paper presents a framework for a new approach in intrusion detection by combining two existing machine learning methods .The IDS based on the new algorithm can be applied as pure SVM, pure CSOACN or their combination by constructing the detection classifier under three different training modes respectively. The initial experiments indicate that performance of their combination is better than pure SVM in terms of higher average detection rate as well as lower rates of both negative and positive false and is better than pure CSOACN in terms of less training time with comparable detection rate and false alarm rates. In this paper, a framework for a new algorithm of intrusion detection is proposed by combining two existing machine learning methods .Based on the new algorithm, an IDS can be developed. The system can be used in different cases by construct the detection classifier under three different training modes SVM training mode is suitable for the time intense case that only one binary classifier is required by training upon a small amount of labeled data. Namely it only needs to distinguish the data of normal from abnormal. On the other hand, the CSOACN training mode is suitable for the preciseness intensive case and can solve multiclass problems upon both label and unlabeled data. The CSVAC mode, which is based on the combination of SVM and CSOACN, can be used to balance the performance of IDS in terms of efficiency and accuracy.

3.4. ACO based Distributed Intrusion Detection System

Intrusion detection is a problem of great significance to protecting information systems security. An intruder may move between multiple nodes in the network to conceal the origin of attack. Distributed intrusion detection and prevention plays an increasingly important role in securing computer networks. To overcome the limitations of conventional intrusion detection systems, alerts are made in distributed intrusion detection system which are exchanged and correlated in a cooperative fashion. It is necessary to develop fast machine learning based intrusion detection algorithms with high detection rates and low false alarm rates, due to the variety of network behaviors and the rapid development of attack fashions. The system has to observe to trigger thousands of alerts per day, in which most are mistakenly triggered by the false identification. So it is difficult for the analyst to correctly identify alerts related to the attack. This paper presents an intelligent learning approach using Ant Colony Optimization (ACO) based distributed intrusion detection system to detect intrusions in the distributed network. The experimental results on the proposed system with the feature extraction algorithm is effective to detect the unseen intrusion attacks with high detection rate and recognize normal network traffic with low false alarm rate. packets flowing through a network are analyzed. The NIDS can detect malicious packets that are designed to overlook a firewall's simplistic filtering rules. NDIS are placed at a strategic point or points within the network to monitor traffic to and from all devices on the network. In a host-based system, (HIDS) the IDS examines all the activity on each individual computer or host. HIDS consists of an agent on a host which identifies intrusions by analyzing system calls, application logs, file-system modifications and other host activities.

3.5. ANTIDS: Self Organized Ant-Based Clustering Model for Intrusion Detection System

Security of computers and the networks that connect them is increasingly becoming of great significance. Computer security is defined as the protection of computing systems against threats to confidentiality, integrity, and availability. Due to the fact that it is almost difficult for a system administrator to recognize and manually intervene to stop an attack, there is an increasing recognition that Intrusion Detection Systems (IDS) should have a lot to earn on following its basic principles on the behavior of complex natural systems, namely in what refers to self-organization, allowing for a real distributed and collective perception of this phenomena. Having that aim in mind, the present work presents a self-organized ANT colony based Intrusion Detection System (ANTIDS) to detect intrusions in a network infrastructure. The performance is compared among conventional soft computing paradigms like Decision Trees (DT), Support Vector Machines (SVM) and Linear Genetic Programming (LGP) to model fast, online and efficient intrusion detection systems. As depicted in Tables 1 and 2, ANTIDS approach has several limitations in what refers to the final recognition rate, obtaining optimal results only for some cases. This is in part due to several reasons. First, the large number of samples used in the present study forces *ACLUSTER* algorithm to be run on a large toroidal space. Empirical studies show that the optimal classification "habitat" area should be in the order of 4 times the number of objects, while the number of ants should be in the order of 1/10 of the number of objects. Rather, it's by large preferably to process and treat parts and streams of data independently, set after set Second, and still in the

present case, the data is poorly uniformly distributed between all the five classes.. In other words, the colony can collectively respond to the perturbation with individuals exhibiting the same behavior. When it comes to artificial agents, this type of flexibility is priceless: it means that the agents can respond to a perturbation without being reprogrammed to deal with that particular instability. Algorithm can work either in unsupervised or supervised mode and finally The self organizing nature of ANTIDS makes them an ideal candidate for distributed IDS.

IV. BACKGROUND

In this section, we present some background knowledge about IDSs. We begin with the introduction of basic concepts and technologies of network security.

4.1. Network security

Being used to transfer valuable and confidential information for a variety of purposes [23], communication networks attract the attention of people who intend to steal information or to disrupt systems. As business practices today are reshaped into e-commerce and data communication, the needs for network security have intensified. There are three basic security concerns that are important to information on a network [24].

• Confidentiality: Loss of confidentiality results when information is read or copied by unauthorized users.
• Integrity: Loss of integrity results when information is modified in unexpected ways.
• Availability: Loss of availability results when information can be erased or become inaccessible by authorized users. On the other hand, there are three security concepts that are related to the people who use the information.
• Authentication: Proving whether a user is who he or she claims to be.
• Authorization: Determining whether a particular user has the privilege to perform certain actions.
• Non-repudiation: Providing protection against an individual falsely denying having performed a particular action. Considering the vulnerabilities and incident trends of networks,

a robust defense requires an adaptable strategy as well as robust tools for the changing environment, well-defined policies and procedures, and continued vigilance. Many organizations have developed a variety of technologies to secure their systems and information against intruders. These technologies protect systems and information, detect unusual or suspicious activities, and respond to events that affect security. Some commonly applied technologies include encryption, timestamps and sequence numbers, and firewalls.

4.2. Network intrusion detection

Intrusion detection [25,26] is the detection of actions that attempt to compromise the integrity, confidentiality, or availability of a resource. It attempts to detect attacks by examining various data records observed through processes on the same network. The data records are split into two categories: host-based data [27], which is audit data that record all system calls in chronologically sorted order; and network-based data [28], which is the network traffic data (i.e., tcp dump). As one of the most promising network security technologies, Intrusion Detection Systems (IDSs) detect a possible intrusion as soon as possible and take appropriate actions. An IDS is a reactive rather than pro-active agent. Although many different IDSs have been developed (for instance, see [4]), their detection schemes generally fall into one of two categories: anomaly detection or misuse detection. Anomaly detectors look for behavior that deviates from normal use of the system whereas misuse detectors look for behavior that matches a known attack scenario. So far, a great deal of time and effort has been invested in IDS. Various approaches to model normal or attack behaviors have been used, for example, statistical approaches, predictive pattern generation neural networks, expert systems, keystroke monitoring, model-based intrusion detection, NSM (Network Security Monitor) [6], autonomous agents [7], fuzzy logic network [8] and data mining [9–1,3].However, it seems that neither anomaly detection nor misuse detection can detect all kinds of intrusion attempts on their own. The future research trend is converging towards a model which is a hybrid of the anomaly and misuse detection models.

V. ACTIVE LEARNING SUPPORT VECTOR MACHINES

To introduce the new algorithm (presented in Section 6), it is necessary to give a brief review of SVMs and the framework of active learning SVMs for intrusion detection.

5.1. Linear SVM trained on separable data

In an SVM, a data point is viewed as a vector in the d dimensional feature space. Assume that all data points belong to either class A or class B. Each training data point xi can be labeled by x_i based on (1):

$Y_{i=}\{-1\ x_i \in$ class A,

$1\ x_i \in$ class B.

Thus, the training data set can be denoted as $D=\{(x_i, y_i)|i=1,2,3\ldots,N\}$.Data points with label 1 and −1 are referred to as positive and negative points, respectively. In the linear separable case, there are many hyperplanes

which might separate the positive from the negative points. The algorithm simply looks for the largest margin separating hyper plane, where the ‘‘margin’’ of a separating hyperplane is defined to be the sum of the distances from the hyperplane to the closest positive and negative points. In order to compute the margin of a separating hyperplane H, consider the hyper planes H_1 and H_2 that contain the closest positive training points and the closest negative training points to H, respectively:

H: $w.x-b=0, x \in R^d$,

H_1: $w.x-b=1, x \in R^d$,

H_2: $w.x-b=-1, x \in R^d$,

where w is the normal to H and b is the distance from H to the origin. Obviously, H, H1, and H2 are parallel. In addition,

$w.x_i-b \geq 1$ for $y_i=1$

$w.x_i-b \leq 1$ for $y_i=-1$

The margin of the separating hyperplane H is equal to the distancebbetweenH1 andH2, which is $2/\|w\|$. Finding the maximum margin can be accomplished by minimizing $\|w\|$.It is clear that the separating hyperplane H will not change if the training points neither on H1 nor on H2 are removed from the training data set. Thus the points that lie on H1 or H2 are critical elements in the training data set, called support vectors.

5.2. **Non-linearly separable cases**

In the case of the data points not being linearly separable, a slack variable can be added to allow for some errors to be made, and the aims are to minimize the number and effect of the errors.

5.3. Intrusion detection based on an active learning SVM

A network connecting record can be described by several features of the connection and is represented as a data point for SVM training. The traditional SVM algorithm is operated over the entire training data set. The number of training data points determines the dimension of the matrix for computing the kernel functions, which influences the time of solving the QP problems. However, SVMs have the property that the points that do not lie on the margin do not need to be involved in the computation. The same decision function is obtained if some of the training data points, excepting the support vectors, are removed. Hence, the number of training data points can be reduced without losing accuracy. An active learning SVM was introduced in [6] to reduce the amount of training data. Initially, an SVM classifier was trained by using only small amount of data points from the whole training data set. The SVM classifier was then gradually modified by adding new data points for the training. After each training process, the output classifier is used to separate the entire data. The recurrence of training a new classifier can stop when a required correct classification rate is obtained.

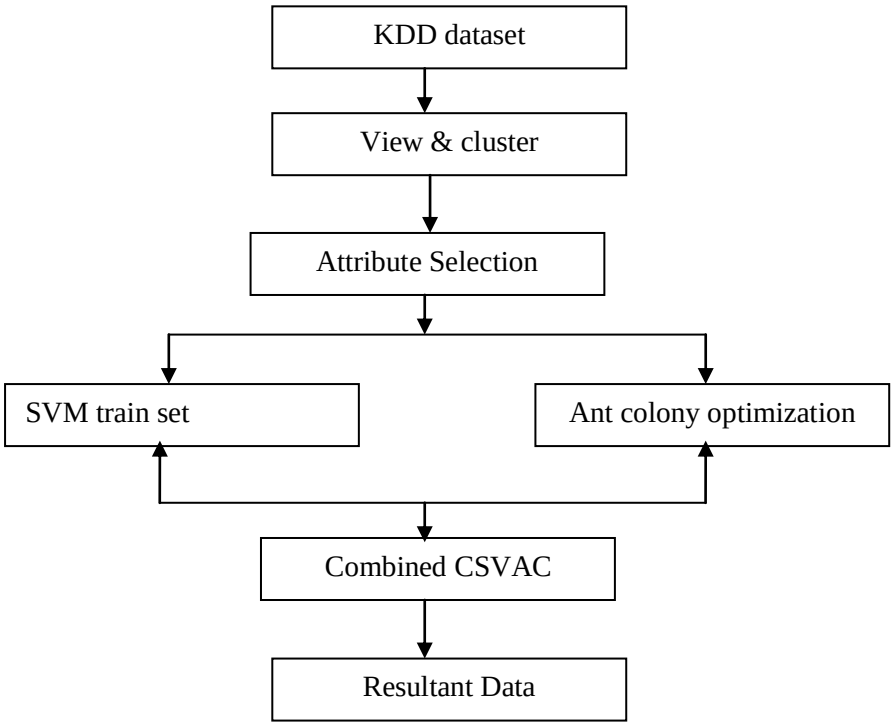

Fig.1. Framework of SVM

Fig.1 shows the framework of SVM active learning. Assume that U is the data set of the entire candidate points for training and T is the data set of points for each SVM training step. There are two independent parts of the active learning machine: f and q. f is the SVM classifying function trained by T. q is a searching function to decide which new data points are selected from U into T for the next SVM training process. We will propose a selection strategy of the searching function q based on a clustering method described in the next section.

VI. CLUSTERING BASED ON A SELF ORGANIZING ANT COLONY NETWORK (CSOACN).

As an ant colony network possesses properties such as flexibility, robustness, decentralization, and self-organization, it can suggest very interesting heuristics. Optimization and control algorithms based on swarm intelligence, including Ant Colony Optimization and Ant Colony Routing, are well known [7]. In the area of data mining, particularly for clustering purposes, there are also many studies using the metaphor of ant colonies [8]. In order to model a self-organized ant colony network, the problem should be described in term of some standard mathematical objects. In the real world of a self-organized ant colony network, a population of ant-like agents moves objects on a two-dimensional grid so as to cluster similar objects into the same region. Object and ant are the two basic entries in the problem. During the process, it is essential to know that whether an object is similar to its neighborhood objects. Swarm similarity is used to describe this property of each object.

6.1. Intrusion detection based on CSOACN

Network connecting records described by several features can be viewed as objects in a CSOACN. These objects belong to different classes (i.e., normal and different kinds of intrusions). As the profiles of both abnormal and normal data are defined as different clusters, an intrusion detection classifier can be constructed with both anomaly and misuse detection pattern by clustering.

VII. COMBINING THE SVM METHOD WITH CSOACNS

We now introduce the new machine learning algorithm, named Combining Support Vector with Ant Colony (CSVAC) that is based on a combination of the modified version of the two algorithms discussed above (i.e., SVM and CSOACN). In this algorithm, SVM and CSOACN are two interactive phases which are taken multiple times. SVM is used to find support vectors and to generate hyper plane that separating normal and abnormal data while a CSOACN is used to find data added to active SVM training set and to finally generate models for normal data as well as for each class of abnormal data.

7.1. SVM

To evaluate the performance of SVM, the component SVM is configured for pure SVM training and the component AC is disabled. Training scheme: Under SVM training mode, the IDS applies the one-vs-all scheme [49] and trains five binary SVM classifiers to distinguish each class of data (Normal, DoS, U2R, R2L, Probe) from the other four. Parameter adjusting: The kernel function used in our experiments is the Gauss function, defined as

$$K(x_i, x_j) = e^{-\frac{(x_i - x_j)^2}{2\sigma^2}}.$$

To select a proper value of σ, we did some testing on the data set D using several different values of σ (0.01, 0.1, 0.5, 1, and 10). The detection rates and training time are shown in Fig. 6(a) and (b), respectively. Considering both accuracy and efficiency, we chose σ = 0.5 as a suitable value for the kernel function. Testing results: Data records in the testing set T1 are classified using the SVM classifiers constructed based on a voting strategy, in which the decision of each binary classification is considered as a vote and the final decision is designated to the class with maximum number of votes. If the maximum vote is not from a unique class, the data point is labeled as ‘‘unknown’’.

7.2. CSOACN

This is the case that the component AC is configured for pure CSOACN clustering and the component SVM is disabled. Training scheme: CSOACN can be directly used for multiclass problems. Under CSOACN training mode, our IDS trains a single classifier which establishes a model for each class of data (Normal, DoS, U2R, R2L, and Probe). Parameter adjusting: As discussed in Section 5, a suitable value of the swarm similarity coefficient β needs to be decided (Eq. (3)).Using several values adjusted from 0.2 to 0.4 with interval 0.05 for β, the detection rates and training times for the set D are shown in (a) and (b). The value β = 0.25 was chosen for further testing. Testing results: By using the CSOACN classifier that was constructed upon the data set D, all data in set T1 can be classified.

7.3. CSVAC

Our new algorithm has both components of SVM and AC enabled. Using values of M adjusted from 1 to 6 with interval 1, the detection rates and training times for the set D are shown in. Considering both efficiency and accuracy, we choose M = 4 to be the value of the parameter. Testing results: By using the CSVAC classifier that was constructed upon the data set D, all data in set T1 can be classified. The testing results are shown in .We also did t-tests on the difference of the mean values with 90% confidence level. The degree of freedom in the t-test for the three algorithms and five observations is defined to be df = 3 ×(5 − 1) = 12. The confidence intervals of the t-scores are given in i = 1, 2, 3, calculated as αi = μi−μ, where μi is the mean value of the measure of algorithm i and μ is the overall mean of the measure across all the three algorithms. The subscripts 1, 2, and 3 represent SVM, CSOACN, and CSVAC, respectively. If the confidence interval does not include zero, we can reject the hypothesis that the difference αi − αj has a zero mean value. From the t-test results, we conclude with 90% confidence that the proposed new CSVAC algorithm is better than both SVM and CSOACN on training time, better than SVM on detection rate, false positive rate and false negative rate, and comparable with CSOACN on these rate measures. The new IDS was also tested using the data set T2 The results are compared with the KDD99 winner and shown in It is noticed that the new algorithm is comparable to the KDD99 winner in terms of average detection rate as well as false positive rate and even better in term of false negative rate.

Table.1 Confidence intervals of *t*-tests.

Measure	Confidence interval	
	a $_{1—}$ a$_2$	$\alpha_2 - \alpha_3$
Training time(s)	(0.659, 1.027)	(2.073, 2.441)
Detection Rate (%)	(−19.648, −3.308)	(−6.250, 10.090)
False Positive (%)	(2.268, 2.853)	(−0.163, 0.423)
False Negative (%)	(19.433, 23.767)	(−2.107, 2.227)

VIII. CONCLUSIONS AND FUTURE WORK

As future work, we are considering integrating the privacy preserving OLAP with the proposed framework in order to improve the effectiveness and the flexibility of IDS system. We also plan to enhance the CSVAC algorithm to generate more SVM classifiers to handle multiclass cases and find ways to convert a nonlinear classification problem to a linear one by applying the recently proposed Maximum Information Coefficient (MIC) method. For further performance analysis, comparisons with other existing algorithms such as those of will be conducted by applying the KDD99 data set but with different distributions and also using other standard benchmark data sets.

REFERENCES

[1]. X.Linclon,A new intrusion detection system using support vector machines and hierarchical clustering.(2004)104-456.
[2]. Y.toan,X,Burges,A Study of Network Intrusion Detection by Applying Clustering Techniques(1999),pp,185-208.
[3]. D.Duan,Network Intrusion Detection by Support Vectors and Ant Colony(2000),851-871.
[4]. V.Ramos,ACO based Distributed Intrusion Detection System(2012),254-654.
[5]. Y.Feng,J.Zhong,C.Ye,ANTIDS:Self Organized Ant-Based Clustering Model for Intrusion Detection System.(2006)247-256.
[6]. Y.Liu,X.Yu,J.X.Huang,Combining integrated sampling with SVM ensembles for learning fron imbalanced datasets,information processing &management 47(4)(2011)617-631.
[7]. C.J.C.Burges,A tutorial on support onvector machines for pattern recognition,Data mining and Knowledge Discovery2(2)(1998)127-167.
[8]. S.Kumar,Classification and detection of computer intrusions,(2005)-(8).
[9]. B.Mukherjee,Network intrusion detection using autonomous agents computer networks 34(4)(2000)567-987.
[10]. J.Han,Data mining concepts and techniques,2012.
[11]. Lin,W.-Y.Lin Intrusion by machines learning.
[12]. S.X.Wu,W.Banzhaf,The use of computational intelligence in intrusion detection systems:(2010).
[13]. C-H.Tang,multi-agent intrusion detection system in industrial network using ant colony clustering approach and unsupervised feature extraction,(2005),51-56.
[14]. S.Kwong ,Ant colony clustering and feature extraction for anomaly intrusion detection,(2006,pp.101-123.
[15]. D.E.Denning an intrusion-detection model,(2000)222-232.
[16]. T.D.Garvey,Model based intrusion detection,(2009).
[17]. S.Freeman,J.Branch,Host-based intrusion detection using SVMs(2002).