Journal of Computer Science 7 (10): 1560-1564, 2011
ISSN 1549-3636

A Novel Local Network Intrusion Detection System Based on Support Vector Machine

Muamer N. Mohammad, Norrozila Sulaiman, Emad T. Khalaf
Faculty of Computer Systems and Software Engineering,
University Malaysia Pahang, Kuantan 26300, Malaysia

Abstract: Problem statement: Past few years have witnessed a growing recognition of intelligent techniques for the construction of efficient and reliable Intrusion Detection Systems (IDS). Many methods and techniques were used for modeling the IDS, but some of them contribute little or not to resolve it. **Approach:** Intrusion detection system for local area network by using Support Vector Machines (SVM) was proposed. First, the intrusion ways and intrusion connecting of Local Area Network were defined for putting forward the design requests on intrusion detection system of LAN. Second, the new method to recognized attack patterns which may give better coverage and make the detection more effective. **Results and Conclusion:** SVM was used as a detection system that recognizes anomalies and raises an alarm. The data that was used in our experiments originated from a campus lab. The result of the evaluation produced a better result in terms of the detection efficiency and false alarm rate.

Key words: Intrusion detection, support vector machine, Radial Bias Function (RBF), Network Intrusion Detection System (NIDS), local area network, Standard Security Mechanism (SSM)

INTRODUCTION

In recent years, the rapid development of artificial intelligent techniques has got a large quantity of algorithms from the fields, such as statistics, pattern recognition, machine learning and database and some algorithms are particularly useful for intrusion detection, such as classification analysis, cluster analysis, association rule analysis and sequential pattern analysis the previous studies show that applying these technologies to intrusion detection is feasible and effective.

Many research focused on Intrusion detection dating back to the work of Anderson (1980) and (Syurahbil *et al*., 2010), which is a model based on the hypothesis that security violations can be detected by monitoring a system's audit records for abnormal patterns of system usage. Intrusion detection is the process of monitoring the events occurring in a computer system or network and analyzing them for sign of intrusion. A Network Intrusion Detection System (NIDS) is a system that emits alerts commensurate with abnormal or unauthorized events in the network (Golzari *et al*., 2001). Snort (Wahid and Zulkarnain, 2011) as a popular NIDS is widely used to audit network packets and compare them with a database of known attack signatures (Ektefa *et al*., 2011; Lundin and Jonsson, 2002; Sodiya *et al*., 2004). This technique appeared to be promising, but there are some problem in structural and system performance. In addition, combining multiple techniques in designing the IDS is a recent event and it needs further improvement. Valdes and Skinner (2001) suggested an approach by using sensor correlation, which means that alarms from different components in the detection system are analyzed and correlated at different levels. Another method to correlate and draw conclusions from data which can be gathered from many distributed sources was multi-sensor data fusion.

In this study, it is hoped that the detection method can be improved by designing a more effective intrusion detection system using intelligent techniques. The collected data are stored for batch-mode analysis or immediately analyzed in real-time environment. Indeed the main advantages of applying SVM to an intrusion detection system lie in that the system can produce an accurate detection model from a mass of audit data automatically to reduce artificial intervention and it can be used to construct an intrusion detection system in various computing environments because of universality of mining process itself.

Corresponding Author: Muamer N. Mohammad, Faculty of Computer Systems and Software Engineering, University Malaysia Pahang, Kuantan 26300, Malaysia

Related background: Intrusion detection is needed as another level of security to protect local network systems. Signature-based analysis is a technique that was proposed earlier. It was widely used in the intrusion detection community to protect a system by using a combination of an alarm that sounds whenever the security sites has been compromised, with Standard Security Mechanism (SSM). Indeed, IDS are also considered as a complementary solution to firewall technology by recognizing attacks against the network that are missed by the firewall.

When a standard security mechanism is taking some actions to prevent the system from a threat, the engineering or a local intrusion detection system might be interested in such information. For this a policy has to be defined, when and how alerts and logging messages are processed so it can respond to the alarm and take the appropriate action, for instance by ousting the intruder, calling the proper external authorities and so on. Intrusion systems are noted for high false alarm rates and considerable research effort is still concentrated on finding effective intrusion, non-intrusion discriminates (Golzari *et al*., 2001; Wahid and Zulkarnain, 2011). It was suggested by Lundin and Jonsson (2002) that techniques should be combined in order to correct some of these problems. Sodiya *et al*. (2004), a strategy that effectively combined strategies of data mining and expert system was used to design an Intrusion Detection System (IDS).

Intrusion detection is critical components of information security system are used to detect suspicious activity both at network and host level. There are two main approaches to design an IDS.. There are two main categories of intrusion detection:

- Misuse based IDS (signature based)
- Anomaly based IDS

In a misuse based intrusion detection system , intrusions are detected by looking for activities that correspond to know signatures of intrusions or vulnerabilities (Golzari *et al*., 2001). While an anomaly based intrusion detection system detect intrusions by searching for abnormal network traffic. The abnormal traffic pattern can be defined either as the violation of accepted thresholds for frequency of events in a connection or as a user's violation of the legitimate profile developed for normal behavior.

One of the most commonly used approaches in expert system based intrusion detection systems is rule-based analysis using Denning's profile model (Golzari *et al*., 2001). Rule-based analysis depends on sets of predefined rules that are provided by an administrator. Expert systems require frequent updates to remain current. This design approach usually results in an inflexible detection system that is unable to detect an attack if the sequence of events is slightly different from the predefined profile (Wahid and Zulkarnain, 2011). Considered that the intruder is an intelligent and flexible agent while the rule based IDSs obey fixed rules. This problem can be tackled by the application of soft computing techniques in IDSs. Soft computing is a general term for describing a set of optimization and processing techniques.

Although support vector machines have become the key techniques for anomaly intrusion detection due to their good generalization nature and the ability to overcome the curse of dimensionality (Lundin and Jonsson, 2002; Sodiya *et al*., 2004), the main issue of SVM technique applied to intrusion detection is its low efficiency.

Theoretical background:

Intrusion detection: An intrusion detection system consists of an audit data collection agent that collects information about the system being observed. This data is either stored or processed directly by the detector. The output is presented to the SSO, where further action will be taken. Normally it involves further investigation into the causes of the alarm.

Over the years, researchers and designers have used many techniques to design intrusion detection systems. However, there are some problems with the present intrusion detection systems which include:

High number of false positives: False alarms are high and attack recognition is not accurate. By lowering thresholds to reduce false alarms raises the number of attacks that get through undetected as false negatives. Improving the ability of an IDS to detect intrusion accurately is the primary problem facing IDS manufactures today.

High number of false negatives: Some intrusions are still undetected in some systems which mean that the IDSs are not able to detect all computer intrusions. Thus, improving the ability of an IDS to detect attacks is another major problem facing by researchers.

Lack of efficiency: IDSs are often required to evaluate events in a real time. This requirement is difficult to meet when a system faced with a very large number of events which is typical in today's networks. Consequently, host-based IDSs often slow down the system and network-based IDSs will drop network packets that they do not have time to process.

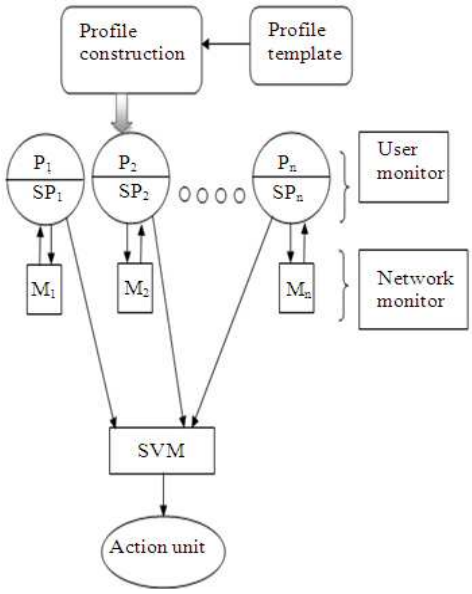

Fig. 1: A SVM model for intrusion detection system

IDS security: Few papers discuss IDS resilience, i.e. the ability of the IDS to resist attacks against itself. One of the papers describes the network IDS. If an attacker is aware that an intrusion detection system exists, the attacker will probably start by studying the IDS to be able to shut it down, cripple it, or circumvent it. The IDS will be the first point of attack, since the attacker can work undisturbed when the IDS is not in operation.

However, in this research, we hope to improve on these previous works to design a more effective intrusion detection system that combines the Support Vector Machine and expert systems (Golzari *et al*., 2001). The interest in this work is to improve detection efficiency by reducing or eliminating false positives and false negatives. IDS security is also a major concern.

Support Vector Machines (SVMs): Several extensions have been proposed to make SVMs suitable to deal with multi-class classification problems (Hsu and Lin, 2002). Although none of the multi-class approaches known in the literature is accepted as a solution to generic problems, SVMs techniques are nowadays mature enough to be applicable to many classification problems (Chen *et al*., 2005).

The SVM approach transforms data into a feature space F that usually has a huge dimension. It is interesting to note that SVM generalization depends on the geometrical characteristics of the training data, not on the dimensions of the input space. Training a support vector machine leads to a quadratic optimization problem with bound constraints and one linear equality constraint. Vapnik (Joachims, 1998) shows how training a SVM for the pattern recognition problem leads to the following quadratic optimization problem (Buntod *et al*., 2010):

Minimize:

$$W(a) = -\sum_{i=1}^{1} a_i + \frac{1}{2}\sum_{i=1}^{1}\sum_{j=1}^{1} y_i y_j a_i a_j k(x_i, x_j) \quad (1)$$

$$\text{Subject to} \sum_{i=1}^{1} y_i a_i \forall_i \text{ :} \leq a_i \leq C \quad (2)$$

Where:

l = The number of training examples

a = A vector of l variables and each component i a corresponds to a training example (xi, yi)

The solution of (1) is the vector a* for which (1) is minimized and (2) is fulfilled.

MATERIALS AND METHODS

System description: A user supports a sequence S if S is contained in the user-sequence for this user. The definition of support is given as the fraction of the total number that exists in the sequence. In the sequential pattern profiling, user daily activities were taken as a sequence and a database for each user containing users' daily sequential patterns was created. Figure 1 shows a SVM model for intrusion detection.

Definitions 1: normal users profiling: Let {U1,U2, . . . ,Un} be the set of users in system S.

Let {T1,T2,...,Tn} be the set of transactions in system $S \ni \forall T_i \in A$. A represents the audit trail or system log.

Let U1 = {T1, T2, . . . ,Tj} ϵ A be a set of transactions made by user 1. The occurrence of U1 in this case, is the number of transactions made by user 1 which is j. Researchers have been seeking for efficient solutions to the problem of creating an effective and dynamic users profile. One of the unique concepts that have been introduced to correct the problem is the use of a monitor.

Definition 2: monitor: Let {U1,U2, …,Un} be a set of users in system $S \ni \exists M_j \forall U_j \in S$, where j = 1, 2, . . ., n. Mj are called monitors which are used to assist in providing effective detection. Monitors are used to keep track of false positive events in each profile and later used to update the profiles.

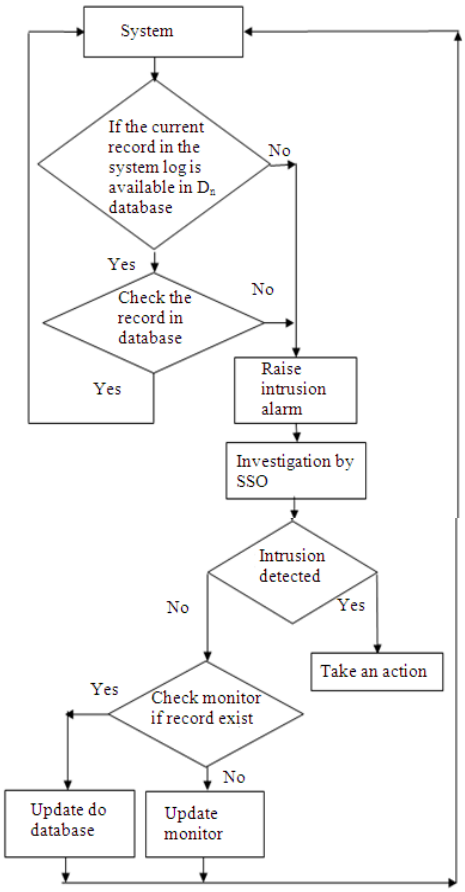

Fig. 2: Over all proposed algorithm for IDS

Definition 3: sequential pattern profiling: An itemset is a non-empty set of items. A sequence is an ordered list of itemsets. We denote a sequence, S by (S1,S2, . . ., Sn), where Sj is an itemset. All activities of a user with the system together can be viewed as a sequence where each activity corresponds to a set of items and the list of activities, ordered by increasing transaction time, corresponds to a sequence formally. Let the activities of a user in a period ordered by increasing transaction-time, be T1,T2, . . .,Tn. Let the set of items in Ti be denoted by itemset (Ti). A user-sequence can then be represented by:

Sequence: (itemset (T1), itemset (T2), . . . itemset (Tn))

Software design: Figure2 shows the SVM algorithm for the detection system in our design. Two major considerations in detection system design are the inference engine the pervious knowledge. The detection begin first check is made to conform from the current record in the system. Second, the SVM system moves forward for sequential pattern relevance, it would return back to the system to decide if it is normal or not to raises an alarm. Third, the intrusive found then it will go to checking and updating the monitor.

RESULTS AND DISCUSSION

Data set: The LAN was operated in a real environment, but being blasted with multiple attacks. For each TCP/IP connection, 41 various quantitative and qualitative features were extracted. Out of this database, a subset of 59261 data was used, of which 20% represent normal patterns. The data was partitioned into two different classes: normal and attack, where the attack is the collection of all different attacks belonging to the four classes. The objective of our SVM experiments is to separate normal and attack patterns. Data points were randomly generated which contain actual attacks and normal usage patterns. Training was done using the Radial Bias Function (RBF) kernel option; an important point of the kernel function is that it defines the feature space in which the training set examples will be classified.

Testing: Our test procedure is outlined in Fig. 1. First, the data was converted and represented by the frequency distribution of system calls. The training data set was then separated into attack data sets and normal data sets, which were then subsequently fed into SVM algorithm. Through the training process SVM predictive models can be built. Then the test data set was fed into SVM predictive models. In our second set of experiments, the data consists of 14000 randomly generated points, with a number of data from each class in proportion to its size. A training set of which 41 features and 13 features each were used respectively. The results were summarized in the following Table 1.

The top-left entry of Table 2 shows that 10864 of the actual “normal” test set were detected to be normal; the last column indicates that 98.5 % of the actual “normal” data points were detected correctly. In the same way, for the attack class 44013 of the actual “attack” test set were correctly detected; the last column indicates that 99.3% of the actual “attack” data points were detected correctly.

The bottom row shows that 99.3% of the test set said to be “normal” indeed were “normal” and 99.9% of the test set classified as “attacks” indeed were attacks.

Table 1: SVM testing results

Testing	Test 1	Test 2	Test 3	Test 4
Test data set	7000.00	7000.00	55000.00	55000.00
Of features	13.00	41.00	13.00	41.00
Accuracy %	99.52	99.53	99.57	99.60
CPU run time	1.06	1.60	10.04	15.44
misclassifications	35.00	33.00	234.00	230.00
False positives	19.00	17.00	102.00	122.00
False negatives	15.00	16.00	132.00	98.00

Table 2: Results of the main test set with 41 features and 55000 data points

Class	Normal	Attack	Accuracy
Normal	10864.00	122.00	98.50%
Attack	98.00	44013.00	99.30%
Accuracy	99.30%	99.50%	

Table 3: Comparing execution time between our proposed model with Chen model

No. of attacks	Execution time (sec) Chen model	Execution time (sec) proposed model
24	18	≈0.00
400	24	≈0.00
1000	32	0.05
3000	38	0.06
5000	49	0.11
7000	104	0.17

In addition, we compared the execution time of the Chen model (Chen *et al*., 2007) with the execution time of our model. The results are given in Table 3 shows that the detection execution time using Chen highly increases when the number of attacks augments. In our proposed model the execution time is somewhat stable, even considering thousands of attacks. On the other hand, our model deduces the class of each detected attack.

CONCLUSION

Nowadays improving the ability of IDS to detect attacks accurately is the primary problem facing in IDS manufactures .It is known that some intrusions are still go undetected in some systems. This shows that the current IDSs still cannot detect all intrusions. A good intrusion detection system should perform with a high precision and a high recall, as well as a lower false positive rate and a lower false negative rate. To consider both the precision and false negative rate is very important as the normal data usually significantly outnumbers the intrusion data in practice. Finally, on the basis of this algorithm, an intrusion detection system model based on pattern matching algorithm is put forward. In addition, the test result shows that the proposed method surely improved detection efficiency by reducing or eliminating false positives and false negatives and reduce the run-time complexity of IDS. The main weakness of our method is its exclusive dependence on SVM performance. Future work, the proposed and discussed approach may be extended in gain to integrating another intelligent techniques (PSO) into one hybrid intelligent-IDS and investigate the possibility and feasibility of implementing this approach in real time intrusion detection environments.

REFERENCES

Buntod, P.C., P. Suksringam and A. Singseevo, 2010. effects of learning environmental education on science process skills and critical thinking of mathayomsuksa 3 students with different learning achievements. J. Soc. Sci., 6: 60-63. DOI: 10.3844/JSSP.2010.60.63

Chen and Yuehui, 2007. Hybrid flexible neural-tree-based intrusion detection systems. Int. J. Intell. Syst. 22: 337-352. DOI: 10.1002/INT.20203

Chen, P.H., C.J. Lin and B. Schkopf, 2005. A tutorial on m-support vector machines. National Taiwan University.

Ektefa, M. F. Sidi, H. Ibrahim, M.A. Jabar and S. Memar, 2011. Comparative study in classification techniques for unsupervised record linkage model. J. Comput. Sci., 7: 341-347. DOI: 10.3844/jcssp.2011.341.347
DOI: 10.3844/JCSSP.2011.341.347

Golzari, S., S. Doraisamy, M.N. Sulaiman and N.I. Udzir, 2001. An efficient and effective immune based classifier. J. Comput. Sci., 7: 148-153. DOI: 10.3844/JCSSP.2011.148.153

Hsu, C.-W., C.-J. Lin, 2002. A comparison of methods for multiclass support vector machines. IEEE Trans. Neural Networks, 13: 415-425. DOI: 10.1109/72.991427

Joachims, T., 1998. Making Large-Scale SVM learning practical. University of Dortmund.

Lundin, E. and E. Jonsson, 2002. Survey of intrusion detection research. Citeulike.

Sodiya, A.S., H.O.D. Longe and A.T. Akinwale, 2004. A new two-tiered strategy to intrusion detection. Inform. Manage. Comput. Security, 12: 27-44. DOI: 10.1108/09685220410518810

Syurahbil, N. Ahmad, M.F. Zolkipli and A.N. Abdalla, 2009. Intrusion preventing system using intrusion detection system decision tree data mining. Am. J. Eng. Applied Sci., 2: 721-725. DOI: 10.3844/AJEASSP.2009.721.725