Imperial Journal of Interdisciplinary Research (IJIR)
Vol-3, Issue-4, 2017
ISSN: 2454-1362, http://www.onlinejournal.in

A Peculiar Amalgam Facet for Two-Level Intrusion Detection Using K-Means + C 5.0

Ms. Meghana Solanki[1] & Mr. Pranav Pathak[2]
[1]Assistant Professor, DYPCOE, Ambi, Talegaon
[2]Assistant Professor, AGCE, Panmalewadi, Satara

Abstract*: The strengths of misuse detection as well as anomaly detection are exploited by combining two intrusion detection systems. From an innovative aspect, in this paper, we achieve high detection rate with a low false positive rate by proposing a hybrid approach. This approach is a two-level hybrid solution. It consists of two anomaly detection segments and a misuse detection segments. In phase1, an anomaly detection method is evolved. It is hired to build the detection segment. The k-means + C 5.0 algorithm becomes very decisive in constructing the two detection segments for phase 2. In this hybrid approach, there is a well coordination between all of the detection segments. The detection segment of phase 1 inclined to build the two detection segments of phase 2.It reduces the false positives as well as false negatives given by the detection segment of phase 1.Experimental outcomes on the KDD'99 dataset demonstrate that the proposed hybrid approach can adequately search out network deviation with a low false positive rate.*

1. Introduction

In recent times, an Internet threat causes the potential damage. It has turn into more genuin problem. It's become necessary to protect users' systems from these threats. Intrusion detection systems (IDSs) are supposed to minimize the serious effect of such invasion [1].

An IDS is a security counter measure. It identifies a set of malicious actions which compact the integrity, confidentiality, or availability of information assets [2]. It monitors network traffic to find out whether a network attack such as a probing attack is targeting any system or not. There are two main approaches to intrusion detection such as misuse detection as well as anomaly detection .These approaches are based on types of analysis. The patterns of known attacks are identified by misuse detection system. Any deviation from the profiles of normal activity is captured by an anomaly detection system. The misuse detection system is build based on attack information. The anomaly detection system is build based on normal behavior information. Both misuse and anomaly detection systems are related to some gain as well shortcomings. The misuse detection system detects known attacks with a low false positive rate. The Unknown attacks may rescue detection because we can't define all possible attacks before we have observed them. The performance of anomaly detection system is better than misuse detection systems in two facets. One facet is the ability to find out new types of attacks. And second facet is that the established profiles of normal behavior are customized not only to system, application but also to network. So it becomes very difficult for a mugger to recognize a exact action that can be done without being exposed. An anomaly detection system also poses some deficiency. It is not able to usually discover boundaries between normal as well as abnormal behavior which lead to a high false positive rate. There is need to complements strengths of both IDS; to do this hybridization of both systems is needed. The hybrid approach consists of misuse as well as anomaly detection segments. This aspect is useful to improve the detection rate but it is not useful with a high false positive rate. When an IDS categorize normal activity as malicious then it becomes a false positive. Such activities turn into a phase of time on incorrect reports. It also potentially avoids real attacks being generated. In this paper, hybrid approach of a two-level is presented from an innovative perspective. In this system the detection segment of phase 1 participate in the building of the two detection segments of phase 2.Also the two detection segments of phase2 can recognize the false negatives as well as false positive produced by the detection segment of phase1. Moreover, an anomaly detection method based on the change of position of cluster centers is proposed and issued to construct the anomaly detection segment in phase1.The k-means + C 5.0 algorithm is employed to build the two detection segment in phase 2.The proposed hybrid approach is evaluated by performing experiments on the KDD'99 benchmark dataset . The experiment outcomes demonstrate that this hybrid aspect poses the ability to find out known as well as unknown attacks with a low false positive rate.

2. Literature Survey

In this paper [3, 4] author presented Hybrid IDS. It consists of methods that combine multiple machine learning techniques, phased or layered approaches as

well as ensemble approaches. In this paper [5] author choose the C4.5 decision tree as well as a one-class support vector machine to construct a misuse detection segment together with anomaly detection segments, respectively. In this paper [6] author provided three-tier IDS with a similar design. They have made an additional use of misuse detection segment subsequently the anomaly detection segment. In this paper [7, 8, and 10] author use an anomaly detection segment to determine suspicious behavior. After they apply a misuse detection segment to find out whether these behaviors are intrusions. In addition, in this paper [8] author introduces a new artificial immune system to search anomalous network connections. Afterwards he uses a Kohonen Self-Organizing Map for attack type classification. In this paper [9] author proposed Next generation intrusion detection expert system (NIDES).It is an early IDS system. It introduces expert rules to find out known attacks in case of a misuse detection segment. And suspicious anomalies are recognized by an anomaly detection segment. Then, a resolver segment is employed for guessing whether to generate alerts. In this paper [10] author proposed Anomaly detection first serial combination (ADFSC). It makes the use of an anomaly detection segment to drain HTTP traffic declared safe. And then afterward the remaining traffic with unknown declaration is tested by the misuse detection segment. In this paper [11] author presented a solution named combined strangeness and isolation measure k-nearest neighbors (CSI- KNN). It makes the use of two anomaly detection segments under a parallel structure as well as a correlation unit. In this paper [9, 12] author proposed the last category of parallel hybrid approaches. In this approach author uses a misuse detection segment as well as an anomaly detection segment are combined in parallel. And then their results are forwarded to a correlation segment before taking t any final decision. In this paper [12] author proposed a hybrid IDS. It uses not only a misuse detection module but also an anomaly detection module. It also uses a decision support system In this paper [13, 14] author proposed an approaches in which the first category use a misuse detection segment to recognize known attacks and in second category he uses an anomaly detection segment to find out either unknown or uncertain attacks and behaviors.

3. A Newly Introduced Hybrid Aspect for Intrusion Detection

Anomaly based methods are attracting the most attention from researchers, because of their capability to find out known along with unknown attacks. Regrettably, these methods many times go through high false positive rates. To find out solution to this problem, some works adopt hybrid construction approaches which use anomaly detection together with misuse detection. However, they face problem with detecting attacks that do not varies significantly from normal network behavior .They hold low false positive rates. Before giving the structure of proposed hybrid approach, we show the basic ideas with the help of example. Consider a dataset shown in Figure1, in which items filled with color of white represent normal behavior and items filled with color of black represent malicious behavior. Malicious items in area1are similar to normal items that bring difficulty to anomaly detection methods. Favorably, they have great resemblance to one another in area1. So, a well-trained misuse detection method can recognize them accurately. In case of a low anomaly threshold, an anomaly detection method should capture malicious items in area 2.

Anomaly detection segment 1 detects network connections initially for the scheduled aspect. Then, the connections which are declared as abnormal will be forwarded to anomaly detection segment 2 for further assessment of attacks or false positives; in parallel, the misuse detection segment in phase 2 further audit the remaining connections which is declared as normal. They will be tagged with the corresponding attacks if they are very identical to the attacks in the training data. Any other way, they are normal connections. therefore, in the proposed hybrid aspect, anomaly detection segment 2 as well as the misuse detection segment are applied to categorize not only the false positives but also the false negatives which is produced by anomaly detection segment 1. A Proposed Hybrid aspect is shown in Figure 2.

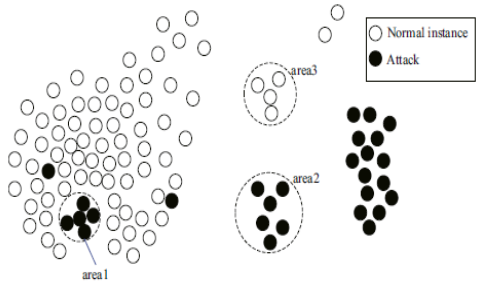

Figure 1.Graphic of relation of item

3.1 Anomaly detection segment 1

Anomaly detection segment1of phase1 is supposed to gain a high detection rate along with high efficiency by considering the structure of the proposed hybrid aspect. Thus, we schedule an innovative anomaly detection method based on the change of cluster centres (ADMBCC).It is used to

build anomaly detection segment1.A cluster is a set of items with the most likely qualities. They are gained by using algorithms such as K-means or CURE. A cluster centre is a "reference item" for ADMBCC. It is defined by a merger of items in a cluster.

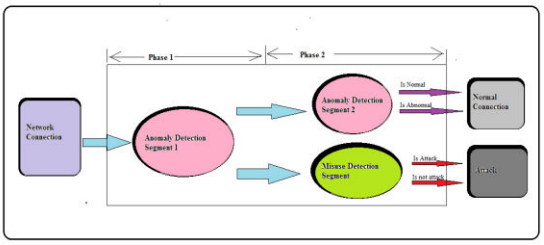

Figure 2. Overview of the revealing method of the proposed hybrid aspect.

3.2 Anomaly detection segment 2

A new arrival item will be categorized as normal or abnormal in phase1. However, the anomaly detection method often leads to false positives. If an item is categorized as abnormal by anomaly detection segment 1, then that item will be further audited by anomaly detection segment 2.It turns out into a decision of either false positive or attack. Therefore, the duty of anomaly detection segment 2constructed by introducing k-means + C 5.0 is to recognize the false positives produced by anomaly detection segment1.

3.3 Misuse detection segment

New arrival item announced normal by anomaly detection segment 1, will be further analyzed by the misuse detection segment. It confirms whether an item is a negative positive or a normal instance. Therefore, the duty of the misuse detection segment in phase 2 is to recognize the negative negatives outputted by anomaly detection segment 1. For this objective, k-means + C 5.0 model is used.

4 Algorithm

K-MEANS+C 5.0 METHOD FOR ANOMALY DETECTION

In this section, we have introduced a hybrid Intrusion detection system by cascading K–Means and C 5.0 decision tree algorithm. This method is divided into training phase & test phase. The k-Means algorithm clubs N data items into k disjoint clusters, where k is a predefined parameter. C 5.0 is an algorithm used to develop a decision tree. The decision trees provoked by C 5.0 are applicable in case of classification or segregation task. C 5.0 is greatly faster than C 4.5. It is more memory efficient. It gets identical outputs to C 4.5 by using smaller decision trees. It supports boosting process. It permits you to weight different cases as well as misclassification types. It automatically scatters the attributes to remove those that may be unhelpful.

K–Means + C 5.0 Algorithm

Selection Phase
Input: Test items Ti, *i* 1,2,3,.....,*N* .
Output: Closest cluster to the test item Ti.
Procedure Selection
Begin
Step 1: For each test item *Ti*.
a. Enumerate the Euclidean distance D(Ti, rj),j=1...k, and find the cluster closest to
Ti.
b. Compute the C 5.0 Decision tree for the closest cluster.
End /*End Procedure*/

Classification Phase
Input: Test item Ti.
Output: Segregate test item Ti as normal or anomaly
Procedure Classification
Begin
Step 1: Assign the test item Ti over the C 5.0 decision tree of the computed
closest cluster.
Step 2: Segregate the test item Ti as normal or anomaly and include it in the cluster.
Step 3: Renew the centre of the cluster.
End /*End Procedure*/

5 Experimental Results

5.1 Datasets

We have done an analysis on the KDD'99 dataset for verifying the effectiveness as well as performance of the proposed hybrid aspect. There are several limitations on the KDD'99 dataset; still it remains not only as a standard but also important dataset. It presents a classic challenge. It is mostly used in the architecture of network intrusion detection. There are 3 labels in KDD'99dataset which are the full training set, the 10% training set along with the test set. There are 41 features which are related to each record in these datasets. A label gives its category. There are the attack records in the KDD'99 dataset. They are generalized to four classes basic attacks, particularly DoS, Prb, U2R and R2L. In the performed evaluation, all of the Normal, U2R and R2L items in the 10% training set are proffered. The normal items are used to construct anomaly

detection segment1 as well as anomaly detection segment 2.The misuse detection segment is constructed by using U2R and R2L items. The test set which consists of 401,517 records in case of corrected KDD dataset is employed in the test phase. The test set encompasses 18 new types of attack that do not lie in the10% training set. The sizes of the training and test datasets are given by Table1. Note that U2R and R2L items were initially filtered out by anomaly detection segment1.Afterward they are used in the construction of misuse detection segment.

Table 1. Size & distribution of records in training & test sets.

Class	Training set size	Test set size
Normal	97,195	72,419
Prb	42	3602
Dos	65	246421
U2R	64	252
R2L	1021	15146
Total	98,345	410043

5.2 Performance Evaluation Metrics

The performance of the new introduced hybrid aspect is tested by using several widely used metrics. These metrics are detection rate (DR), true negative rate(TNR), false positive rate(FPR) and accuracy(ACC).There are four basic metrics which are used in the calculation of these metrics such as, true positives(TP), true negatives(TN), false positives (FP), and false negatives(FN). DR, TNR, FPR and ACC are retrieved by

DR=TP/ (TP + FN)
TNR=TN/ (TN+FP)
FPR=FP/ (FP+TN)
ACC= (TN+TP)/ (TN+TP+FN+FP)

5.3 Experimental Results

Results with KDD'99 dataset

Performance of detection for the newly introduced hybrid aspect in case of the binary classes i.e. normal and attack, on the test set from the KDD'99 dataset is shown in this zone. The required parameters are cluster numbers, number of items etc. To achieve a better observation, our baseline is the detection performance of anomaly detection segment 1(ADMBCC).The comparison of the detection performances between ADMBCC and newly introduced hybrid aspect on the test set is given by Table2. The detection performances on different connection classes are detailed in Table3. For every comparison, the best performance is given in boldface.

Table 2. Detection performances of ADMBCC and hybrid aspect on the test set.

Detection aspect	DR (%)	FPR (%)	ACC (%)
ADMBCC	91.59	3.05	92.31
Hybrid Aspect	**91.96**	**0.61**	**93.59**

Table 3. Detection performances of ADMBCC and hybrid aspect on the five classes.

Detection aspect	Normal (%)	Dos (%)	Prb (%)	U2R (%)	R2L (%)
ADMBCC	98.40	**97.44**	**98.74**	**84.34**	14.03
Hybrid Aspect	**99.33**	97.04	97.99	78.97	**15.01**

6. Conclusions

In this newly introduces aspect, we are integrating misuse as well as anomaly detection methods. A hybrid aspect gets profit from both of their firmness. In this paper, a hybrid aspect of a two-level system is presented from an innovative perspective. In this system the detection segment of phase 1 participate in the building of the two detection segments of phase 2.Also the two detection segments of phase2 can recognize the false negatives as well as false positive produced by the detection segment of phase1. Moreover, an anomaly detection method based on the change of position of cluster centers is proposed and issued to construct the anomaly detection segment in phase1.The k-means + C 5.0 algorithm is employed to build the two detection segment in phase 2.The proposed hybrid aspect is evaluated by performing experiments on the KDD'99 benchmark dataset .

7. References

[1] H.J. Liao, C.H.R. Lin, Y.C. Lin, K.Y. Tung, "Review: Intrusion detection system: a comprehensive review", J. Netw. Comput. Appl.36(2013)16–24.
[2] P. Dokas, L. Ertoz, V. Kumar, A. Lazarevic, J. Srivastava, P.N. Tan, "Data mining for network intrusion detection", in: Proceedings of NSF Workshop on Next Generation Data Mining, 2002, p. 15.
[3] C.F.Tsai,Y.F. Hsu, C.Y. Lin, W.Y. Lin, "Intrusion detection by machine learning : A review", ExpertSyst.Appl.Int.J.36(2009)11994–12000.
[4] C. Xiang, P.C. Yong, L.S. Meng, "Design of multiple-level hybrid classifier for intrusion detection system using Bayesian clustering and decision trees", Pattern Recognit.Lett.29(2008)918–924.

[5] G. Kim, S. Lee, S. Kim, "A novel hybrid intrusion detection method integrating anomaly detection with misuse detection", ExpertSyst.Appl.41(2014)1690–1700.
[6] T.S. Hwang, T.J. Lee, Y.J. Lee, "A three-tier IDS via data mining approach, in: Proceedings of Annual ACM Workshop on Mining Network Data", Minenet' 07, 2007,pp.1–6.
[7] D. Barbara, J. Couto, S. Jajodia, L. Popyack, N. Wu, "ADAM: detecting intrusions by data mining", in: Proceedings of IEEE Workshop on Information Assurance and Security, 2001, pp.11–16.
[8] S.T. Powers, J. He, "A hybrid artificial immune system and Self Organising Map for network intrusion detection", Inf.Sci.(NY)178(2008)3024–3042.
[9] D. Anderson, T. Frivold, A. Valdes, Next-Generation "Intrusion Detection Expert System (Hides): A Summary", Contract, 1995.
[10] E. Tombini, H. Debar, L. Me, M. Ducasse, "A serial combination of anomaly and misuse IDSes applied to HTTP traffic", in: Proceedings of Twentieth Annual Computer Security Applications Conference, 2005, pp.428–437.
[11] L. Kuang, M. Zulkernine, "An anomaly intrusion detection method using the CSI- KNN algorithm", in: ACM Symposiumon Applied Computing, 2008, pp.921–926.
[12] O. Depren, M. Topallar, E. Anarim, M. K. Ciliz, "An intelligent intrusion detection system (IDS) for anomaly and misuse detection in computer networks", Expert Syst. Appl.29(2005)713–722.
[13] Y.Liao, V.R. Vemuri, "Use of K-Nearest Neighbor classifier for intrusion detection", Comput. Secur. 21(2002)439–448.
[14] J. Zhang, M. Zulkernine,"A hybrid network intrusion detection technique using random forests", in: First International Conference on Availability, Reliability and Security, 2006. ARES2006, 2006, pp.262–269.