Unknown Attack Detection by Multistage One-Class SVM Focusing on Communication Interval

Shohei Araki[1], Yukiko Yamaguchi[2], Hajime Shimada[2], and Hiroki Takakura[2]

[1] Nagoya University, Graduate School of Information Science, Aichi, Japan
araki@net.itc.nagoya-u.ac.jp
[2] Nagoya University, Information Technology Center, Aichi, Japan
{yamaguchi,shimada,takakura}@itc.nagoya-u.ac.jp

Abstract. Cyber attacks have been more sophisticated. Existing countermeasures, e.g, Intrusion Detection System (IDS), cannot work well for detecting their existence. Although anomaly-based IDS is considered to be promising approach to detect unknown attacks, it still lacks the ability to distinguish sophisticated attacks from trivial known ones. Therefore, we applied multistage one-class Support Vector Machine (OC-SVM) to detect such serious attacks. At the first stage, two training data are retrieved from traffic archive. The one is used for training OC-SVM and then, attacks are obtained from the another. Also testing data from real network are examined by the same OC-SVM and attacks are extracted. The attacks from the traffic archive are used for training OC-SVM at the second stage and those from real network are analyzed. Finally, we can obtain unknown attacks which are not stored in archive.

Keywords: Intrusion Detection System, anomaly detection, network security.

1 Introduction

In recent years, a threat of cyber attacks over the Internet has become a serious issue. An attacker attacks computer systems and networks in order to steal confidential information for earning money in black market. These cyber attacks become more varied and sophisticated year by year.

To detect cyber attacks, Intrusion Detection System (IDS) plays an important role. There are two types of IDS by detection method: a signature-based IDS and an anomaly-based IDS. The signature-based IDS detects attacks by pattern matching of predefined signatures. It has high detection rate, but it cannot detect unknown attacks. On the other hand, an anomaly-based IDS learns normal behavior of network traffic and extracts abnormal values of network traffic behavior as attacks. Although the anomaly-based IDS can detect unknown attacks, it has some problems. Above anomaly-based methods cannot tell whether the detected attacks are known attacks or not. Furthermore, it is impossible to show the seriousness of each unknown attack which affects on the network. Especially,

C.K. Loo et al. (Eds.): ICONIP 2014, Part III, LNCS 8836, pp. 325–332, 2014.

in case of a targeted attack, specially-crafted tools are used along with other conventional tools which cause the barrage of trivial attacks. A network administrator has to search the most serious attack among huge amount of unknown attacks. Therefore we need some method to extract such serious attack.

In this paper, in order to solve above problems, we introduce 6 new features to Kyoto2006+ Dataset [2]. Also new method is proposed to detect serious attacks by using multistage one-class Support Vector Machine (OC-SVM [6]). The multistage OC-SVM uses three sets of traffic, two sets retrieved from a traffic archive and one extracted from real network. At the first stage, OC-SVM learns older archive set and then analyzes newer archive set and one from real network. At the second stage, OC-SVM learns outlier traffic from the newer archive set and analyzes that from the real network. As a result, extracted traffic from outlier of the real network which does not exist in the newer set can be extracted, and we should pay attention to it as possible serious attack.

We evaluated our method using Kyoto2006+ Dataset and 6 new features. The results show that our method detects attacks with higher accuracy than by using only conventional features and successfully extracted unknown attacks.

2 Related Works

Several researcher proposed anomaly-based IDS methods using OC-SVM. Eskin et al. [3] proposed a method of anomaly detection. They compared methods based on OC-SVM, clustering and k-nearest neighbor and show higher detection rate at OC-SVM than others. Prdrisci et al. [5] proposed a method that extracts features from payload and detects attacks using OC-SVM.

An approach of unknown attack detection, Song et al. [8] proposed a method that extracts new features from alerts of a signature-based IDS and detects unknown attacks. 0-day attack shows quite irregular behavior from known attacks concerning packet size and communication interval because an attacker confirms the effectiveness of the attack. These irregular characteristic often raises alerts of the signature-based IDS so that they can detect 0-day attack by analyzing alerts. This method uses alerts of a signature-based IDS, so it cannot detect unknown attacks that a signature-based IDS does not report any alert.

Above anomaly-based methods cannot detect whether the detected attacks are known attacks or not. In other words, these methods cannot extract only unknown attacks. Because various countermeasures are deployed against known attacks, a network administrator has to dedicate to finding the existence of unknown attacks. Therefore, we propose a multistage OC-SVM method to extract such unknown attacks.

3 Multistage OC-SVM

The overall process of proposed method is composed of following steps (Fig. 1).

1. Features Extraction from past traffic data and a real network.

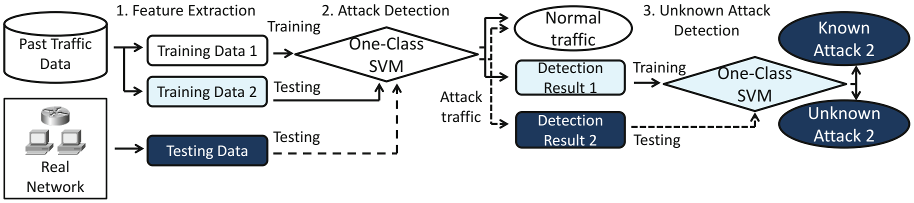

Fig. 1. Overall process

2. Attack detection by the first stage OC-SVM.
3. Classification to known attacks and unknown attacks by the second stage OC-SVM.

3.1 Feature Extraction

At first, we perform feature extraction. The features are extracted by each TCP sessions. Training Data 1 and Training Data 2 are extracted from past traffic data. Training Data 1 is extracted at older date than Training Data 2. Testing Data is extracted from a real network.

Kondo et al. [4] have shown that bot infected computers show different behavior from normal behavior about packet size and communication interval. By generating histograms for reflecting these characteristics, their method detects malicious sessions with Command & Control server by using SVM.

In addition to features of Kyoto2006+ Dataset, this paper introduces the following 6 new features from traffic data. Here, "duration" means the time length of a TCP session.

1. The number of received bytes divided by duration
2. The standard deviation of 1. concerning the past 100 sessions which have the same source IP addresses of the current session
3. The number of sent bytes divided by duration
4. The standard deviation of 3. concerning the past 100 sessions which have the same source IP addresses of the current session
5. The communication interval between the current and the previous sessions which have the same destination IP address.
6. The standard deviation of 5. in the past 100 times sessions

These features are intended to represent characteristics on communication interval. Our proposed method, therefore, analyzes 12 conventional and 6 new features to detect unknown attacks.

3.2 Attack Detection

The first stage classifier of OS-SVM is used to distinguish attack sessions from normal ones. By leaning normal traffic, the classifier creates a hypersphere that includes the large majority of normal sessions. If the classification assigns a testing session inside the hypersphere, the session is considered normal. Otherwise it is considered attack. In order to effectively use OC-SVM, it is mandatory to choose proper parameter ν which adjusts the radius of the hypersphere. For example, if ν is set to 0.1, OC-SVM calculates a hypersphere excluding 10% of data.

Our method requires three different time of traffic data, i.e., Training Data 1, Training Data 2 and Testing Data. Both training data are retrieved from traffic archive and the testing data is collected from the real network. Training data 1 is older than Training Data 2. By using Training Data 1, the first stage classifier learns normal traffic and then, evaluates Training Data 2 and Testing Data to obtain their outlier traffic. The outlier traffic from Training Data 2 and Testing Data is used for the second stage analysis describe in Section 3.3.

3.3 Extraction of Serious Attacks

As shown in Fig. 1, the second stage classifier is trained by using the outlier sessions from Training Data 2. After the training, the classifier obtains hypersphere that includes attacks observed by the time of Training Data 2.

Then the outlier sessions from Testing Data are examined. If the classifier assigns a session outside of the hypersphere, the session can be considered newly unknown attack. Because such a session has not been observed previously, it can be considered that a zero-day attack or a targeted attack causes the session.

By adopting our multi-stage OC-SVM, we can extract the most outlier sessions from the real network. By considering its degree of the outlier, the sessions can be considered the most hazardous and should be deeply analyzed as soon as possible. It means that our method can tell the network administrator the priority to take action to zero-day attacks.

4 Evaluation

4.1 Dataset

Because it is difficult to prepare labeled traffic data in a real network environment, we used Kyoto2006+ Dataset for evaluation. It is obtained from honeypot networks of Kyoto University. In this dataset, 24 features are generated by each TCP session. 24 features consist of 14 conventional features and 10 additional features. The former 14 features are based on features of KDDCup1999 [1] Dataset. These features consist of duration, service type, the number of connections whose source IP address and destination IP address are the same, a rate of "SYN" errors, and so on. The latter 10 features consist of signature-based IDS alerts, IP address, port number, start time of the session. We use 12 conventional features except 2 non-numeric attributes in the former 14 features. We do not use the latter 10 features for detection because these features are non-numeric attribute. In addition to these features, we use 6 features that mentioned in Sec. 3.1.

Table 1. Detection performance when $\nu = 0.06$

		Jan. 20th, 2008		Jan. 30th, 2008	
		True Classification			
		Attack	Normal	Attack	Normal
Detection Result	Attack	656	4,082	475	2,846
	Normal	40	63,168	221	70,257
Detection Rate		94.25%		68.22%	
False Positive Rate		6.07%		3.89%	

In this dataset, most of traffic data are attack traffic so that it does not represent the ratio of attack traffic in practical network. So, we adjusted attack rate to 1% because it is very few in general environment networks.

4.2 Evaluation of the First Stage Classifier

Overview. We selected November 1st in 2007 as training data. The number of normal sessions is 37,730. We selected January 20th and January 30th in 2008 as testing data. The former data have 67,250 normal sessions and 696 attack sessions and the latter data have 73,103 normal sessions and 834 attack sessions. Attack sessions are adjusted to approximately 1% of the normal sessions.

We use 12 conventional features and newly extracted 6 features explained in Section 3.1. For comparison, we also evaluate detection ratio by using only the 12 conventional features.

We define detection rate and false positive rate as following expressions.

$$\text{Detection Rate} = \frac{\#\text{ of sessions classified as an attack}}{\#\text{ of all attack sessions}}$$

$$\text{False Positive Rate} = \frac{\#\text{ of normal sessions classified as an attack}}{\#\text{ of all normal sessions}}$$

Results and Analysis. Fig. 2 shows that ROC curves under ν=\{0.01, 0.02, 0.04, 0.06, 0.08, 0.10\}. This figure also shows the detection results of OC-SVM using only conventional 12 features in order to confirm the effectiveness of new features. The results show that the additional 6 features contributes to obtaining high detection rate and low false positive rate.

Table 1 shows the detailed results of detection on January 20th and 30th, 2008 under $\nu = 0.06$. The detection rate on January 30th is lower than the rate on January 20th. Although parameter ν is the same on two days, the detection rate is significantly different. Accordingly, it is necessary to estimate the appropriate parameter ν to perform high attack detection all dates.

There are some false positive sessions that cannot be found in the result by using only conventional features. Table 2 shows an example of false positive sessions in the result of proposed method. In Table 2, IP address indicates the

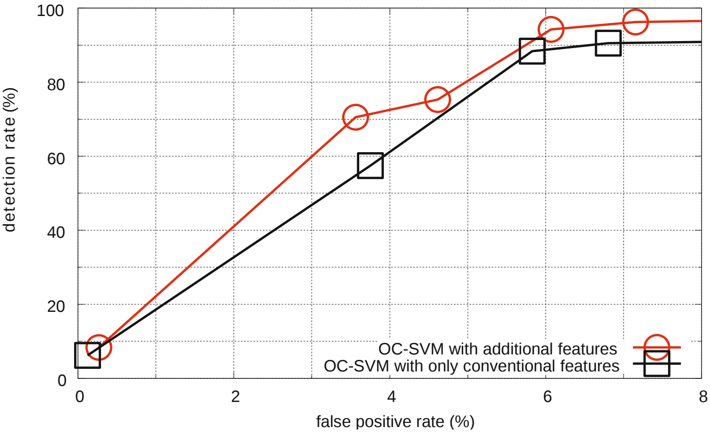

Fig. 2. ROC curve on January 20th, 2008

Table 2. Example of false positive sessions

Time	Duration	Received	Sent	SrcIP:port	DstIP:port
15:21:32	3.86	1,861,381,796	0	6403:9449	0f19:25
15:21:32	7.35	3,513	175	2133:3757	0f19:25
15:21:32	14.07	3,696	175	6110:49062	0f19:25
15:21:32	35.91	0	0	6028:33349	0f19:25
15:21:33	23.95	3,576	175	6110:48371	0f19:25

lower 16 bits of sanitized IPv6 address in Kyoto2006+. Since the 1st line in Table 2 shows that the received bytes of the session is over one million, it heavily affects on its following 100 sessions, i.e., their feature 2 described in Sec. 3.2. For this reason, sessions after huge traffic session were detected as attacks by mistake. To solve this problem, we need to filter out data such a noise before extracting features and training by OC-SVM. One of the filtering methods is grid-based data splitting algorithm [7]. By using such a method, we can exclude a noise data that has too extreme value.

4.3 Evaluation of the Second Stage Classifier

Overview. As similar to Fig. 1, we selected the attack sessions that were detected by the first stage classifier from January 20th to 27th in 2008 as Detection Result 1 for training data of the second stage classifier. Similarly, we selected from February 1st to 7th in 2008 as Detection Result 2 for testing data. We use training data and testing data for one week (longer than previous subsection) to obtain sufficient amount of data because the number of attacks is comparatively small. Although the training data includes normal session data as false positive, they are treated as attacks. Training Data 1 for the first stage classifier is November 1st in 2007 as same as in the previous subsection.

Table 3. Result of unknown attack detection

		True Classification	
		Unknown	Known
Detection Result	Unknown	107	18,559
	Known	43	30,952
Detection Rate		71.33%	
False Positive Rate		37.48%	

Table 4. Result of unknown attack detection by training only attacks

		True Classification	
		Unknown	Known
Detection Result	Unknown	120	10,370
	Known	30	39,141
Detection Rate		80.00%	
False Positive Rate		20.94%	

For the first stage classifier, we set the parameter ν to 0.05 in order to obtain moderate detection rate without extreme high false positive rate at any day. On the other hand, it is expected difference between known and unknown attacks is relatively small. In order to enhance the performance of the second stage classifier, we set larger value to v, i.e., 0.2.

As similar to the previous section, we define detection rate and false positive rate as follows. In this evaluation, normal sessions as false positive caused by the first stage classifier are also treated as known attack sessions.

$$\text{Detection Rate} = \frac{\text{\# of sessions classified as an unknown attack}}{\text{\# of all unknown attack sessions}}$$

$$\text{False Positive Rate} = \frac{\text{\# of attack sessions classified as an unknown attack}}{\text{\# of all known attack sessions}}$$

Results and Analysis. Table 3 shows a result of unknown attack detection. The result shows much higher false positive rate than the detection of the first stage classifier.

From these results, we suspected that the training data for the second stage classifier contain normal sessions as false positive caused by the first stage classifier. These noise session may affect the performance of pthe second stage classifier.

To prove this hypothesis, at first, we counted the number of sessions that were actual attack sessions and actual normal sessions in the training data for second stage classifier. The number of actual attack sessions are 3,547 and the number of actual normal sessions are 21,165. Although the first stage classifier can effectively detect attacks, there still remain huge amount of normal sessions in Detecting Result 1.

To prove effectiveness of our idea, next, we extracted 3,547 actual attack sessions from the Detection Result 1 and trained the second stage classifier by them. As same as the previous evaluation, we set the parameter ν to 0.2. Table 4 shows a result with new training data. The detection rate becomes 80.00% and the false positive rate becomes 20.94% which is much better result than that of Table 3. From this result, we can conclude the high possibility that the false positive of normal sessions becomes noise data. Therefore, if the first classifier has less false positive, we can improve unknown attack detection performance.

5 Conclusion

In this paper, we presented a method to detect unknown attacks using feature extraction and multistage OC-SVM. We added 6 new features based on communication interval, and applied them to OC-SVM.

We evaluated the proposed method with Kyoto2006+ Dataset and the features. By comparing with the first stage classifier without the new features, the classifier with these features shows higher precision rate in detecting attacks .c In the second stage classifier, our method detects unknown attacks, although there is a high false positive rate. A signature-based IDS cannot detect unknown attacks so that current network administrators take too much time and effort to find unknown attacks by analyzing all suspicious sessions. For these reasons, it can be said that our method has enough advantage by limiting on the number of suspicious unknown attacks.

For future works, we need to improve the detection rate of unknown attacks in the second stage classifier. Therefore, we have to perform filtering and clustering in order to reduce an affect of noise data. We also have to extract more effective features th at reflect unique characteristic of unknown attacks.

Acknowledgment. This work is supported by R&D of detective and analytical technology against advanced cyber-attack, administered by the Ministry of Internal Affairs and Communications.

References

1. KDD Cup 1999 Dataset, `http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html`
2. Kyoto2006+ Dataset, `http://www.takakura.com/Kyoto_data/`
3. Eskin, E., Arnold, A., Prerau, M., Portnoy, L., Stolfo, S.: A geometric framework for unsupervised anomaly detection. In: Applications of Data Mining in Computer Security, pp. 77–101. Springer (2002)
4. Kondo, S., Sato, N.: Botnet traffic detection techniques by C&C session classification using SVM. In: Miyaji, A., Kikuchi, H., Rannenberg, K. (eds.) IWSEC 2007. LNCS, vol. 4752, pp. 91–104. Springer, Heidelberg (2007)
5. Perdisci, R., Gu, G., Lee, W.: Using an ensemble of one-class svm classifiers to harden payload-based anomaly detection systems. In: Sixth International Conference on Data Mining, ICDM 2006, pp. 488–498. IEEE (2006)
6. Schölkopf, B., Platt, J., Shawe-Taylor, J., Smola, A., Williamson, R.: Estimating the support of a high-dimensional distribution. Neural Comput. 13(7), 1443–1471 (2001)
7. Song, J., Ohira, K., Takakura, H., Okabe, Y., Kwon, Y.: A clustering method for improving performance of anomaly-based intrusion detection system. IEICE - Trans. Inf. Syst. E91-D(5), 1282–1291 (2008)
8. Song, J., Takakura, H., Kwon, Y.: A generalized feature extraction scheme to detect 0-day attacks via IDS alerts. In: The 2008 International Symposium on Applications and the Internet (SAINT 2008), pp. 55–61 (2008)