2012 IEEE/IPSJ 12th International Symposium on Applications and the Internet

Unknown Attacks Detection Using Feature Extraction from Anomaly-based IDS Alerts

Masaaki Sato
School of Engineering
Nagoya University
Nagoya, 464-8601 Japan
satou@net.itc.nagoya-u.ac.jp

Hirofumi Yamaki
Information Technology Center
Nagoya University
Nagoya, 464-8601 Japan
yamaki@itc.nagoya-u.ac.jp

Hiroki Takakura
Information Technology Center
Nagoya University
Nagoya, 464-8601 Japan
takakura@itc.nagoya-u.ac.jp

***Abstract*—Intrusion Detection Systems (IDSs) play an important role detecting various kinds of attacks and defend our computer systems from them. There are basically two main types of detection techniques: signature-based and anomaly-based. A signature-based IDS cannot detect unknown attacks because a signature has not been written. To overcome this shortcoming, many researchers have been developing anomaly-based IDSs. Although they can detect unknown attacks, there is a problem that they just classify network traffic into normal or abnormal. Therefore, IDS operators have to manually inspect IDS alerts to classify them into known attacks or unknown attacks. Because there are a lot of alerts related to known attacks, it is difficult to extract only unknown attacks from them. In this paper, we present a method that automatically detects unknown attacks from an anomaly-based IDS alerts. We evaluate our method using Kyoto2006+ dataset.**

***Keywords*-intrusion detection system; anomaly detection; unknown attacks;**

I. Introduction

Cyber attacks against our computer systems and networks become more complicated and diverse. Especially, they are suffer from unknown attacks because there are no security advisories, patches or signatures. For early detection of such attacks, Intrusion Detection Systems (IDSs) play an important role. There are two main types of IDSs: signature-based and anomaly-based. Although a signature-based approach is the most widely used in commercial IDSs, we cannot write a signature to detect unknown attacks beforehand. On the other hand, anomaly-based IDSs can detect unknown attacks, but they have problems that a low detection rate and a high false positive rate. To overcome these shortcoming, many researchers have been developing high performance anomaly-based IDSs.

Although Anomaly-based IDSs (AIDSs) can detect unknown attacks, they still have problems except for a detection performance. Since an AIDS just classifies network traffic into normal or abnormal, AIDS operators have to manually inspect an alert to identify whether an unknown attack exists or not. Moreover, AIDS reports a lot of alerts related to known attacks. It is very difficult to manually detect only unknown attacks from AIDS alerts.

In this paper, we present a method that automatically extracts only unknown attacks from AIDS alerts. We modified the existing feature extraction method proposed by Song et al.[1], and also add new features; duration, source bytes and destination bytes. Then, we apply one-class SVM to them. In our experiment, we evaluated our method against two types of unknown attacks. The one is an unknown attack observed in previous research which raises many irrelevant signature-based IDS alerts. The other is an attack that raises Antivirus alerts but does not raise signature-based IDS alerts. We used Kyoto2006+ dataset[2], and our experimental results show that it is capable of detecting unknown attacks detected in previous research and other attacks which signature-based IDSs cannot detect.

The rest of the paper is organized as follows. In Sect.II, we present related work. In Sect.III, we present our method in detail. In Sect.IV, we show the results of experiments and analyze them. Finally, we present concluding remarks and suggestions for future study.

II. Related Work

In recent years, many researchers have been developing anomaly-based IDSs that have a high detection rate and a low false positive rate. Song et al.[3] proposed a method based on clustering and multiple one-class SVM considering one or more normal patterns in training data. Kishimoto et al.[4] proposed a method combining multiple classifiers to cope with different network traffic trends.

Song et al.[1] proposed a method that detects unknown attacks from signature-based IDS alerts using a characteristic that unknown attacks induce many difference kinds of alerts. In general, it is said that signature-based IDSs cannot detect unknown attacks because it is impossible to write signatures beforehand. However, attackers sometimes inject old code into their new program and make IDSs raise many meaningless alerts to hide their real activities. In many cases, their new exploit codes are also crafted by combination of old exploit codes, and thus, each old exploit code raises their corresponding IDS alerts. From the above characteristics on unknown attacks, Song et al. generated

978-0-7695-4737-4/12 $26.00 © 2012 IEEE
DOI 10.1109/SAINT.2012.51

new 7 statistical features using only the 6 basic features of signature-based IDS alerts; detection time, source address and port, destination address and port, and signature name. They apply one-class SVM to new 7 statistical features and detect unknown attacks from signature-based IDS alerts.

Although their method can detect unknown attacks that raise unnatural combinations of signature-based IDS alerts, it cannot detect an unknown attack where it does not raise any IDS alerts. In order to detect such attacks, we extract 10 features from AIDS alerts because they include attacks that signature-based IDSs cannot detect. Among 10 features, 7 features are based on Song's method and 3 features are extracted from AIDS alerts; duration, source bytes, destination bytes.

III. Proposed Method

The overall process of proposed method is composed of following 4 steps (Fig.1).

- Step 1: Anomaly Detection. Detect attack traffic from Kyoto2006+ dataset.
- Step 2: Feature Extraction. Extract 10 features from anomaly-based IDS alerts and make training data and testing data.
- Step 3: Training. Applying one-class SVM to training data.
- Step 4: Testing. Analyze testing data with one-class SVM Model and classify it into unknown or known attacks.

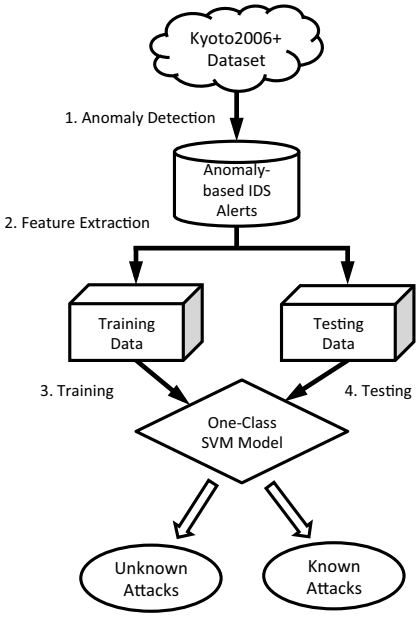

Figure 1. All Process

A. Kyoto2006+ Dataset

In our research, we use Kyoto2006+ dataset[2] that is obtained from a honeypot networks of Kyoto University. In the honeypot networks, several types of honeypots are deployed over 5 different networks which are inside and outside of Kyoto University. There are some different OS (e.g. Windows XP with different patches, Vista, Solaris), network printers and home information appliances (e.g. TV, Video Recorder). Kyoto2006+ dataset deploys a mail server in the same network to collect for normal traffic.

From traffic data of the network, Kyoto 2006+ dataset extracts 14 conventional features and 10 additional features for each session. The former 14 features are extracted based on KDDCup 1999[5] dataset that is widely used for performance evaluation in intrusion detection system. The latter 10 features are extracted for more effective investigation. For example, signature-based IDS alerts, Antivirus alerts, source IP address and port number, time the session was started and so on.

B. Anomaly Detection (Step 1)

We use an anomaly detection method proposed by Song et al.[3] that uses clustering and multiple one-class SVM. The algorithm is as follows.

- Training Phase
  1) Filtering: filter out attack data from the training data.
  2) Clustering: separate the filtered data into k clusters.
  3) Modeling: apply one-class SVM to each k clusters.
- Testing Phase
  1) Dividing: assign the testing data to the closest cluster.
  2) Classifying: classify the test data as normal or attack using corresponding one-class SVM model.

Table I shows a sample of AIDS alerts. Source and Destination IP Address is sanitized from real IPv4 address to one of unique IPv6 address. Signature_IDS_Alerts indicates corresponding signature ID. Parenthesis indicates the number of the same alert observed during the connection.

Table I
Example of Anomaly-based IDS alert

Feature Name	Value
Start_Time	01:23:45
Duration	6.78 sec
Source bytes	123 bytes
Destination bytes	456 bytes
Source_IP_Address	fd58::1234
Source_Port	3521
Destination_IP_Address	fd58::5678
Destination_Port	80
Signature_IDS_Alerts	2008(1)

C. Feature Extraction (Step 2)

After anomaly detection, we extract 10 features from AIDS alerts. We assume the following characteristics of unknown attacks.

1) The number of unknown attacks is very few. Almost all of anomaly-based IDS alerts are triggered by known attacks.
2) Unknown attacks are executed against only a certain destination port. Because if an attacker found a new vulnerability of a certain application, he has to send his new exploit code to corresponding destination port.
3) Executions of unknown attacks take long period. As attackers develop and improve a new exploit code, duration of an attack and code size is frequently changed, and intervals of attacks are irregular.
4) To hide a new exploit code, attackers sometimes inject old codes into their program and make signature-based IDSs raise many meaningless alerts. In many cases, their new exploit codes are also crafted by combination of old exploit codes, and thus, each old exploit code raises their corresponding IDS alerts.

First, we directly use 3 features in AIDS alerts; duration, source bytes and destination bytes. These features are based on third characteristic. If unknown attacks trigger alerts, these features are irregular for each alert. Otherwise, these features have approximately constant values.

The rest of features are extracted from AIDS alerts using a method proposed by Song[1]. They extract 7 statistical features refer to last N alerts whose source IP address and destination port are the same to current alert, e.g. the number of alerts whose destination address is the same to the current alert and rate of the number of alerts whose alert types are different from the current alert. This is based on second characteristic of unknown attacks. Because their method is specialized in feature extraction from signature-based IDS alerts, it cannot extract appropriate feature values from AIDS alerts. For example, consider a situation where a current alert triggers several signatures and past 5 alerts do not trigger any signatures(Table II). From fourth characteristic of unknown attacks, it is highly possible that a current alert is an unknown attack because it triggers some different signatures at the same time. The existing method extracts rate of the number of alerts whose alert types are different from the current alert. In this example, the value is 0. Therefore we changed this feature to the number of different kinds of alerts among past N alerts including a current alert. The value of this feature increases when signature-based IDS raises many different kinds of the alerts, in this example, the value is 4. In this way, we changed the definition of 2 features to extract appropriate feature values from AIDS alerts.

Table II
EXAMPLE OF FEATURE EXTRACTION

Time	Src_Host	Dst_Host	Sig_ID
12:01:30	2.x.x.10 : 3512	10.x.x.5 : 443	0
12:01:35	2.x.x.10 : 3512	10.x.x.2 : 443	0
12:02:10	2.x.x.10 : 3078	10.x.x.9 : 443	0
12:02:42	2.x.x.10 : 5395	10.x.x.5 : 443	0
12:02:48	2.x.x.10 : 2341	10.x.x.5 : 443	0
Current Alert			
12:03:04	2.x.x.10 : 4522	10.x.x.5 : 443	5, 12, 30, 45

Song's method also have another problem that if the number of corresponding alerts does not exceed N, values of all features become 0. Because of first characteristic of unknown attacks, if N increases, it is impossible to extract features from unknown attacks. To avoid this problem, we extract features from alerts that do not exceed N.

D. Detection of Unknown Attacks (Step 3 and 4)

1) Training Phase: In our method, we apply one-class SVM[6] to the above 10 features to detect unknown attacks from anomaly-based IDS alerts. One-class SVM seeks a hypersphere that includes most of training data within it. Because almost all of alerts is known attacks, inside the hypersphere can be considered known attacks, while outside is unknown attacks. When one-class SVM seeks a hypersphere, we can use parameter v which adjusts the radius of the hypersphere. For example, if v is 0.1, one-class SVM seeks a hypersphere excluding 10% of training data. If v is 0.9, it seeks a hypersphere excluding 90% of training data. In our research, we used LIBSVM[7] that is a library for SVM to carry out the experiments with one-class SVM.

2) Testing Phase: In the testing phase, we compare testing data with SVM model. If a data instance is inside the hypersphere, the data is regarded as known attacks. Otherwise, it is regarded as unknown attacks.

IV. EXPERIMENTAL RESULTS

A. Overview

At the beginning, we prepare AIDS alerts using Song's method written in Sect.III-B. As a training data, we use those of November 1st 2007. There are 37,970 nomal instances and 37,730 attack instances. We adjusted attack ratio to 1% because it is very small in general network traffic. As a testing data, we select January 29th, 30th. There are 64,911, 75,103 normal instances and 23,852, 33,018 attack instances, respectively.

Among 14 conventional features in Kyoto2006+ dataset, 2 features (Service[1] and Flag[2]) are not numerical features. Although Song's method cannot use discrete features, these two features may contain important information to identify attack traffic from normal traffic. Thus, we assigned a numerical value to each discrete value and compared detection

[1]the connection's service type
[2]the state of the connection at the time the summary was written

Table III
DETECTION PERFORMANCE USING ONLY 12 NUMERICAL FEATURES

Date	2008/1/29	2008/1/30
DR	90.4%	89.9%
FPR	9.3%	10.9%
NUM	27629	37856

Table IV
DETECTION PERFORMANCE USING 12 FEATURES AND FLAG

Date	2008/1/29	2008/1/30
DR	93.5%	89.9%
FPR	4.2%	6.3%
NUM	25307	34436

Table V
DETECTION PERFORMANCE USING 12 FEATURES AND SERVICE

Date	2008/1/29	2008/1/30
DR	70.6%	80.2%
FPR	5.3%	8.3%
NUM	20288	32760

Detection Time	Src_Host : Port	Dst_Host : Port	Size	Signature IDS
2008/01/30 20:30:15	***.it :30198	Solaris : 80	1647	CVE-1999-0874 CVE-2005-2090 CVE-2005-2088
2008/01/30 20:30:18	***.it : 30172	Solaris : 80	66421	MS-01-016
2008/01/30 20:30:56	***.it :32660	Solaris : 80	1647	CVE-1999-0874 CVE-2005-2090 CVE-2005-2088

Figure 2. Example of Unknown Attacks

performance of using only 12 numerical features and that of also using Flag or Service feature. We set four parameters of Song's method, α, β, k and v, to 0.01, 100, 10 and 0.01, respectively. DR, FPR and NUM in the table represent detection rate, false positive rate and the number of alerts.

By comparing between Tables III and IV, Flag feature improves detection rate and false positive rate, but from Table V, Service feature greatly degrades detection rate. Therefore, we utilize 12 numerical features and Flag feature for anomaly detection.

Next, we extract features written in Sect.III-C from AIDS alerts. We evaluate detection performance of our proposed method from the following two viewpoints.

1) Unknown attacks observed in previous research which raise many signature-based IDS alerts. On January 30th 2008, there are such ones that attack apache on Solaris. Fig.2 shows a part of these attacks. In first row, signature-based IDS triggered three signatures. CVE-1999-0874, CVE-2005-2090 and CVE-2005-2088 indicate vulnerabilities of Microsoft Internet Information Services, Tomcat and Apache, respectively. This is an unnatural situation because these combinations of alerts are irrelevant and old vulnerability of different years are mixed. Therefore, these attacks make signature-based IDSs raise many meaningless alerts to hide their real activities. These unknown attacks were also detected by Song's method.
2) Attacks that raise Antivirus alerts but do not raise signature-based IDS alerts. In case of Kyoto2006+ dataset, Antivirus was directly applied to session data. If Antivirus detected some malicious codes in the session, signature-based IDS should detect them. Therefore, such attacks are regarded as unknown attacks because appropriate signatures for them were not prepared. Note that, Song's method cannot detect these attacks because they do not trigger any signatures.

We regard the above two attacks as unknown attacks and the others as known attacks. We used alerts on January 29th 2008 as the training data, and those on January 30th 2008 as the testing data.

Table VI
PERFORMANCE COMPARISON OF THE PROPOSED METHOD

	Unknown (1)	Unknown (2)	(1) and (2)
Total	13	17	30
Detection Num	11	13	24
Detection Rate	84.6%	76.5%	80.0%

Table VII
DETECTION RESULT

		True Classification	
		Unknown	Known
Detection Result	Unknown	24	1,380
	Known	6	33,026

B. Results

Tables VI and VII are detection results when we set two parameter N and v to 5 and 0.001. Table VI shows the performance against each unknown attack written in Sect.IV-A (1) and (2). Each detection rate is 84.6%, 76.5% and that of both attacks is 80.0%. From Table VII, we detected 1,404 alerts as unknown attacks. Although 1,380 alerts were mistakenly classified into unknown attacks, i.e., false positive, our method can be practically deployed in real environment. Without our method, IDS operators have to manually analyze all AIDS alerts (34,436). Because our method significantly reduces the number of alerts which require manual analysis, it effectively supports their operational overhead.

C. Analysis

In the proposed method, there are 2 parameters, N and v. We performed experiments to investigate whether these parameters affect detection performance or not. First, we set $v = 0.001$ and changed N = 5, 20, 50, 100. Next, we set N = 5 and changed v = 0.001, 0.005, 0.01, 0.02. The experimental results are shown in Table VIII and IX. From Table VIII, we can observe that higher N degrades the detection rate and the false positive rate. In this case,

Table VIII
PERFORMANCE VARIATION ACCORDING TO N

N	DR	FPR
5	**80.0%**	**4.0%**
20	**73.3%**	**4.2%**
50	63.3%	4.4%
100	63.3%	9.0%

Table IX
PERFORMANCE VARIATION ACCORDING TO v

v	DR	FPR
0.001	**80.0%**	**4.0%**
0.005	**80.0%**	**5.1%**
0.01	80.0%	6.9%
0.02	93.3%	20.2%

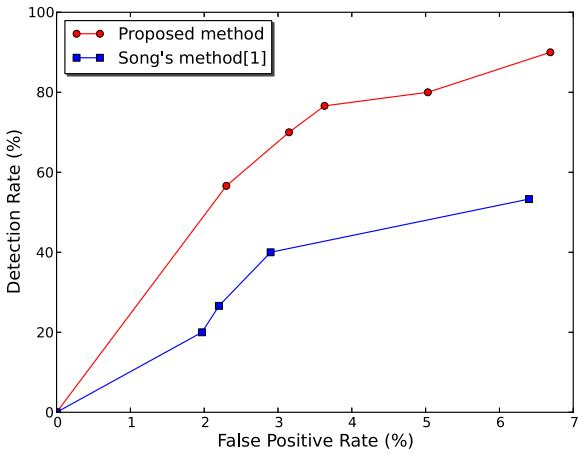

Figure 3. ROC curve over Kyoto2006+ dataset

the performance is the best at N = 5. If we set N under 20, performance of the proposed method is almost constant. From Table IX, higher v improves detection rate, but greatly degrades false positive rate. In one-class SVM, v adjusts the radius of the hypersphere to filter out unknown attacks from training data. We should set v within $0.0001 \sim 0.001$ because the ratio of unknown attacks in the training data is very small.

To verify the effectiveness of the proposed method, we compared our method with Song's method[1] because it can also reduce IDS alerts and extract unknown attacks. We used alerts on January 29th 2008 as the training data, and those on January 30th 2008 as the testing data. Fig.3 shows ROC curve of these methods over Kyoto2006+ dataset. From Fig.3, we can see that the proposed method has a high detection rate compared to Song's method. Table X shows the performance of Song's method against each unknown attack written in Sect.IV-A (1) and (2) when parameter N and v are set to 5 and 0.001. Each detection rate is 69.2%, 17.6% and that of both attacks is 40.0% . From Table X, Song's method cannot detect unknown attacks written in Sect.IV-A (2), which raise Antivirus alerts but do not raise any signature-based IDS Alerts. Therefore, the proposed method has a higher detection performance against unknown attacks which do not trigger any signature-based IDS alerts.

Table X
PERFORMANCE COMPARISON OF SONG'S METHOD

	Unknown (1)	Unknown (2)	(1) and (2)
Total	13	17	30
Detection Num	9	3	12
Detection Rate	69.2%	17.6%	40.0%

V. CONCLUSION

In this paper, we present a method to detect unknown attacks using feature extraction from anomaly-based IDS alerts. We extract 10 features based on the existing method proposed by Song et al., and apply one-class SVM to them.

We have evaluated the proposed method on Kyoto2006+ dataset. The result shows that it detects not only unknown attacks observed in previous research, but also ones that the existing method cannot detect.

For future work, we need to improve the detection rate of unknown attacks that do not trigger any signature-based IDS alerts. Therefore, we have to analyze behaviors of unknown attacks more precisely, and extract other features from anomaly-based IDS alerts.

ACKNOWLEDGMENT

This work was partially supported by Strategic Information and Communications R&D Promotion Programme (091603006).

REFERENCES

[1] J. Song, H. Takakura, and Y. Kwon, "A generalized feature extraction scheme to detect 0-day attacks via ids alerts," in *Applications and the Internet, 2008. SAINT 2008. International Symposium on*. IEEE, 2008, pp. 55–61.

[2] "Kyoto2006+ dataset," http://www.takakura.com/Kyoto_data/.

[3] J. Song, H. Takakura, Y. Okabe, and Y. Kwon, "Unsupervised anomaly detection based on clustering and multiple one-class svm," *IEICE transactions on communications*, vol. 92, no. 6, pp. 1981–1990, 2009.

[4] K. Kishimoto, H. Yamaki, and H. Takakura, "Improving performance of anomaly-based ids by combining multiple classifiers," in *Applications and the Internet (SAINT), 2011 IEEE/IPSJ 11th International Symposium on*. IEEE, 2011, pp. 366–371.

[5] "The third international knowledge discovery and data mining tools competition dataset kdd99-cup," http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html.

[6] B. Schölkopf, J. Platt, J. Shawe-Taylor, A. Smola, and R. Williamson, "Estimating the support of a high-dimensional distribution," *Neural computation*, vol. 13, no. 7, pp. 1443–1471, 2001.

[7] C. Chih-Chung and L. Chih-Jen, "Libsvm: a library for support vector machines," http://www.csie.ntu.edu.tw/~cjlin/libsvm/.